Generalising Kripke Semantics for Quantified Modal Logics

Wolfgang Schwarz
Unfinished draft, 16/09/2013

We turn now to what is arguably one of the least well behaved modal languages ever proposed: first-order modal logic.
[Blackburn and van Benthem 2007]

Contents

1 Introduction

Modal logic has outgrown its philosophical origins. What used to be the logic of possibility and necessity has become topic-neutral, with applications ranging from the validation of computer programs to the study of mathematical proofs. Along the way, *modal predicate logic* has lost its role as the centre of investigation, to the point that it is hardly mentioned in standard textbooks such as [Blackburn et al. 2001]. One reason for this development is that on the standard semantics going back to Kripke, modal predicate logic has turned out to be ill-behaved in several respects. Most notably, applications of the canonical model technique for proving completeness only work for highly restricted classes of models and logical systems, and many interesting systems – e.g., all systems in between quantified S4.3 and S5 ([Ghilardi 1991]) – have been shown to be incomplete, even if their propositional fragment is complete.

Another reason for the decline of modal predicate logic is that propositional modal logic itself has emerged as a fragment of first-order predicate logic, with the domain of "worlds" playing the role of "individuals". As emphasized in [Blackburn et al. 2001], the distinctive character of modal logic is not its subject matter, but its *perspective*. Statements of modal logic describe relational structures from the inside perspective of a particular node. Kripke-style modal predicate logic therefore looks like a somewhat cumbersome hybrid, combining an internal perspective on one class of objects (the domain of the modal operators) with an external perspective on a possibly different class of objects (the domain of quantification).

Nevertheless, this hybrid perspective is useful and natural for many applications – not least the traditional philosophical topics of epistemic, deontic and temporal logic (and the logic of "metaphysical necessity", if there is such a thing). For instance, when reasoning about time, it is natural to take a perspective that is internal to the structure of times, so that what is true at one point may be false at another, but external to the structure of sticks and stones and people existing at the various times.

The main complexities that arise for this approach concern *de re* statements like $\exists x \Diamond Fx$. If we assume that each world or time has an associated domain of individuals that exist at the relevant world or time, we can read $\exists x \Diamond Fx$ as true at w iff w's domain of individuals has a member that satisfies Fx at some point w' accessible from w. Accordingly, if some individual at w does not exist at w', then $\exists x \Diamond \neg \exists y(x = y)$ is true (at w), despite the fact that $\neg \exists y(x = y)$ is not satisfiable in standard predicate logic. This is why modal predicate logics often use free logic as their predicate logic base (see e.g. [Garson 1984]).

A more serious problem for the standard interpretation of modal formulas is that it makes identity non-contingent. For example, if the domain of w contains two individuals x and y, then there can be no other point from the perspective of which these two individuals are identical: $\forall x \forall y(x \neq y \supset \Box x \neq y)$ is valid on the standard semantics. But arguably it is possible to believe of two individuals that they are one and the same, and then we may not want $\forall x \forall y(x \neq y \supset \Box x \neq y)$ to be a theorem of epistemic logic. Moreover, $\forall x \forall y(x \neq y \supset \Box x \neq y)$ is not provable by the standard axioms and rules of first-order logic combined with those of the basic modal logic K. Standard Kripke semantics therefore cannot serve as model theory for one of the simplest modal predicate logics.

David Lewis [1968] once proposed an alternative to Kripke semantics that (among other things) allows for contingent identity. On this account, the *de re* formula $\exists x \Diamond Fx$ is true at w if the domain of w contains an object whose *counterpart* at some accessible point w' satisfies Fx. If two objects at w have the same counterpart at w', $\exists x \exists y(x \neq y \wedge \Diamond(x = y))$ becomes true.

Unfortunately, Lewis combined his proposal with two further ideas, which have prevented its wider adoption in logic and philosophy. First, he held that ordinary individuals never exist at more than one possible world, so that an individual's counterparts at other worlds are never identical to the original individual. The analogous temporal position, defended in [Sider 2001], holds that ordinary individuals are temporarily unextended stages that have other stages as temporal counterparts. This is not how we intuitively think of sticks or stones or persons.

Second, Lewis effectively swapped the traditional, hybrid perspective of Kripke semantics for a thoroughly internal perspective, where statements are evaluated not relative to worlds, but relative to individuals at worlds (see esp. [Lewis 1986: 230-235]). This has curious consequences for the logic determined by his semantics. For instance, the 'necessity of existence', $\Box \exists y(x = y)$, comes out valid, while basic distribution principles such as $\Box(A \wedge B) \supset \Box A$ become invalid (see [Hazen 1979], [Woollaston 1994]). These consequences can be explained, and the internalist approach has been developed into a powerful model-theoretic framework in the hands of Silvio Ghilardi, Giancarlo Meloni and Giovanna Corsi (see [Ghilardi and Meloni 1988], [Ghilardi and Meloni 1991], [Ghilardi 2001] [Corsi 2002], [Braüner and Ghilardi 2007: 591–616]). Nevertheless, it goes against the traditional conception of modal predicate logic.

In this essay, I will investigate a model theory that combines the familiar picture of Kripke semantics with the idea that individuals are tracked across worlds by a counterpart relation. Since the domain of individuals at different worlds need not be distinct, standard Kripke semantics emerges as the special case where counterparthood is identity. As we will see, lifting this constraint on counterprathood results in a simple and intuitive framework that overcomes several shortcomings of standard Kripke semantics.

My proposal is indebted to [Kutz 2000], which develops a similar framework. ([Kracht and Kutz 2002] summarizes of the main results in English.) Unfortunately, the logics proposed by Kutz are incomplete and his completeness proof invalid. Here these problems will be repaired. Unlike Kutz, I will also offer a model theory for negative free logics without outer domains, and for languages with individual constants and object-language substitution operators. (To incorporate individual constants, Kracht and Kutz [2005] switch from counterpart semantics to what Schurz [2011] calls *worldline semantics*, where quantifiers range over functions from worlds to individuals; see also [Kracht and Kutz 2007].)

The basic idea of counterpart semantics can also be motivated more abstractly, by the fact that it takes into account cross-world relations between individuals. Consider a temporal application, where the "worlds" are moments in time. An individual at one time may be an ancestor, or a cause, or an inspiration of an individual at a later time. Such relations connecting individuals from different worlds are nowhere to be found in a Kripke model. In counterpart semantics, they can be represented as counterpart relations.

Several alternatives to Kripke semantics have been proposed to overcome some of these limitations, such as the meta-frame semantics of [Skvortsov and Shehtman 1993] or the category-theoretical semantics of [Ghilardi and Meloni 1988]. However, the semantical and philosophical perspicuity of Kripke semantics is largely absent in these proposals. (Also Ghilardi's models require S4, so they are not fully general.)

In this essay, I will present a conceptually simple and philosophically well-motivated generalisation of Kripke semantics that fixes both the proof-theoretical and the semantical problems. My proposal is in the tradition of counterpart-theoretic semantics, but closer to [Kutz 2000] and [Kracht and Kutz 2002] than to [Ghilardi and Meloni 1988], [Corsi 2002] and [Braüner and Ghilardi 2007].

The basic idea in counterpart-theoretic semantics is that a *de re* formula like $\exists x \Diamond Fx$ is to be evaluated at a world w by looking for an accessible world w' where *some counterpart* of the individual denoted by 'x' at w satisfies F. It should be evident how this helps with the problem of the simplest quantified modal logic: by letting two individuals at one world have the same counterpart at another world, (NNI) can be rendered invalid. More philosophical arguments in support of counterpart-theoretic interpretations of modal and temporal discourse can be found e.g. in [Lewis 1986], [Sider 2001] and [?]. In the latter paper, I also argue that using such an interpretation does not automatically bring on board various philosophically controversial features of Lewis's counterpart theory. For example, I will not impose a requirement of disjoint domains.

2 Counterpart models

Like Kripke semantics, counterpart semantics comes in several flavours. A well-known choice in Kripke semantics is whether the domain of individuals that exist at a world is constant or variable. In counterpart semantics, strict identity across worlds is less important; the more relevant choice is whether every individual that exists at a world should have a counterpart at every other world.

If one allows for individuals to lack counterparts at some worlds, the next question is what can be said about things that don't exist. The alternatives are well-known from free logic. One option is that if x doesn't exist at w, then every atomic predication Fx is false at w. This is known as *negative*, or *single-domain* semantics. Alternatively, one may hold that non-existence is no bar to satisfying predicates, so Fx may be true at some worlds where x doesn't exist and false at others. The extension of F at a world must therefore be specified not only for things that exist at that world, but also for things that don't exist. This is known as *positive*, or *dual-domain* semantics. Both approaches are attractive for certain applications, so I will explore them in tandem.

In positive models, terms are never genuinely empty. Worlds are associated with an *inner domain* of individuals existing at that world, and an *outer domain* of individuals which, although they need not exist, may still fall in the extension of atomic predicates. Every individual at any world has at least one counterpart at every accessible world, if only in the outer domain. In negative models, we want to do without the somewhat ghostly outer domains. Here terms can go genuinely empty. However, in a modal setting, it is not enough to require that whenever x is empty, then atomic predications involving x are false. We must also require that modal predications of the form $\Diamond Fx$ are false for empty x. If $\Diamond Fx$ could be true and $\Diamond Fy$ false although x and y are both empty, we would have to introduce outer domains after all, so that the outer referent of x has an F-counterpart while that of y does not.

As a consequence, the logic of single-domain counterpart models validates some principles that are not derivable from the standard axioms and rules of negative free logic combined with those of the basic modal logic K, notably

(NA) $\neg Ex \supset \Box\neg Ex$,

(TE) $x = y \supset \Box(Ex \supset Ey)$.

Here Ex abbreviates $\exists y(x = y)$. (NA) reflects the fact that non-existent objects don't have any counterparts. (TE) says that if x is identical to y, and x has a counterpart at some accessible world, then y also has a counterpart at that world. If we had outer domains, an individual could have some existing and some non-existing counterparts at a world, which would render (TE) false.

We can offer counterpart models for negative modal predicate logics without (NA) and (TE). These are dual-domain models in which the extension of all predicates, including identity, is restricted to the inner domain. (NA) would then require that individuals which only figure in the outer domain of a world never have counterparts in the inner domain of another world. (TE) would require that if an individual in the inner domain of a world has a counterpart in the inner domain of another world, then all its counterparts at that world are in the inner domain. The two requirements are obviously independent and non-trivial. Hence the axioms (NA) and (TE) are independent of one another and of the standard principles of basic negative free logic combined with K.

((TE) is false at w if $V_w(x) \in D_w$ has both an existing and a non-existing counterpart at some accessible world w'; in this case, some w'-image V' of V at w assigns the non-existing individual to y and the existing individual to x, rendering $Ex \supset Ey$ false at w' under V', wherefore $\Box(Ex \supset Ey)$ is false at w under V.)

I would still need to prove that all axioms of NK are valid in my dual domain models.

In a sense (NA) rules out terms for individuals that only exist at other worlds: if x is empty, then it remains empty whatever world we look at. Why do we nevertheless have a free logic then? Because we need empty names at worlds considered as counterfactual: we want to allow for $Ex \wedge \Diamond\neg Ex$. We can also allow for empty names at the actual world, it's just that not much can be said to distinguish among them. I will return to the possibility of ruling out such names in section 10. In later sections, I will also look at dual-domain counterpart models for negative logics in which one can distinguish non-existent individuals and in which $\neg Ex \supset \Box\neg Ex$ is not valid.

It is important to develop a negative model theory where we have absolute non-existence of individuals at other worlds. For this is what creates the deviant logic in Lewis's semantics and why Ghilardi et al move to typed languages. I want to show that the present approach works even if we let things go out of existence.

At first glance, single-domain, negative semantics looks philosophically much more attractive than positive semantics. Including an outer domain makes quantification over the inner domain look like restricted quantification: quantifiers only range over the existing things, not over all the things. So there are things (in the outer domain) that don't exist. This sounds Meinongian. Moreover, properties allegedly instantiated by non-existing objects tend to be "intensional" properties like *being remembered* whose status as genuine properties is problematic anyway, or they are dubious abstractions from complex sentences, such as *not giving a speech*. Surely it is not the case that Albert Einstein is giving a speech today. But does it follow that Einstein today has the property of not giving a speech? This is at least questionable. If Einstein doesn't exist today, then one might well say that it is neither true that he gives a speech, nor that he has the property of not giving a speech.

However, things look different in a modal context, where the relevant "non-existing" objects are objects existing at other worlds (or times). Then it does not look absurd after all to treat ordinary quantification as a kind of restricted quantification: restricted to what exists at the actual world. We just saw that a modal semantics without outer domains may force us to ban meaningful names for aliens. But it seems that we can very well have names that denote past people so that e.g. "Albert Einstein was friends with Kurt Gödel" is true, while "Adolf Hitler was friends with Kurt Gödel" is false. But if we can name past objects, why can't we quantify over them? Can't we add unrestricted quantifiers? What would such quantifiers range over in counterpart semantics? Formally, they would simply range over the total domain, which contains actual counterparts for all objects at any world. Forbes suggests that these actual counterparts of alien objects might be identified with the alien objects themselves. (One thing this rules out is an alien object having two counterparts at the actual world.) Another motivation for dual domains comes from sentences like

> It might have been the case that Mount Everest doesn't exist while some other mountain exists which, despite not co-existing with Mount Everest, *could* have co-existed with Mount Everest,
>
> $\Diamond(\neg\exists y(y=e) \wedge \exists x\Diamond(\exists y(y=x) \wedge \exists y(y=e)))$,

which seems to require that we can locate a Mount Everest counterpart at a world twice removed from actuality, although there is no counterpart at the intermediate world. Intuitively, the Everest counterpart here should be *once* removed from the actual Everest. Perhaps we are running into expressive limitations of (single-indexed) QML here. In [?], I outline a single-domain negative semantics that attempts to take into account the present worries. It differs in various ways from the semantics considered here, e.g. by taking into account multiple counterpart relations and hybrid operators.

Another question which arises specifically for counterpart-theoretic accounts is whether we want to allow for what Allen Hazen calls "internal relations" (see [Hazen 1979: 328–330], [Lewis 1986: 232f.]). Suppose Queen Elizabeth II is necessarily the daughter of George VI. Does that mean that at every world, every counterpart of Elizabeth is the daughter of every counterpart of George? Arguably not, if we want to allow individuals to have multiple counterparts at some worlds. Consider a possible world that embeds two copies of the actual world, a "left" copy and a "right" copy. Plausibly, both Elizabeth II and George VI have two counterparts in this world, but each of the Elizabeth counterparts is only daughter to one of the George counterparts.

To model this sort of situations, we need to allow for different ways of locating the individuals from one world at another world. Formally, we will have multiple counterpart relations. One relation will link Elizabeth and George to their counterparts in the left copy, another to their counterparts in the right copy. $\Box Gab$ will be true iff, *on every counterpart relation*, all counterparts of a G-related to all counterparts of b.[1]

One might think that (NE) is invalid for single-domain structures with multiple counterpart relations. For now an individual $V_w(x)$ can have an existing counterpart relative to one counterpart relation and no counterpart at all relative to another. But $x = y \supset \Box(Ex \supset Ey)$ is still valid. That's because $w, V \Vdash \Box(Ex \supset Ey)$ means that for all worlds w' and counterpart relations C linking w and w', $Ex \supset Ey$ is true at w' under every interpretation V' that assigns to all variables one of their C-counterparts (or nothing if there is none). If Ex is true relative to some such V' at some w', then so is Ey. The point is that we're not allowed to use different counterpart relations to evaluate different variables within a formula.

Here is the official definition. As usual, a *model* combines an abstract *structure* with an *interpretation* of our language on that structure. The relevant structures are defined as follows.

1 [Hazen 1979] introduces models with multiple counterpart relations, but stipulates that each relation is actually an injective function, in order to validate the necessity of identity and get a traditional logic of 'actually'. Multiple counterpart relations are also used in [Kutz 2000] and [Kracht and Kutz 2002]. It turns out that the introduction of multiple counterpart relations makes little difference to the base logic. In particular, the logic of all positive or negative counterpart models is exactly the same either way. However, multiple counterpart relations will help in the construction of canonical models for stronger logics in section 7, where we will run into a form of Hazen's "problem of internal relations".

Definition 2.1 (Counterpart structure)

A *counterpart structure* is a quintuple $\mathcal{S} = \langle W, R, U, D, K \rangle$, consisting of

1. a non-empty set W (of "points" or "worlds"),
2. a binary ("accessibility") relation R on W,
3. a ("outer domain") function U that assigns to each point $w \in W$ a set U_w,
4. a ("inner domain") function D that assigns to each point $w \in W$ a set $D_w \subseteq U_w$, and
5. a ("counterpart-inducing") function K that assigns to each pair of points $\langle w, w' \rangle \in R$ a non-empty set $K_{w,w'}$ of binary ("counterpart") relations $C \subseteq U_w \times U_{w'}$

such that *either* (i) $D = U$, *or* (ii) all counterpart relations are "total" in the sense that if $C \in K_{w,w'}$, then for each $d \in U_w$ there is a $d' \in U_{w'}$ with dCd'. In case (i), $\mathcal{S}$ is a *single-domain* structure, in case (ii) it is a *total* structure.

If all counterpart relations are total and also $D = U$, then the structure is both single-domain and total.

In negative models, it can happen that wRw' but no individual in U_w has any counterpart at w', either because U_w is empty or because the members of U_w all don't have counterparts at w'. In either case, there will be exactly one counterpart relation $C \in K_{w,w'}$, namely the empty relation containing no ordered pairs at all.

At first, the "counterpart-inducing function" with its associated many counterpart relations may look unfamiliar. Think of this as constructed from a Lewisian counterpart relation in two steps. First, we want to drop Lewis's requirement of disjoint domains, so that an individual can occur in the domain of many or all worlds. It is then not enough to just specify which individuals are counterparts of other individuals. For example, we want to allow for d at w to have d' as its only counterpart at w', and d'' at w', although d' also exists at w''. Now is d' a counterpart of d? In effect, counterparthood turns into a four-place relation between one individual at one world and another (or the same) individual at another world. It proves convenient to represent this by associating each pair of worlds with a "local" counterpart relation between the individuals in the associated domains. Secondly, these local counterpart relation give way to sets of relations in order to allow for internal relations.

The following terminology might help to make all this look more familiar. I will say that in a given model, *d' at w' is a counterpart of d at w* iff there is a $C \in K_{w,w'}$ such that dCd'. Similarly, a pair of individuals *$\langle d'_1, d'_2 \rangle$ at w' is a counterpart of $\langle d_1, d_2 \rangle$ at w* iff there is a $C \in K_{w,w'}$ such that $d_1 C d'_1$ and $d_2 C d'_2$. And so on for larger sequences. (Note that an identity pair $\langle d, d \rangle$ at w has $\langle d'_1, d'_2 \rangle$ at w' as counterpart iff there is a $C \in K_{w,w'}$ such that d is C-related to both d'_1 and d'_2.) As we will see, the interpretation of modal formulas can be spelled out in terms of this "counterpart relation" between sequences rather than the function $K_{w,w'}$ on which it is officially based.

That counterparthood should be extended to sequences is suggested in [Lewis 1983] and [Lewis 1986], in response to the problem of internal relations. In the present framework, counterparthood between sequences is a derivative notion. This has the advantage that it immediately rules out some possibilities that would cause trouble. For example, it can never happen that a pair $\langle d_1, d_2 \rangle$ at w has $\langle d'_1, d'_2 \rangle$ at w' as counterpart although by itself, d_1 at w does not have d'_1 at w' as counterpart. Similarly, it can never happen that $\langle d_1, d_2 \rangle$ at w has $\langle d'_1, d'_2 \rangle$ at w' as counterpart while $\langle d_2, d_1 \rangle$ at w does not have $\langle d'_2, d'_1 \rangle$ at w' as counterpart. We also don't have to worry about "gappy" sequences that arise when some things fail to have counterparts. And we automatically get a sensible answer to the question which sequences, in general, matter for the evaluation of a modal formula $\Box A$ at a world – the sequence of individuals denoted by terms in $\Box A$, in order of appearance?, or in alphabetical order of the corresponding terms?, including repetitions if a term occurs more than once?, or should we also consider sequences involving individuals not mentioned in $\Box A$?

A new possibility arising from multiple counterpart relations is that in negative models, we could have one relation linking d to d' and another relation on which d has no counterpart at all. In positive models, d could be linked to an "inner" individual by one relation and to an "outer" individual by another. Arguably, this should be allowed. For example, it might be that some George counterparts are childless. Relative to this choice of George counterpart, 'Elizabeth' is empty, while it is non-empty relative to other choices of George counterpart.

What consequences does this have for the logic? Does it render (TE) invalid?

Next, we define interpretations of the language of quantified modal logic on counterpart structures. To this end, we first have to say what that language is.

> Definition 2.2 (Languages of QML)
> A set of formulas $\mathcal{L}$ is a *standard language of quantified modal logic* if there are distinct sets of symbols $Var(\mathcal{L})$ (the *variables* of $\mathcal{L}$), $Pred(\mathcal{L})$ (the *predicates* of $\mathcal{L}$), $\{\neg, \supset, \forall, =, \Box\}$ such that $Var(\mathcal{L})$ is countably infinite and $\mathcal{L}$ is generated by the rule
>
> $$Px_1 \ldots x_n \mid x = y \mid \neg A \mid (A \supset B) \mid \forall x A \mid \Box A,$$
>
> where $P \in Pred(\mathcal{L})$, $x, y, x_1, \ldots \in Var(\mathcal{L})$ and every predicate P is limited to a fixed number $n \geq 0$ of variables, called the *arity* of P.

Some notational conventions: I will often use '$\mathcal{L}$' for a fixed but arbitrary language, 'x', 'y', 'z', 'v' (sometimes with indices or dashes) for members of $Var(\mathcal{L})$, and 'F', 'G', 'P' for members of $Pred(\mathcal{L})$ with arity 1, 2 and n, respectively. Formulas involving '$\land$',

'$\vee$', '$\leftrightarrow$', '$\exists$' and '$\Diamond$' are defined by the usual metalinguistic abbreviations. The order of precedence among connectives is $\neg, \wedge, \vee, \supset$; association is to the right. So $A \wedge B \supset C \supset D$ is $(A \wedge B) \supset (C \supset D)$. For any variable x, 'Ex' abbreviates '$\exists y(y = x)$', where y is the alphabetically first variable other than x. '$A_1 \wedge \ldots \wedge A_n$' stands for '$A_1$' if $n = 1$, or for '$(A_1 \wedge \ldots \wedge A_{n-1}) \wedge A_n$' if $n > 1$, or for an arbitrary tautology $\top$ (say, '$x = x \supset x = x$') if $n = 0$. For any expression or set of expressions A, $Var(A)$ is the set of variables in (members of) A, and $Varf(A)$ is the set of variables with free occurrences in (members of) A.

> Definition 2.3 (Interpretation)
> Let $\mathcal{S} = \langle W, R, U, D, C \rangle$ be a counterpart structure and $\mathcal{L}$ a language of quantified modal logic. An *interpretation function V for $\mathcal{L}$ on $\mathcal{S}$* is a function that assigns to each world $w \in W$ a function V_w such that
> (i) for every predicate P of $\mathcal{L}$, $V_w(P) \subseteq U_w^n$,
> (ii) $V_w(=) = \{\langle d, d \rangle : d \in U_w\}$, and
> (iii) for every variable x of $\mathcal{L}$, $V_w(x)$ is either undefined or in U_w.
>
> If $V_w(x)$ is undefined for some w and x, then V is called *partial*, otherwise it is *total*.

For zero-ary predicates P, clause (i) says that $V_w(P) \subseteq U_w^0$. For any U_w, there is exactly one "zero-tuple" in U_w^0, which we may identify with the empty set. So U_w^0 has exactly two subsets, the empty set $\emptyset = 0$ and the unit set of the empty set $\{\emptyset\} = 1$. It is convenient to think of these simply as truth-values.

> Definition 2.4 (Counterpart model)
> A *counterpart model $\mathcal{M}$* for a language $\mathcal{L}$ consists of a counterpart structure $\mathcal{S}$ together with an interpretation function V for $\mathcal{L}$ on $\mathcal{S}$ such that either $\mathcal{S}$ is single-domain or both $\mathcal{S}$ and V are total. In the first case, $\mathcal{M}$ is a *negative* model; in the second case, it is a *positive* model.

Thus a counterpart model is effectively a collection of free first-order models, with relations R and C that link models and their domains. Since modal operators shift the point of evaluation from one world to another along the accessibility relation R, it never matters what counterparts an individual from one world has at another world unless that world is accessible. This is why I officially stipulated in definition 2.1 that counterparthood is only defined between accessible worlds. As a consequence, one could actually drop the accessibility relation R from structures $\langle W, R, U, D, K \rangle$, since R can be recovered from K: $\langle w, w' \rangle \in R$ iff $\langle w, w' \rangle \in Dom(K)$. In positive models, we can also say that wRw' iff there are $d \in U_w, d' \in U_{w'}$ such that for some $C \in K_{w,w'}$, dCd'. However, in negative models, all individuals at w may fail to have counterparts even at accessible worlds w'; there may even be no individuals at all in U_w. In either case, $K_{w,w'}$ will be the empty set, but it will still be defined.

Note that in a negative model, $D_w = U_w$ can be empty. In positive models, D_w may be empty, but U_w must have at least one member, since $V_w(x) \in U_w$.

Variables are non-rigid in the sense that their interpretation is world-relative. However, we will see at the end of this section that the truth-value of a formula A at a world w never depends on what V assigns to variables at worlds $w' \neq w$. For instance, when we evaluate $\Diamond Fx$ at w, we do not check whether Fx is true at some accessible world w', i.e. whether $V_{w'}(x) \in V_{w'}(F)$. Rather, we check whether some individuals at w' that are counterpart-related to $V_w(x)$ are in $V_{w'}(F)$. $V_{w'}(x)$ only enters the picture when we evaluate formulas relative to w'. If we had a designated centre world w_{c} in each model, we could drop the world-relativity of V for individual variables.

Now let's specify how formulas of $\mathcal{L}$ are evaluated at worlds. For the semantics of quantifiers, we need the concept of an x-variant of V.

> Definition 2.5 (Variant)
> Let V and V' be interpretations on a structure $\mathcal{S}$. V' is an *x-variant of V on w* if V' differs from V at most in the value assigned to x at w. V' is an *existential* x-variant of V on w if in addition, $V'(x) \in D_w$.

$\forall xA$ will then be true at a world w under V iff A is true at w under all existential x-variants V' of V on w. This rule allows us to dispose with assignment functions and to use free variables as individual constants, which makes the semantics slightly simpler. You may have noticed that individual constants are not explicitly mentioned in definition 2.2. However, unlike e.g. in [Kripke 1963] and [Lewis 1968], the lack of individual constants plays no important role in the present account. Whenever you want to use an individual constant, simply use a variable that never gets bound. If you want, you may also add a clause to the syntax to the effect that a certain class of variables cannot be bound, and call these variables 'individual constants'. Nothing hangs on this way of handling quantifiers. If you prefer a more traditional treatment with assignment functions and a clear separation between constants and variables, it is trivial to translate between the two approaches (see [Bostock 1997: 81–90]).

At worlds considered as counterfactual, variables denote counterparts of the things they originally denoted. So we need an operation that shifts the value of terms to the counterparts of their original value.

Definition 2.6 (Image)
Let V and V' be interpretations on a structure $\mathcal{S}$. V' is a *w'-image of V at w* (for short, $V_w \triangleright V'_{w'}$) iff

(i) for every world w in $\mathcal{S}$ and predicate P, $V_w(P) = V'_w(P)$, and

(ii) there is a $C \in K_{w,w'}$ such that for every variable x, if $V_w(x)$ is C-related to some $d \in U_{w'}$, then $V'_{w'}(x)$ is some such d, otherwise $V'_{w'}(x)$ is undefined.

If (i) holds, I will also say that V and V' *agree on all predicates.*

$V'_{w'}(x)$ can only be undefined in negative models. In positive models, this cannot happen because $V_w(x)$ is always defined and counterpart relations are total.

Remember that even if no member of U_w has any counterpart at w', $K_{w,w'}$ isn't empty: it still contains the empty set. Consequently, there will still be a w'-image of V at w, namely "empty" interpretation V' with $V'_{w'}(x) = undef$ for all x. This is important e.g. for the interpretation of $\Box\neg Fx$ or $\Box P_0$ with P_0 zero-ary.

When going through proofs, I often need to unpack the truth clause for $\Box B$: "$w, V \Vdash \Box B$ iff $w', V' \Vdash B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$, i.e. for all w', V' such that wRw' and there is a counterpart relation $C \in K_{w,w'}$ such that for all variables z, $V_w(z) C V'_{w'}(z)$." This always looks complicated. It would probably be better to move the interpretation of variables (and names) out of V and not make them world-relative. (The world relativity is never seriously used anyway.) A point of evaluation is then a pair w, s where $s : Var \to U_w$ is a simple assignment function. A w'-image of an assignment function s at w is defined like for V. ... Does that help?

Definition 2.7 (Truth)
The relation $w, V \Vdash_{\mathcal{S}} A$ ("A is *true at w in $\mathcal{S}$ under V*") between a world w in a structure $\mathcal{S}$, an interpretation function V on $\mathcal{S}$, and a sentence A is defined as follows.

$$
\begin{array}{ll}
w, V \Vdash_{\mathcal{S}} Px_1 \ldots x_n & \text{iff } \langle V_w(x_1), \ldots, V_w(x_n) \rangle \in V_w(P). \\
w, V \Vdash_{\mathcal{S}} \neg A & \text{iff } w, V \nVdash_{\mathcal{S}} A. \\
w, V \Vdash_{\mathcal{S}} A \supset B & \text{iff } w, V \nVdash_{\mathcal{S}} A \text{ or } w, V \Vdash_{\mathcal{S}} B. \\
w, V \Vdash_{\mathcal{S}} \forall x A & \text{iff } w, V' \Vdash_{\mathcal{S}} A \text{ for all existential } x\text{-variants } V' \text{ of } V \text{ on } w. \\
w, V \Vdash_{\mathcal{S}} \Box A & \text{iff } w', V' \Vdash_{\mathcal{S}} A \text{ for all } w', V' \text{ such that } wRw' \text{ and } V_w \triangleright V'_{w'}.
\end{array}
$$

I will drop the subscript $\mathcal{S}$ when the structure is clear from context.

Using the terminology introduced above, the clause for the box means that if $A(x)$ is a formula in which only x occurs freely, then in positive models, $\Box A(x)$ is true at w iff all counterparts of

$V_w(x)$ at all accessible worlds satisfy $A(x)$, i.e. iff $A(x)$ is true at all accessible worlds under all assignments of counterparts of $V_w(x)$ to 'x'. Likewise, $\Box A(x, y)$ is true at w iff all counterparts of the pair $\langle V_w(x), V_w(y)\rangle$ at all accessible worlds satisfy $A(x, y)$. And so on. In negative models, if $V_w(x)$ has no counterpart at some accessible world w' relative to some counterpart relation between w and w', the truth of $\Box A(x)$ at w requires that $A(x)$ is also true at w' under an interpretation that leaves 'x' empty.

In dual-domain structures, we could also introduce *outer* quantifiers that range over members of U_w instead of D_w. In Kripke semantics, outer quantifiers are typically understood as "possibilist" quantifiers that range over all possible individuals. Not so in counterpart semantics. (Although with Forbes we could have a corresponding requirement on outer domains, which is only plausible if inner domains are disjoint. xxx?)

As usual, truth at all worlds in a structure (or a class of structures) under all interpretations is called *validity*.

Definition 2.8 (Validity)
A set of $\mathcal{L}$-formulas Γ is *positively valid* in a set Σ of total counterpart structures if $w, V \Vdash_{\mathcal{S}} A$ for all $A \in \Gamma$, $\mathcal{S} = \langle W, R, U, D, L\rangle \in \Sigma$, $w \in W$ and total interpretations V on $\mathcal{S}$.

Γ is *negatively valid* in a set Σ of single-domain counterpart structures if $w, V \Vdash_{\mathcal{S}} A$ for all $A \in \Gamma$, $\mathcal{S} = \langle W, R, U, D, L\rangle \in \Sigma$, $w \in W$ and interpretations V on $\mathcal{S}$.

Definition 2.9 ((Semantic) Consequence)
Let $\mathbb{C}$ be a set of models or structures. A formula A is a *(local) consequence* of a set of formulas Γ *in* $\mathbb{C}$ iff for all worlds w in all models in $\mathbb{C}$, whenever all members of Γ are true at w, then so is A. Two formulas A and B are *(locally) equivalent in* $\mathbb{C}$ iff they are consequences of one another in $\mathbb{C}$.

Now we can prove that the value that V assigns to variables at other worlds never matters when evaluating formulas at a given world. This follows from the following lemma.

Lemma 2.10 (Coincidence lemma)
Let A be a sentence in a language $\mathcal{L}$ of quantified modal logic, w a world in a structure $\mathcal{S}$, and V, V' interpretations for $\mathcal{L}$ on $\mathcal{S}$ such that V and V' agree on all predicates, and $V_w(x) = V'_w(x)$ for every variable x that is free in A. (In this case, I will say that V and V' *agree at w on the variables in A*.) Then

$$w, V \Vdash_{\mathcal{S}} A \text{ iff } w, V' \Vdash_{\mathcal{S}} A.$$

Proof by induction on A.

1. For atomic formulas, the claim is guaranteed directly by definition 2.7.
2. A is $\neg B$. $w, V \Vdash \neg B$ iff $w, V \nVdash B$ by definition 2.7, iff $w, V' \nVdash B$ by induction hypothesis, iff $w, V' \Vdash \neg B$ by definition 2.7.
3. A is $B \supset C$. $w, V \Vdash B \supset C$ iff $w, V \nVdash B$ or $w, V \Vdash C$ by definition 2.7, iff $w, V' \nVdash B$ or $w, V' \Vdash C$ by induction hypothesis, iff $w, V' \Vdash B \supset C$ by definition 2.7.
4. A is $\forall x B$. By definition 2.7, $w, V \Vdash \forall x B$ iff $w, V^* \Vdash B$ for all existential x-variants V^* of V on w. Each such x-variant V^* agrees at w with the x-variant V'^* of V' on w such that $V'^*(x) = V^*(x)$ on all variables in B. Conversely, each existential x-variant V'^* of V' on w agrees at w with the x-variant V^* of V on w with $V^*(x) = V'^*(x)$ on all variables in B. So by induction hypothesis, $w, V^* \Vdash B$ for all existential x-variants V^* of V on w iff $w, V'^* \Vdash B$ for all existential x-variants V'^* of V' on w, iff $w, V' \Vdash \forall x B$ by definition 2.7.
5. A is $\Box B$. By definition 2.7, $w, V \Vdash \Box B$ iff $w', V^* \Vdash B$ for all w', V^* such that wRw' and $V_w \triangleright V^*_{w'}$, where $V_w \triangleright V^*_{w'}$ means that there is a $C \in K_{w,w'}$ such that for every variable x, either $V_w(x) C V^*_{w'}(x)$ or $V_w(x)$ has no C-counterpart at w' and $V^*_{w'}(x)$ is undefined. Since $V_w(x) = V'_w(x)$ for all variables x, each w'-image of V at w agrees with some w'-image of V' on all variables in B and vice versa. So by induction hypothesis, $w', V^* \Vdash B$ for all w', V^* such that wRw' and $V_w \triangleright V^*_{w'}$ iff $w', V'^* \Vdash B$ for all w', V'^* such that wRw' and $V_w \triangleright V'^*_{w'}$, iff $w, V' \Vdash \Box B$ by definition 2.7. ∎

Corollary 2.11 (Locality lemma)
If two interpretations V and V' on a structure $\mathcal{S}$ agree on all predicates and if for all variables x, $V_w(x) = V'_w(x)$, then for any formula A, $w, V \Vdash_{\mathcal{S}} A$ iff $w, V' \Vdash_{\mathcal{S}} A$.

Proof Immediate from lemma 2.10. ∎

Here is an alternative and perhaps more perspicuous way to formulate the interpretation of modal formulas, using counterpart relations between sequences.

Definition 2.12 (Sequence)
Given a set X and a number $n \in \mathbb{N}$, an *n-ary sequence on X* is a partial function $s : \{i : 1 \le i \le n\} \to X$. If a sequence is undefined for some member of its domain, then it is *gappy*, otherwise it is *non-gappy*. (There is exactly one 0-ary sequence on X: the empty set.)

Definition 2.13 (Sequential counterparthood)
For any counterpart structure $\mathcal{S} = \langle W, R, U, D, K \rangle$ and number $n \in \mathbb{N}$, define R^n as the binary relation that holds between two pairs $\langle w, s \rangle, \langle w', s' \rangle$ with $w, w' \in W$ and s, s' n-ary sequences on $U_w, U_{w'}$ respectively iff for some $C \in K_{w,w'}$ and all $i \leq n$, either $s_i C s'_i$ or there is no d with $s_i C d$ in which case s'_i is undefined. Let R^* be the union of all R^n. R^* is the *sequential counterpart relation of* $\mathcal{S}$.

(Note that $R^0 = R$.)

Lemma 2.14 (Sequential semantics)
Let $A(x_1, \ldots, x_n)$ be a formula with free variables $x_1, \ldots, x_n$ (where n may be zero). Given an interpretation V on a structure $\langle W, R, U, D, K \rangle$ and a world $w \in W$, let $V_w(x_1, \ldots, x_n)$ be the sequence of values assigned by V_w to $x_1, \ldots, x_n$ (i.e. the n-ary sequence s on U_w with $s_1 = V_w(x_1), \ldots, s_n = V_w(x_n)$, in some fixed "alphabetical" order of $x_1, \ldots, x_n$.).

The clause for the box in definition 2.7, viz.

$$w, V \Vdash_{\mathcal{S}} \Box A(x_1, \ldots, x_n) \text{ iff } w', V' \Vdash_{\mathcal{S}} A \text{ for all } w', V' \text{ such that } wRw' \text{ and } V_w \triangleright V'_{w'}.$$

is equivalent to

$$w, V \Vdash_{\mathcal{S}} \Box A(x_1, \ldots, x_n) \text{ iff } w', V' \Vdash_{\mathcal{S}} A(x_1, \ldots, x_n) \text{ for all } w', V' \text{ such that } \langle w, V_w(x_1, \ldots, x_n) \rangle R^* \langle w', V'_{w'}(x_1, \ldots, x_n) \rangle,$$

where R^* is the sequential counterpart relation of $\mathcal{S}$.

Proof By definitions 2.7 and 2.6, $w, V \Vdash_{\mathcal{S}} \Box A(x_1, \ldots, x_n)$ iff

(1) $w', V' \Vdash_{\mathcal{S}} A(x_1, \ldots, x_n)$ for all w', V' such that wRw' and for some $C \in K_{w,w'}$ and all variables x, either $V_w(x)$ is C-related to $V'_{w'}(x)$ or there is no d to which $V_w(x)$ is C-related in which case $V'_{w'}(x)$ is undefined.

By lemma 2.11, it doesn't matter what V' assigns to variables not in $A(x_1, \ldots, x_n)$. So (1) is equivalent to

(2) $w', V' \Vdash_{\mathcal{S}} A(x_1, \ldots, x_n)$ for all w', V' such that wRw' and for some $C \in K_{w,w'}$ and all variables $x_i \in x_1, \ldots, x_n$, either $V_w(x_i)$ is C-related to $V'_{w'}(x_i)$ or there is no d to which $V_w(x_i)$ is C-related in which case $V'_{w'}(x_i)$ is undefined.

By definition 6.6, $\langle w, V_w(x_1, \ldots, x_n) \rangle R^* \langle w', V'_{w'}(x_1, \ldots, x_n) \rangle$ iff wRw' and for some $C \in K_{w,w'}$ and all variables $x_i \in x_1, \ldots, x_n$, either $V_w(x_i)$ is C-related to $V'_{w'}(x_i)$ or there is no d to which $V_w(x_i)$ is C-related in which case $V'_{w'}(x_i)$ is undefined. So (2) is equivalent to

(3) $w', V' \Vdash_{\mathcal{S}} A(x_1, \ldots, x_n)$ for all w', V' such that $\langle w, V_w(x_1, \ldots, x_n) \rangle R^* \langle w', V'_{w'}(x_1, \ldots, x_n) \rangle$.
■

Negative models can in a sense be "simulated" by positive models: starting with any negative model, we can create a positive model by adding a "null individual" o to the outer domain U_w of every world w; o never satisfies any atomic predicates and serves as referent of previously empty terms.

Definition 2.15 (Positive transpose)
The *positive transpose* $\mathcal{S}^+$ of a counterpart structure $\mathcal{S} = \langle W, R, U, D, K \rangle$ is the structure $\langle W, R, U^+, D, K^+ \rangle$ with U^+ and K^+ constructed as follows. Let o be an arbitrary individual (say, the smallest ordinal) not in $\bigcup_w D_w$. For all $w \in W$, $U_w^+ = U_w \cup \{o\}$. For all $\langle w, w' \rangle \in R$, $K^+_{w,w'}$ is the set of relations $C^+ \subseteq U_w^+ \times U_{w'}^+$ such that for some $C \in K_{w,w'}$, $C^+ = C \cup \{\langle d, o \rangle : d \in U_w^+$ and there is no $d' \in U_{w'}$ with dCd' $\}$. (So in particular, C^+ contains $\langle o, o \rangle$.)

If V is a (partial or total) interpretation on a structure $\mathcal{S}$, then the *positive transpose* V^+ of V is the interpretation on $\mathcal{S}^+$ that coincides with V except that whenever $V_w(x)$ is undefined for some w and x, then $V_w^+(x) = o$.

Lemma 2.16 (Truth-preservation under transposes)
If $\mathcal{M}$ is a counterpart model consisting of a structure $\mathcal{S} = \langle W, R, U, D, K \rangle$ and an interpretation V of $\mathcal{L}$ on $\mathcal{S}$, and if $\mathcal{S}^+ = \langle W, R, U^+, D, K^+ \rangle$ and V^+ are the positive transposes of $\mathcal{S}$ and V respectively, then for any world $w \in W$ and formula A of $\mathcal{L}$,

$$w, V \Vdash_{\mathcal{S}} A \text{ iff } w, V^+ \Vdash_{\mathcal{S}^+} A.$$

Proof by induction on A.

1. A is $Px_1 \dots x_n$. By definition 2.7, $w, V \Vdash_{\mathcal{S}} Px_1 \dots x_n$ iff $\langle V_w(x_1), \dots, V_w(x_n) \rangle \in V_w(P)$. If all $V_w(x_i)$ are defined, then $V_w^+(x_i) = V_w(x_i)$ and $V_w^+(P) = V_w(P)$ by definition 2.15, and so $\langle V_w(x_1), \dots, V_w(x_n) \rangle \in V_w(P)$ iff $\langle V_w^+(x_1), \dots, V_w^+(x_n) \rangle \in V_w^+(P)$, i.e. iff $w, V^+ \Vdash_{\mathcal{S}^+} Px_1 \dots x_n$. If some $V_w(x_i)$ is undefined, then $\langle V_w(x_1), \dots, V_w(x_n) \rangle$ is undefined and not in $V_w(P)$; so $w, V \nVdash_{\mathcal{S}} Px_1 \dots x_n$. Moreover, in this case $V_w^+(x_i) = o$ and since $V_w(P) = V_w^+(P)$ never contains any tuples involving o, $\langle V_w^+(x_1), \dots, V_w^+(x_n) \rangle \notin V_w^+(P)$. So either way, $w, V \nVdash_{\mathcal{S}} Px_1 \dots x_n$ iff $w, V^+ \Vdash_{\mathcal{S}^+} Px_1 \dots x_n$.
2. A is $\neg B$. $w, V \Vdash_{\mathcal{S}} \neg B$ iff $w, V \nVdash_{\mathcal{S}} B$ by definition 2.7, iff $w, V^+ \nVdash_{\mathcal{S}^+} B$ by induction hypothesis, iff $w, V^+ \Vdash_{\mathcal{S}^+} \neg B$ by definition 2.7.
3. A is $B \supset C$. $w, V \Vdash_{\mathcal{S}} B \supset C$ iff $w, V \nVdash_{\mathcal{S}} B$ or $w, V \Vdash_{\mathcal{S}} C$ by definition 2.7, iff $w, V^+ \nVdash_{\mathcal{S}^+} B$ or $w, V^+ \Vdash_{\mathcal{S}^+} C$ by induction hypothesis, iff $w, V^+ \Vdash_{\mathcal{S}^+} B \supset C$ by definition 2.7.

4. A is $\forall xB$. By definition 2.7, $w, V \Vdash_{\mathcal{S}} \forall xB$ iff $w, V' \Vdash_{\mathcal{S}} B$ for all existential x-variants V' of V on w. These x-variants V' of V correspond one-one to the x-variants ${V^+}'$ of V^+ with ${V^+}'_w(x) = V'_w(x)$. Moreover, for each such pair $V', {V^+}'$, $\langle \mathcal{S}^+, {V^+}' \rangle$ is the positive transpose of $\langle \mathcal{S}, V' \rangle$. By induction hypothesis, $w, V' \Vdash_{\mathcal{S}} B$ iff $w, {V^+}' \Vdash_{\mathcal{S}^+} B$. And so $w, V \Vdash_{\mathcal{S}} \forall xB$ iff $w, {V^+}' \Vdash_{\mathcal{S}^+} B$ for all existential x-variants ${V^+}'$ of V^+ on w, iff $w, V^+ \Vdash_{\mathcal{S}^+} \forall xB$ by definition 2.7.

5. A is $\Box B$. Assume $w, V \Vdash_{\mathcal{S}} \Box B$. By definition 2.7, this means that $w', V' \Vdash_{\mathcal{S}} B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$. We need to show that $w', {V^+}' \Vdash_{\mathcal{S}^+} B$ for all $w', {V^+}'$ with wRw' and $V^+_w \triangleright {V^+}'_{w'}$. So let $w', {V^+}'$ be such that wRw' and $V^+_w \triangleright {V^+}'_{w'}$. Since V^+ is a total interpretation and $\mathcal{S}^+$ a total structure, this means that for every variable x there is a $C^+ \in K^+_{w,w'}$ with $V^+_w(x) C^+ {V^+}'_{w'}(x)$. Let V' be the interpretation on $\mathcal{S}$ that coincides with ${V^+}'$ except that $V'_w(x)$ is undefined for every world $w \in W$ and variable x for which ${V^+}'_w(x) = o$. Let $C = \{\langle d, d' \rangle \in C : d \neq o \text{ and } d' \neq o\}$. Now assume there are d, d' with $V_w(x) = d$ and dCd'. Then $V^+_w(x) = d$ and dC^+d' and thus ${V^+}'_{w'}(x) \neq o$, for $\langle d, o \rangle \in C^+$ only if there is no d' with $\langle d, d' \rangle \in C$. So $V'_{w'}(x) = {V^+}'_{w'}(x)$, and $V_w(x) C V'_{w'}(x)$. On the other hand, assume there are no d, d' with $V_w(x) = d$ and dCd', either because $V_w(x)$ is undefined or because $V_w(x) = d$ and the only d' with $\langle d, d' \rangle \in C^+$ is o. Either way, then ${V^+}'_{w'}(x) = o$, and so $V'_{w'}(x)$ is undefined. So for all variables x, if there are d, d' with $V_w(x) = d$ and dCd' then $V_w(x) C V'_{w'}(x)$, otherwise $V'_{w'}(x)$ is undefined. Since $C \in K_{w,w'}$ by construction of K^+ (definition 2.15), this means that $V_w \triangleright V'_{w'}$. But ${V^+}'$ is the positive transpose of V'. So we've shown that whenever $V^+_w \triangleright {V^+}'_{w'}$, then there is a V' such that ${V^+}'$ is the positive transpose of V' and $V_w \triangleright V'_{w'}$. We know that $w', V' \Vdash_{\mathcal{S}} B$. So by induction hypothesis, $w', {V^+}' \Vdash_{\mathcal{S}^+} B$. That is, for each $w', {V^+}'$ with wRw' and $V^+_w \triangleright {V^+}'_{w'}$, $w', {V^+}' \Vdash_{\mathcal{S}^+} B$. By definition 2.7, this means that $w, V^+ \Vdash_{\mathcal{S}^+} \Box B$.

   In the other direction, assume $w, V^+ \Vdash_{\mathcal{S}^+} \Box B$. That is, $w', {V^+}' \Vdash_{\mathcal{S}^+} B$ for each $w', {V^+}'$ with wRw' and $V^+_w \triangleright {V^+}'_{w'}$. We have to show that $w', V' \Vdash_{\mathcal{S}} B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$. So let w', V' be such that wRw' and $V_w \triangleright V'_{w'}$. The latter means that there is a $C \in K_{w,w'}$ such that for every variable x, either $V_w(x) C V'_{w'}(x)$ or $V_w(x)$ has no C-counterpart at w' and $V'_{w'}(x)$ is undefined. Let ${V'}^+$ be the positive transform of V'. Let $C^+ = C \cup \{\langle d, o \rangle : d \in U^+_w \text{ and there is no } d' \in U_{w'} \text{ with } dCd' \}$. By definition 2.15, $C^+ \in K^+_{w,w'}$. For any variable x, if $V_w(x) C V'_{w'}(x)$, then both $V_w(x)$ and $V'_{w'}(x)$ are defined and thus $V^+_w(x) = V_w(x)$ and ${V'}^+_{w'}(x) = V'_{w'}(x)$ by definition 2.15; moreover, then $V^+_w(x) C^+ {V'}^+_{w'}(x)$ since $C \subseteq C^+$. On the other hand, if $V_w(x)$ has no C-counterpart at w', so that $V'_{w'}(x)$ is undefined, then by construction of C^+ and V^+_w, $V^+_w(x)$ (which equals $V_w(x)$ if $V_w(x)$ is defined, else o) has o as C^+-counterpart at w'; and ${V'}^+_{w'}(x) = o$; so again $V^+_w(x) C^+ {V'}^+_{w'}(x)$. So for every variable x, there is a $C^+ \in K^+_{w,w'}$ with $V^+_w(x) C^+ {V'}^+_{w'}(x)$, and so $V^+_w \triangleright {V'}^+_{w'}$. Now we know that $w', {V^+}' \Vdash_{\mathcal{S}^+} B$ for all $w', {V^+}'$ with wRw' and $V^+_w \triangleright {V^+}'_{w'}$. Hence $w', {V'}^+ \Vdash_{\mathcal{S}^+} B$. By induction hypothesis, $w', V' \Vdash_{\mathcal{S}} B$. So we've shown that whenever wRw' and $V_w \triangleright V'_{w'}$, then $w', V' \Vdash_{\mathcal{S}} B$. By definition 2.7, this means that $w, V \Vdash_{\mathcal{S}} \Box B$. ■

3 Substitution

Before we look at logical systems for our models, we need to talk a little bit about substitution.

Modal operators shift the point of evaluation: whether $\Box A$ or $\Diamond A$ is true at w often depends on what is true at other points w'. In counterpart semantics, when the point of evaluation is shifted from w to w', the semantic value of every individual constant and variable shifts to the counterpart of the previous value, following some counterpart relation C. If an individual at w has no C-counterpart at w', the relevant terms become empty. If an individual has multiple C-counterparts, we might think of the corresponding terms as becoming "ambiguous", denoting all the counterparts at the same time. To verify $\Box Fx$, we require that Fx is true at all accessible worlds under all "disambiguations".

An important question now is whether these disambiguations are uniform or mixed: should $\Box Gxx$ be true iff at all accessible worlds (for all counterpart relations C) all x counterparts are G-related to *themselves* (uniform) or to *one another* (mixed)? On the mixed account, $\Box x = x$ becomes invalid, as does $\Box(Fx \lor \neg Fx)$, even if x exists at all worlds. (One can still have the corresponding unboxed principles if one restricts Necessitation.) Moreover, the semantics becomes more complicated because a mixed disambiguation cannot be represented by a standard interpretation function; so if we say that $\Box A$ is true relative to interpretation V iff A is true at all accessible worlds under all interpretation functions V' suitably related to V, we automatically get uniform disambiguations. Thus I have used uniform disambiguations in the previous section. (See section **??** for the alternative route.)

The present issue might remind you of the old observation that a sentence like 'Brutus killed himself' can be understood either as an application of a *monadic* predicate 'killing himself' to the subject Brutus, or as an application of the binary 'killing' to Brutus and Brutus. Peter Geach once suggested a syntactic mechanism for distinguishing these readings, by introducing an operator $\langle z : x, y\rangle$ that turns a binary expression into a unary expression: while Gxy is satisfied by pairs of individuals as values of x and y, $\langle z : x, y\rangle Gxy$ is satisfied by a single individual for z. The operator $\langle z : x, y\rangle$, which might be read 'z is an x and a y such that' acts as a quantifier that binds both x and y.

A similar trick can be used in our modal context. On the uniform reading, $\Box x = x$ says that all counterparts of x are self-identical at all accessible worlds. To say that at all accessible worlds (and under all counterpart relations), all x-counterparts are identical to all x-counterparts we could instead say $\langle x : y, z\rangle \Box y = z$. The effect of $\langle x : y, z\rangle$ is to introduce two variables y and z that co-refer with x. By using distinct but co-refering variables in a modal context, we can express relations between possibly distinct counterparts; by using the same variable, we make sure that the same counterpart must be assigned to every occurrence.

With $\langle x : y, z \rangle \Box y = z$, we actually end up with *three* co-referring variables: y and z are made to co-refer with x, but we also have x itself. Thus the job can also be done with $\langle x : y \rangle \Box x = y$ – read: 'x is a y such that ...'. (See xxx on Lowe and Geach.)

To see the use of this operator, consider the following two sentences, which look at first glance like simple applications of universal instantiation.

(1) $\forall x \Box Gxy \supset \Box Gyy$;

(2) $\forall x \Diamond Gxy \supset \Diamond Gyy$.

Suppose for a moment that we have at most one counterpart relation from any world to another, so that we can ignore the quantification over counterpart relations. The first formula then says that if all things x are such that all x-counterparts are G-related to all y-counterparts, then all y-counterparts are G-related to themselves. That must be true. (2), however, is not valid. If all things x are such that some x-counterpart is G-related to some y-counterpart, it only follows that some y-counterpart is G-related to some y-counterpart; it does not follow that some y-counterpart is G-related to itself.

With the two distinct variables x and y, the formula $\Diamond Gxy$ looks at arbitrary combinations of x-counterparts and y-counterparts, even if $x = y$. By contrast, $\Diamond Gyy$ only looks at single y counterparts and checks whether one of them is G-related to itself. To prevent this accidental "capturing" of y in the consequent of (2), we can use the Geach quantifier:

(2′) $\forall x \Diamond Gxy \supset \langle y : x \rangle \Diamond Gxy$

Having multiple counterpart relations makes no essential difference to these considerations. Since Gyy is a formula of type $A(y)$, $\Diamond Gyy$ is true at w iff Gyy is true at some accessible world under some assignment of a y-counterpart to 'y'. On the other hand, given $x = y$, $\Diamond Gxy$ is true at w iff there is an accessible world at which some counterpart of the pair $\langle x, y \rangle$ $(= \langle x, x \rangle = \langle y, y \rangle)$ satisfies Gxy. (Officially: iff there is some accessible w', counterpart relation $C \in K_{w,w'}$ and variant V' of the original interpretation function V such that $V'_{w'}(x)$ is a C-counterpart of $V_w(x)$, $V'_{w'}(y)$ is another C-counterpart of $V_w(x)$ $(= V_w(y))$, and Gxy is true relative to w', V'.) More abstractly, since the point of evaluation always shifts along some particular counterpart relation, nothing very interesting happens when an individual has several counterparts relative to different counterpart relations. In a sense, the real points of evaluations are worlds-under-a-counterpart-relation.

The same issues arise for Leibniz' Law. In the pair

(3) $x = y \supset \Box Gxy \supset \Box Gyy$;

(4) $x = y \supset \Diamond Gxy \supset \Diamond Gyy$,

only the first sentence is valid. In (4), the substituted variable y again gets captured by the other occurrence of y in the scope of the diamond. To avoid this, we should write

$$(4') \qquad x = y \supset \Diamond Gxy \supset \langle y : x \rangle \Diamond Gxy$$

in place of (4). The Geach quantifier $\langle y : x \rangle$ thus functions as an *object-language substitution operator*.

If we want to use this operator, we have to extend the language of quantified modal logic.

> Definition 3.1 (Languages of QML with substitution)
> A *language of quantified modal logic with substitution* is like a standard language of quantified modal logic (definition 2.2) except that there is a further logical constant $\langle \, : \, \rangle$ with the construction rule that whenever x, y are variables and A is a formula of $\mathcal{L}$, then $\langle y : x \rangle A$ is a formula of $\mathcal{L}$.

As for the semantics: just as $\forall x A$ is true relative to an interpretation V iff A is true relative to all x-variants of V (on the relevant domain), $\langle y : x \rangle A$ is true relative to V iff A is true relative to the x-variant of V that maps x to $V(y)$. In our modal framework:

> Definition 3.2 (Semantics for the substitution operator)
> $w, V \Vdash \langle y : x \rangle A$ iff $w, V' \Vdash A$, where V' is the x-variant of V on w with $V'_w(x) = V_w(y)$.

Note that V' need not be an *existential* x-variant of V on w.

The coincidence lemma 2.10 is easily adjusted to languages with substitution, and corollary 2.11 follows as before. I won't go through the whole proof again. Here is the only new step in the induction:

> A is $\langle y : x \rangle B$. $w, V \Vdash \langle y : x \rangle B$ iff $w, V^* \Vdash B$ where V^* is the x-variant of V on w with $V^*_w(x) = V_w(y)$. Let V'^* be the x-variant of V' on w with $V'^*_w(x) = V'_w(y)$. Then V^* and V'^* agree at w on all variables in B, so by induction hypothesis, $w, V^* \Vdash B$ iff $V'^*, w \Vdash B$. And this holds iff $w, V' \Vdash \langle y : x \rangle B$ by the semantics of $\langle y : x \rangle$. ■

If we were to restrict the x-variant to D_w, we'd get something like this.

Definition: inner substitution

$w, V \Vdash \langle y | x \rangle A$ iff $V(y) \in D_w$ and $w, V' \Vdash A$, where V' is the (existential) x-variant of V on w with $V'_w(x) = V_w(y)$.

Unlike $\langle y : x\rangle A$, $\langle y|x\rangle A$ is easily definable, as $\exists z(y{=}z \wedge A)$.

Inner substitution does not in general satisfy the substitution lemma. It does so only relative to worlds and interpretations that verify Ey for all variables y. If $\neg Ey$, then $[y/x]A$ comes out false, whether A is Fx or $\neg Fx$ or $\Box(p \supset p)$. In this respect, $\langle y|x\rangle A$ functions more like a lambda expression $(\lambda x.A)(y)$.

The question is whether the use of $[y/x]$ in (FUI) and (LL) really requires a proper substitution operator, or whether we could just use $\langle y|x\rangle$ instead. For instance, (FUI) would then become

$$\forall xA \supset Ey \supset \exists z(y{=}z \wedge A).$$

This seems harmless, because the antecedent Ey guarantees anyway that y exists. Things more difficult for (LL).

In local classical systems, we do have Ey as a theorem. So the substitution lemma holds with $\langle y|x\rangle$ for initial interpretations V, but not for "counterfactual" images V'. Nevertheless, the relevant versions of (UI) and (LL) remain sound. $\Box(FUI)$ is also sound. $\Box(LL)$ is sound only on negative interpretations. Otherwise $\Box x = y$ and $\Box\neg Fx$ can be true and $\Box\exists x(x = y \wedge \neg Fx)$ false if $V(x)$ and $V(y)$ have the same unique counterpart at all worlds, all of which are $\neg F$, and some of which don't exist. OTOH, $\Box(\forall LL)$ still holds.

Many Substitution Principles involving modal operators fail with inner substitution. In particular, we lose "Continuity" for counterfactual interpretations V: $[y/x]\Box\neg Fx$ doesn't imply that y exists at all worlds, but $\Box[y/x]\neg Fx$ does. Indeed, (Cont) turns into

$$\exists y(y{=}x \wedge \Box A) \supset \Box\exists y(y{=}x \wedge A),$$

which entails the necessity of existence by using a tautology for A.

(Note BTW that the existential suggestion as it stands won't work for $\langle x : x\rangle$, which *should* be a redundant operator, certainly we don't want $\langle x : x\rangle A$ to be $\exists x(x{=}x \wedge A)$, i.e. $\exists xA$. It also needs to be fixed for sequences. E.g. $\langle x, y : y, x\rangle A$ shouldn't be $\exists x\exists y(x{=}y \wedge y{=}x \wedge A)$, nor $\exists x(x{=}y \wedge \ldots)$. The problem is that the variables on the right of the column need to be bound without capturing any variables on the left. So if the RHS variables $\underline{y}$ occur on the left, I need to first rename $\underline{y}$ to some new variables $\underline{y}'$. (See below on polyadic substitution quantifiers. The fix mentioned there can also be applied to $\langle x : x\rangle$, where it yields $\langle x : z\rangle\langle z : x\rangle$.))

When we go through the canonical model section with $\langle y|x\rangle$ in place of $\langle y : x\rangle$, we run into trouble. Consider the crucial clause $A = \Box B$ of the truth lemma. From RTL, we want to show that if $\Box B \in w$, then $w, V \Vdash \Box B$. The proof below goes like this.

Assume $w, V \nVdash \Box B$. So $w', V' \nVdash B$ for some w', V' such that wRw' and V' is such that for all variables y, if there is a $z \in V_w(y)$ with $z^\tau{=}z^\tau \in w'$, then there is a $z \in V_w(y)$ with $z^\tau \in V'_{w'}(y)$, else $V'_{w'}(y)$ is undefined.

Let $*$ be a substitution that maps each variable y in B for which there is a $z \in V_w(y)$ with $z^\tau \in V'_{w'}(y)$ to some such $z \in V_w(y)$ with $z^\tau \in V'_{w'}(z)$, and that maps every other variable to itself. So if $y \in Var(B)$ and $V'_{w'}(y)$ is defined, then $(*y)^\tau \in V'_{w'}(y)$, and so $V'_{w'}(y) = [(*y)^\tau]_{w'} = V^{\tau,*}_{w'}(y)$. If $V'_{w'}(y)$ is undefined, then $V^{\tau,*}_{w'}(y) = V^{\tau}_{w'}(y)$ is undefined, as otherwise $V^{\tau}_{w'}(y) = [y^\tau]_{w'} \neq \emptyset$ and $y^\tau = y^\tau \in w'$. So V' and $V^{\tau,*}$ agree at w' on all variables in B. By the coincidence lemma, $w', V^{\tau,*} \nVdash B$.

Suppose for reductio that $\Box B \in w$. Let $y_1, \ldots, y_n$ be the variables y in B with $(*y)^\tau \in V'_{w'}(y)$. For each such y, $*y \in V_w(y)$, and so $y = *y \in w$. By (LL^g), xxx which generalises (LL^s)

for sequences, $\langle *y_1, \ldots, *y_n : y_1, \ldots, y_n \rangle \Box B \in w$. By (S$\Box$), $\Box \langle *y_1, \ldots, *y_n : y_1, \ldots, y_n \rangle B \in w$. So by construction of R, $(\langle *y_1, \ldots, *y_n : y_1, \ldots, y_n \rangle B)^\tau \in w'$. By induction hypothesis, then $w', V \Vdash (\langle *y_1, \ldots, *y_n : y_1, \ldots, y_n \rangle B)^\tau$, and by the transformation lemma 3.13, $w', V^\tau \Vdash \langle *y_1, \ldots, *y_n : y_1, \ldots, y_n \rangle B$. And then $w', V^{\tau \cdot [*y_1, \ldots, *y_n / y_1, \ldots, y_n]} \Vdash B$ by lemma 3.15. But $[*y_1, \ldots, *y_n : y_1, \ldots, y_n]$ is $*$. (If $(*y)^\tau \notin V'_{w'}(y)$, then $V'_{w'}(y)$ is undefined, and there is no $z \in V_w(y)$ with $z^\tau \in V'_{w'}(y)$, so then $*y = y$.) So $w', V^{\tau \cdot *} \Vdash B$. Contradiction.

With inner substitution, this doesn't work. Here, too, we want $y = x \wedge \Box\neg Gxy \Rightarrow \langle y : x \rangle \Box\neg Gxy \Rightarrow \Box\neg Gyy$, without $\Box Ey$. But we don't want $y = x \wedge \Box\Diamond\neg Gxy \Rightarrow \Box\Diamond\neg Gyy$. (Or, for $\langle y : x \rangle \Box B$ formulas: we want $\forall x \Box\neg Gxy \Rightarrow \langle y : x \rangle \Box\neg Gxy \Rightarrow \Box\neg Gyy$, without $\Box Ey$.) We'd need some further axiom to go from $\langle y : x \rangle \Box\neg Gxy$ to $\Box\neg Gyy$, but not from $\langle y : x \rangle \Box\Diamond\neg Gxy$ to $\Box\Diamond\neg Gyy$.

And what about the last step, in which substitution semantics is applied even though some of the y^* may not even be self-identical at w under V^τ, let alone be in $V^\tau_w(E)$?

Substitution operators turn out to have significant expressive power. As [Kuhn 1980] shows (in effect), if a language has substitution operators, it no longer needs variables or individual constants in its atomic formulas: instead of Fx, we can simply say F, with the convention that the implicit variable is always x (more generally, the first one is x, the second y, etc.); Fy turns into $\langle y : x \rangle F$, Gyz into $\langle y : x \rangle \langle z : y \rangle G$. Similarly, $\forall x F x$ can be replaced by $\forall F$, and $\forall y Gxy$ by $\forall \langle y : z \rangle \langle x : y \rangle \langle z : x \rangle G$. $\langle y : z \rangle \langle x : y \rangle \langle z : x \rangle$ is the swapping operator $\langle x, y : y, x \rangle$. $\forall y Gxy$ is equivalent to $\langle x, y : y, x \rangle \forall x Gyx$. Directly, $\forall y Gxy$ is true at a sequence $abc \ldots$ iff Gxy is true at $ab'c \ldots$ for all b'. Moreover, Gxy is true at $ab'c$ iff $\langle x, y : y, x \rangle Gxy$ is true at $b'ac \ldots$. So we also don't need different quantifiers for different variables any more. In this essay, however, I will not exploit this power of substitution operators – mainly for the sake of familiarity. Our languages with substitution operators will still have ordinary formulas $Px_1 \ldots x_n$ and quantifiers $\forall x, \forall y$, etc.

Substitution operators are close cousins of lambda abstraction and application, as introduced to modal logic in [Stalnaker and Thomason 1968]. Lambda abstraction converts a formula A into a predicate $(\lambda x.A)$, which can then be applied to a singular term y to form a new formula $(\lambda x.A)y$. Semantically, $(\lambda x.A)y$ is true under an interpretation V at a world w iff A is true under the x-variant V' of V on w with $V'_w(y) = V_w(x)$. Thus $(\lambda x.A)y$ is just another way of writing $\langle y : x \rangle A$.

The standard use of lambda abstraction in modal logic is to resolve an ambiguity in modal formulas with non-rigid terms. In English, sentences like 'the pope could have been Italian' have a *de dicto* reading and a *de re* reading. On the *de dicto* reading, the sentence says that it could have been that whoever is pope is Italian; on the *de re* reading, the sentence says of the actual pope that *he* could have been Italian. The ambiguity arises because 'the pope' is a non-rigid designator: relative to every world w, it picks out whoever is pope at w. If α is such a non-rigid designator, then the obvious interpretation of $\Diamond F\alpha$ is the *de dicto* reading: $\Diamond F\alpha$ is true at w iff for some accessible world w', the referent of α at w' falls in the extension of F at w'. To express the *de re* reading, lambda abstraction can be used: $(\lambda x.\Diamond Fx)\alpha$.

In counterpart semantics, the distinction between rigid and non-rigid designators is less straightforward than in standard Kripke semantics. By the rules of the previous section, $\Diamond Fx$ is true at w under interpretation V iff some counterpart of $V_w(x)$ at some accessible w' is in $V_{w'}(F)$. If counterparthood is identity, then x functions as a standard rigid designator. On the other hand, if counterparthood is defined so the pope at w has all and only the popes at other worlds as counterpart, and $V_w(x)$ is the pope at w, then x looks much like 'the pope' in English. This is not a defect. A common argument for counterpart treatments of ordinary modal discourse is precisely that there are multiple legitimate ways of re-identifying an individual at our world at other possible worlds (see e.g. [Lewis 1986: 251–255]).

On the other hand, it would be implausible to explain the ambiguity of 'the pope could have been Italian' as an ambiguity in the counterparthood relation associated with 'the pope'. After all, there is no such ambiguity in 'Jorge Mario Bergoglio could have been an Italian'. Definite descriptions like 'the pope' are non-rigid in a sense that does not carry over to names like 'Jorge Mario Bergoglio'. If we wanted to add non-rigid singular terms α that function like 'the pope' to our modal language, we should say that $\Diamond F\alpha$ is true at w under V iff there is some accessible w' for which $V_{w'}(\alpha) \in V_{w'}(F)$: on its non-rigid interpretation, the referent of 'the pope' at a world w' does not depend on its referent at the actual world w; the term is in a sense directly evaluated at w'.

By the semantics of the previous section, all designators are in this sense rigid. Thus definite descriptions like 'the pope' are not adequately formalised as individual constants in our modal language. We could add a new category of non-rigid "constants", but I will not do consider such an expansion, partly for the sake of simplicity and partly because I think an adequate treatment of descriptions should treat them as composites of a determiner ('the') and a predicate, rather than as atomic terms.

So we don't need lambda abstraction or substitution operators to distinguish *de dicto* and *de re*, because everything is always *de re*. Our main use of substitution operators is rather to distinguish between "mixed" and "uniform" assignments of counterparts. This comes in handy when we look at substitution principles such as Leibniz's Law.

It is well-known that in first-order logic, careless substitution of variables can cause accidental capturing. For example,

(5) $\quad x = y \supset (\exists y(x \neq y) \supset \exists y(y \neq y))$

is not a valid instance of Leibniz's Law, because the variable y gets captured by the quantifier $\exists y$. There are two common ways to respond. One is to define substitution as a simple replacement of variables, and restrict principles like Leibniz's Law:

(LL$^-$) $\;x = y \supset (A \supset [y/x]A)$ provided x is free in A and y is free for x in A, where y is *free for x in A* if no occurrence of x in A lies in the scope of a y-quantifier. The other response is to use a more sophisticated definition of substitution on which $\exists y(y \neq y)$ does not count as a proper substitution instance of $\exists y(x \neq y)$.

Informally, $[y/x]A$ should say about y exactly what A says about x. More precisely, proper substitutions should satisfy the following condition, sometimes called the "substitution lemma":

$$w, V \Vdash [y/x]A \text{ iff } w, V^{[y/x]} \Vdash A,$$

where $V^{[y/x]}$ is like V except that it assigns to x (at any world) the value V assigns to y.

This goal can be achieved by applying the substitution to an alphabetic variant of the original formula in which the bound variables have been renamed so that capturing can't happen. Thus before x is replaced by y in $\exists y(x \neq y)$, the variable y is made free for x by renaming all bound occurrences of y by some new variable z. Instead of (5), a legitimate instance of Leibniz's Law then is

(5′) $\quad x = y \supset (\exists y(x \neq y) \supset \exists z(y \neq z)).$

The following definition uses this idea to prevent capturing of variables by quantifiers.[2]

DEFINITION 3.3 (CLASSICAL SUBSTITUTION)
For any variables x, y, z let $[y/x]z$ be y if $z = x$ (i.e., z is the same variable as x), otherwise $[y/x]z$ is z. For formulas A, define $[y/x]A$ as A if $x = y$; otherwise

$$[y/x]Px_1 \dots x_n = P[y/x]x_1 \dots [y/x]x_n$$

$$[y/x]\neg A = \neg[y/x]A;$$

$$[y/x](A \supset B) = [y/x]A \supset [y/x]B;$$

$$[y/x]\forall zA = \begin{cases} \forall v[y/x][v/z]A & \text{if } z = y \text{ and } x \in \mathit{Varf}(A), \text{ or } z = x \text{ and } y \in \mathit{Varf}(A), \\ \forall [y/x]z[y/x]A & \text{otherwise,} \end{cases}$$

where v is the alphabetically first variable not in $\mathit{Var}(A), x, y$.

$$[y/x]\langle y_2 : z\rangle A = \begin{cases} \langle [y/x]y_2 : v\rangle [y/x][v/z]A & \text{if } z = y \text{ and } x \in \mathit{Varf}(A) \text{ or } z = x \text{ and } y \in \mathit{Varf}(A), \\ \langle [y/x]y_2 : [y/x]z\rangle [y/x]A & \text{otherwise,} \end{cases}$$

where v is the alphabetically first variable not in $\mathit{Var}(A), x, y, y_2$.

$$[y/x]\Box A = \Box[y/x]A.$$

2 See [Bell and Machover 1977: 54-67], [Gabbay et al. 2009: 87–103] for alternative definitions with the same goal. An unconventional aspect of the present definition is that $[y/x]A$ can replace bound occurrences of x. For example, $[y/x]\forall xFx$ is $\forall yFy$. This has the advantage of generalising nicely to polyadic substitutions and especially transformations, defined below.

There are reasons why I chose this particular definition, rather than various alternatives, like that of Bell and Machover. The first reason concerns the proof of the substitution lemma (lemma 3.9 below); the second the generalisation to polyadic substitutions and transformations (definition 3.11).

Perhaps the simplest alternative would be to always replace any bound variable by a new variable before applying the substitution:

$$\begin{aligned}[y/x]\forall zA \; &= \forall v[y/x][v/z]A, \\ &\text{where } v \text{ is the alphabetically first variable not in } \mathit{Var}(A), x, y.\end{aligned}$$

So $[y/x]\forall zFzx$ would be $\forall vFvy$. This satisfies the substitution lemma, but it makes it hard to prove that lemma: to show that $w, V^{[y/x]} \Vdash \forall zA$ iff $w, V \Vdash [y/x]\forall zA$, i.e. iff $w, V \Vdash \forall v[y/x][v/y]A$, we have to show that $w, V^{[y/x]\prime} \Vdash A$ for every z-variant $V^{[y/x]\prime}$ of V iff $w, V' \Vdash [y/x][v/y]A$ for every v-variant V' of V. And for that, we need some rather general form of the alpha-conversion lemma (lemma 3.8 below). But to prove the alpha-conversion lemma, we need the substitution lemma: we have to show that if $w, V \Vdash \forall xA$ and B is an alphabetic variant of A, then $w, V \Vdash \forall z[z/x]B$, where z is new. So suppose $w, V \nVdash \forall z[z/x]B$. Then there is an existential z-variant V^* of V on w such that $w, V^* \nVdash [z/x]B$. Since B is an alphabetic variant of A, $[z/x]B$ is an alphabetic variant of $[z/x]A$. By induction hypothesis, $w, V^* \nVdash [z/x]A$. Let V^+ be the (existential) x-variant of V^* on w with $V^+_w(x) = V^*_w(z)$. So V^+ agrees with $(V^*)^{[z/x]}$ about all variables at w. By the locality lemma (2.11), $w, V^+ \Vdash A$ iff $w, (V^*)^{[z/x]} \Vdash A$. Now we want to say – and here we need the substitution lemma, at least for cases where the substituted variable is new – that $w, (V^*)^{[z/x]} \Vdash A$ iff $w, V^* \Vdash [z/x]A$. Then we have $w, V^+ \nVdash A'$. Letting V' be the z-variant of V^+ on w with $V'_w(z) = V_w(z)$, it would follow by the coincidence lemma (2.10) that $w, V' \nVdash A$. But V' is an existential x-variant of V on w. So by definition 2.7, $w, V \nVdash \forall xA$.

The Bell and Machover definition provides a way out of this conundrum. It goes, in effect, like this.

$$[y/x]\forall zA \; = \begin{cases} \forall z[y/x]A & \text{if } z \notin \{x, y\}, \\ \forall zA & \text{if } z = x, \\ \forall v[y/x][v/z]A & \text{if } z = y \neq x, \end{cases}$$
$$\text{where } v \text{ is the alphabetically first variable not in } \mathit{Var}(A), x, y.$$

(For $\langle y_2 : z\rangle A$, the variable y_2 gets replaced by $[y/x]y_2$ in all three clauses.)

The second clause ensures that bound variables are left alone: $[y/x]\forall xA$ is simply $\forall xA$. The third clause prevents capturing in $[y/x]\forall yA$ by turning it into $[y/x]\forall v[v/y]A$. (Since bound variables are left alone, this is the only way capturing can arise: y can't get captured in $[y/x]\forall xA(y)$.)

Now we can prove a preliminary version of the substitution lemma, for the case where $y \notin \mathit{Var}(A)$, without using the alpha-conversion lemma: assume $A = \forall zB$. Since $y \notin A$, either $z = x$ or $z \notin \{x, y\}$. Assume first that $z = x$. So $A = [y/x]A = \forall xB$. By definition 2.7, $w, V^{[y/x]} \Vdash \forall xB$ iff $w, V^{[y/x]\prime} \Vdash B$ for all existential x-variants $V^{[y/x]\prime}$ of $V^{[y/x]}$ on w. These $V^{[y/x]\prime}$ are precisely the existential x-variants of V on w. So $w, V^{[y/x]} \Vdash \forall xB$ iff $w, V' \Vdash B$ for all existential x-variants V' of V on w, iff $w, V \Vdash \forall xB$ by definition 2.7, iff $w, V \Vdash [y/x]\forall xB$. Alternatively, assume $z \notin \{x, y\}$. By definition 2.7, $w, V^{[y/x]} \Vdash \forall zB$ iff $w, V^{[y/x]\prime} \Vdash B$ for all existential z-variants $V^{[y/x]\prime}$ of $V^{[y/x]}$ on w. These $V^{[y/x]\prime}$ are precisely the functions $V'^{[y/x]}$ where V' is an existential z-variant of

V on w. So $w, V^{[y/x]} \Vdash \forall zB$ iff $w, V'^{[y/x]} \Vdash B$ for all existential z-variants V' of V on w. By induction hypothesis, $w, V'^{[y/x]} \Vdash B$ iff $w, V' \Vdash [y/x]B$. So $w, V^{[y/x]} \Vdash \forall zB$ iff $w, V' \Vdash [y/x]B$ for all existential z-variants V' of V on w, iff $w, V \Vdash \forall z[y/x]B$ by definition 2.7, iff $w, V \Vdash [y/x]\forall zB$.

The downside of the Bell and Machover definition is that its generalisation to polyadic substitutions σ and especially to transformations (injective substitutions) is rather complicated and inelegant. Since bound variables are left alone, we have to define $\sigma(\forall zA)$ as $\forall z\sigma'(A)$, where σ' is like σ except that $\sigma'(z) = z$. Now capturing happens if there is a variable $x \in A$ such that $\sigma(x) = \sigma'(z) = z$. In this case, $\sigma(\forall zA) = \sigma'(\forall zA) = \sigma'(\forall v[v/z]A)$, where v is the first variable not in $\sigma'(A)$. Then there's no capturing in $\sigma'(\forall v[v/z]A)$, so $\sigma'(\forall v[v/z]A) = \forall v\sigma''([v/z]A)$, where σ'' is like σ' except that $\sigma''(v) = v$. Since z does not occur in $[v/z]A$, we can simplify this slightly by letting σ'' be like σ except that $\sigma''(v) = v$. So the generalisation is

$$\sigma(\forall zA) = \begin{cases} \forall z\sigma^z(A) & \text{if there is no variable } x \in \mathit{Var}(A) \text{ with } \sigma(x) = z, \\ \forall v\sigma^v([v/z](A)) & \text{otherwise,} \end{cases}$$

where σ^z (σ^v) is like σ except that $\sigma^z(z) = z$ ($\sigma^v(v) = v$), and v is the alphabetically first variable not in $\sigma^z(A)$.

This in itself is a bit inelegant. It gets worse if we consider injective substitutions, i.e. transformations. Transformations have the advantage that they never give rise to capturing as long as they are indiscriminately applied to all variables in a formula. But on the Bell and Machover definition, bound variables are either left alone or replaced by some entirely new variable. So if τ maps x to y and y to x, then $\tau(\forall yFxy)$ is $\forall vFyv$ rather than simply $\forall xFyx$. We end up with a needlessly complicated account of transformations that is quite unintuitive and complicates the proof of the transformation lemma 3.13. (Although one could probably get lemma 3.13 as a special case of the substitution lemma generalised to polyadic substitutions; but I never prove this lemma.)

My own definition happily replaces bound variables: $[y/x]\forall zA$ is simply $\forall[y/x]z[y/x]A$ unless the substitution would lead to capturing of a previously free variable. Since capturing can't happen with transformations, applying a transformation to a formula really amounts to blindly renaming all variables.

Since we replace bound variables, $[y/x]\forall zA$ can lead to capturing in several ways:

(i) a free occurrence of x gets replaced by y in the scope of a y quantifier, as in $[y/x]\forall yFx \to \forall yFy$; in this case x is free in $\forall zA$ (and thus $x \neq z$), and $[y/x]z = y$ (which, since $x \neq z$, reduces to $z = y$).

(ii) a free occurrence of y in the scope of $\forall x$ ends up in the scope of $\forall y$, e.g. $[y/x]\forall xFy$; i.e. y is free in $\forall zA$ (and thus $z \neq y$) and $[y/x]z = y$ (which reduces to $z = x$).

There are also parallel cases with bound occurrences, as in $[y/x]\forall y\forall xFxy \to \forall y\forall yFyy$. But if we apply the substitution piecemeal, we have $[y/x]\forall y\forall xFxy \to \forall y[y/x]\forall xFxy$, and now the *free* occurrence of y in $\forall xFxy$ gets captured; so this is actually an instance of (ii).

Summarising (i) and (ii), we have: $z = y$ and x is free in $\forall zA$, or $z = x$ and y is free in $\forall zA$. Equivalently: $z = y \neq x$ and x is free in A, or $z = x \neq y$ and y is free in A. Equivalently: z is one of x, y, and the other variable is free in $\forall zA$. Equivalently: $z \in \{x, y\}$ and $\emptyset \neq \{x, y\} \backslash \{z\} \subseteq FV(\forall zA)$. Equivalently: $x \neq y$, z is one of x, y, and the other variable is free in A. By moving the case of $x = y$ out of the recursive definition, stipulating that $[y/x]A = A$, I can simplify slightly to: $z = x$ and $y \in \mathit{Varf}(A)$, or $z = y$ and $x \in \mathit{Varf}(A)$.

What is the generalisation to polyadic substitutions σ? Again, capturing in $\sigma(\forall zA)$ could happen either because some free variable x in A gets mapped onto the bound variable, i.e. to $\sigma(z)$, or the bound variable z gets mapped onto some free variable in A, i.e. some free variable x in A is such that $\sigma(z) = \sigma(x)$. The two cases now coincide. Either way, capturing happens if there is a free variable x in $\forall zA$ such that $\sigma(x) = \sigma(z)$. If this happens, we have to replace $\forall zA$ by an alphabetic variant $\forall v[v/z]A$, and make sure σ applied to $\forall v[v/z]A$ does not touch v:

$$\sigma(\forall zA) = \begin{cases} \forall v\sigma'([v/z]A) & \text{if there is an } x \text{ free in } \forall zA \text{ with } \sigma(x) = \sigma(z), \\ \forall\sigma(z)\sigma(A) & \text{otherwise,} \end{cases}$$

where σ' is like σ except that $\sigma'(v) = v$, and v is the alphabetically first variable not in $\sigma'(A)$.

For the record, another alternative that I was using for a while was this:

$$[y/x]\forall zA = \begin{cases} \forall v[y/x][v/z]A & \text{if } z \in \{x, y\}, \\ \forall z[y/x]A & \text{otherwise,} \end{cases}$$

where v is the alphabetically first variable not in $\mathit{Var}(A), x, y$.

This turns $[x/y]\forall yFy$ into $\forall zFz$, rather than $\forall xFx$. It suffers from the same problem as the simple first definition that always substitutes bound variables.

For a quicker presentation of the results in this study, it might be best to go with the naive definition of substitution and limit principles like (LL) to cases where the substituted variable x is free for y in A. The generalisation of $[y/x]$ to polyadic substitutions is then trivial, and it yields the same result for transformations than my own definition. So the transformation lemma should remain unaffected. Modal freedom would then be strengthened to require that y is free for x, and the logic would otherwise remain the same. Presumably the completeness proof would go through just as well.

To evaluate the clause for $\langle y_2 : z\rangle A$, several cases have to be kept in mind. Let x, y, z, y_2 be mutually distinct unless stated otherwise.

1. $z = y$: $[y/x]\langle y_2 : y\rangle A$.
2. $z = y$ and $y_2 = x$: $[y/x]\langle x : y\rangle A$.
3. $z = x$: $[y/x]\langle y_2 : x\rangle A$.
4. $z = x$ and $y_2 = x$: $[y/x]\langle x : x\rangle A$.
5. $y_2 = x$: $[y/x]\langle x : z\rangle A$.
6. $y_2 = y$: $[y/x]\langle y : z\rangle A$.

It helps to always ask what $\langle y_2 : z\rangle A$ says of x, and check that $[y/x]\langle y_2 : z\rangle A$ says the same of y (ignoring modal contexts, or assuming y is modally free for x in A).

In standard predicate logic, the last two clauses are empty because there is neither a box nor a substitution quantifier; I've added them because we'll need them later. The clauses for the substitution quantifier are exactly parallel to those for the universal quantifier, and the underlying motivation is the same. For example, $[y/x]\langle y_2 : y\rangle x \neq y$ is $\langle y_2 : z\rangle y \neq z$ rather than $\langle y_2 : y\rangle y \neq y$.

The clause for the box treats substitution into modal contexts as unproblematic. However, as Lewis [1983] observed, in counterpart semantics modal operators effectively function as unselective binders that capture all variables in their scope. As a consequence, our definition of substitution does not satisfy the substitution lemma. For instance, $w, V^{[y/x]} \Vdash \Diamond Gxy$ does not imply $w, V \Vdash \Diamond Gyy$. Informally, $\Diamond Gxy$ says about the individual x that at some world, one of its counterparts is G-related to some y-counterpart; $\Diamond Gyy$ does not say the same thing about y, for it says that at some world, one of y's counterparts is G-related to *itself*.[3]

The situation is then similar to that of naive substitution in first-order logic. We can either restrict principles like Leibniz's Law or adjust the definition of substitution so that it satisfies the substitution lemma.

A suitable restriction for substitution principles is that when y is substituted for x in A, y must be *modally free for* x in A, in the following sense.

> DEFINITION 3.4 (MODAL SEPARATION AND MODAL FREEDOM)
> Two variables x and y are *modally separated in* a formula A if no free occurrences of x and y in A lie in the scope of the same modal operator.
>
> (It is allowed that x and y both occur freely in the scope of a modal operator in A, as long as at least one of them is bound from outside that operator, as in $A = \forall x \Box Gxy$.)
>
> y is *modally free for x in A* if either (i) $x = y$, or (ii) x and y are modally separated in A, or (iii) A has the form $\Box B$ and y is modally free for x in B.

For example, x and y are modally separated in $\Box Fx \supset \Diamond Fy$ and in $\forall x \Box Gxy$. y is modally free for x in $\Box x = y$ and $\Box\Box\neg Gxy$ and $\Box\Diamond\neg\exists x Gxy$, but not in $\Box\Diamond\neg Gxy$. Correspondingly,

$x = y \supset (\Box x = y \supset \Box y = y)$ and
$x = y \supset (\Box\Box\neg Gxy \supset \Box\Box\neg Gyy)$ and
$x = y \supset (\Box\Box\neg\exists x Gxy \supset \Box\Box\neg\exists z Gyz)$

3 People sometimes complain that counterpart semantics doesn't validate the necessity of identity

(NI) $\quad x = y \supset (\Box x = x \supset \Box x = y)$,

which is allegedly an instance of Leibniz's Law. But you can't just look at the shape of a formula to see whether it is a legitimate instance of Leibniz's Law. If $\Box$ is synonymous to $\forall y$, then (NI) is a tautological variant of (5) and clearly invalid. Similarly, there can be no argument about whether (NI) is a legitimate instance of Leibniz's Law in counterpart semantics. It definitely isn't.

are valid, while

$$x = y \supset (\Box\Diamond\neg Gxy \supset \Box\Diamond\neg Gyy)$$

is invalid. The first example is a tautological variant of Kutz's $x = y \supset \Diamond x = x \supset \Diamond x = y$.

A natural first shot might be:

> $x = y \supset A \supset [y/x]A$, provided y and x do not both occur in the scope of a diamond in A.

But diamonds come in many disguises. E.g. $\Diamond\neg Fxy$ can also be expressed as $\neg(\top \supset \Box Gxy)$, or as $\Box Gxy \supset \bot$. (Our substitution system handles that correctly: $x = y \supset (\Box Gxy \supset \bot) \Rightarrow \langle y : x\rangle(\Box Gxy \supset \bot) \Rightarrow (\langle y : x\rangle\Box Gxy \supset \langle y : x\rangle\bot) \Rightarrow \neg\langle y : x\rangle\Box Gxy$ [by propcal] $\Rightarrow \langle y : x\rangle\neg\Box Gxy \Rightarrow \langle y : x\rangle\Diamond\neg Gxy$.)

If we don't want to restrict the substitution principles, can we redefine substitution so that it satisfies the substitution lemma? In standard languages of quantified modal logic, this is not easy, as we will see in a moment, after we've spelled out some special conditions under which classical substitution (as defined in definition 3.11) does its job even in counterpart models.

Definition 3.5 (Interpretation under substitution)
For any interpretation V of a language $\mathcal{L}$ on a structure $\mathcal{S}$ and variables x, y of $\mathcal{L}$, $V^{[y/x]}$ is the interpretation that is like V except that for any world w in $\mathcal{S}$, $V_w^{[y/x]}(x) = V_w(y)$.

Lemma 3.6 (Restricted substitution lemma, preliminary version)
Let A be a sentence in a language $\mathcal{L}$ (with or without substitution), $\mathcal{S}$ a counterpart structure for $\mathcal{L}$, w a world in $\mathcal{S}$, V an interpretation on $\mathcal{S}$. Then

$$w, V^{[y/x]} \Vdash_{\mathcal{S}} A \text{ iff } w, V \Vdash_{\mathcal{S}} [y/x]A, \text{ provided } y \notin \mathit{Var}(A).$$

Proof If $y = x$, then $[y/x]A = A$ and $V^{[y/x]} = V$, so the result is trivial. Assume then that $y \neq x$. The proof is by induction on A.

1. A is $Px_1 \ldots x_n$. $w, V^{[y/x]} \Vdash Px_1 \ldots x_n$ iff $\langle V_w^{[y/x]}(x_1), \ldots, V_w^{[y/x]}(x_n)\rangle \in V_w^{[y/x]}(P)$ by definition 2.7, iff $\langle V_w([y/x]x_1), \ldots, V_w([y/x]x_n)\rangle \in V_w(P)$ by definition 3.5, iff $w, V \Vdash P[y/x]x_1 \ldots [y/x]x_n$ by definition 2.7, iff $w, V \Vdash [y/x]Px_1 \ldots x_n$ by definition 3.3.
2. A is $\neg B$. $w, V^{[y/x]} \Vdash \neg B$ iff $w, V^{[y/x]} \nVdash B$ by definition 2.7, iff $w, V \nVdash [y/x]B$ by induction hypothesis, iff $w, V \Vdash [y/x]\neg B$ by definitions 2.7 and 3.3.

3. A is $B \supset C$. $w, V^{[y/x]} \Vdash B \supset C$ iff $w, V^{[y/x]} \nVdash B$ or $w, V^{[y/x]} \Vdash C$ By definition 2.7, iff $w, V \nVdash [y/x]B$ or $w, V \Vdash [y/x]C$ by induction hypothesis, iff $w, V \Vdash [y/x](B \supset C)$ by definitions 2.7 and 3.3.

4. A is $\forall zB$. Since $y \notin Var(A)$, $[y/x]A = \forall[y/x]z[y/x]B$. Assume first that $z \neq x$. By definition 2.7, $w, V^{[y/x]} \Vdash \forall zB$ iff $w, V^{[y/x]\prime} \Vdash B$ for all existential z-variants $V^{[y/x]\prime}$ of $V^{[y/x]}$ on w. These $V^{[y/x]\prime}$ are precisely the functions $V'^{[y/x]}$ where V' is an existential z-variant of V on w. An existential z-variant $V^{[y/x]\prime}$ of $V^{[y/x]}$ on w is an interpretation that maps

   a) z at w to an arbitrary member of D_w, z at w' to $V_{w'}(z)$,
   b) x at any world w' to $V_{w'}(y)$, and
   c) every other variable v at any w' to $V_{w'}(v)$.

   This also characterises the interpretations $V'^{[y/x]}$ where V' is an existential z-variant of V on w. So, $w, V^{[y/x]} \Vdash \forall zB$ iff $w, V'^{[y/x]} \Vdash B$ for all existential z-variants V' of V on w. By induction hypothesis, $w, V'^{[y/x]} \Vdash B$ iff $w, V' \Vdash [y/x]B$. So $w, V^{[y/x]} \Vdash \forall zB$ iff $w, V' \Vdash \forall z[y/x]B$ by definition 2.7, iff $w, V \Vdash [y/x]\forall zB$ by definition 3.3.

   Alternatively, assume $z = x$. By definition 3.3, $[y/x]\forall zB$ is $\forall y[y/x]B$. Assume $w, V \nVdash \forall y[y/x]B$. By definition 2.7, then $w, V' \nVdash [y/x]B$ for some existential y-variant V' of V on w. By induction hypothesis, then $w, V'^{[y/x]} \nVdash B$. Let V^* be the (existential) x-variant of V on w with $V^*_w(x) = V'^{[y/x]}_w(x) = V'_w(y)$. V^* is a y-variant on w of $V'^{[y/x]}$, and y is not free in B, so by the coincidence lemma 2.10, $w, V^* \nVdash B$. But V^* is also an existential x-variant of $V^{[y/x]}$ on w. So $w, V^{[y/x]} \nVdash \forall xB$ by definition 2.7.

   Conversely, assume $w, V^{[y/x]} \nVdash \forall xB$. By definition 2.7, then $w, V^* \nVdash B$ for some existential x-variant of $V^{[y/x]}$ (and thus V) on w. Let V' be the (existential) y-variant of V on w with $V'_w(y) = V^*_w(x)$. Then $V'^{[y/x]}$ and V^* agree at w on all variables except y; in particular, $V'^{[y/x]}_w(x) = V'_w(y) = V^*_w(x)$. Since y is not free in B, by the coincidence lemma 2.10, $w, V'^{[y/x]} \nVdash B$. By induction hypothesis, $w, V' \nVdash [y/x]B$. And since V' is an existential y-variant of V on w, then $w, V \nVdash \forall y[y/x]B$ by definition 2.7.

5. A is $\langle y_2 : z\rangle B$. Since $y \notin Var(A)$, $[y/x]A = \langle [y/x]y_2 : [y/x]z\rangle[y/x]B$ by definition 3.3. By definition 3.2, $w, V \Vdash \langle [y/x]y_2 : [y/x]z\rangle[y/x]B$ iff $w, V' \Vdash [y/x]B$, where V' is the $[y/x]z$-variant of V on w with $V'_w([y/x]z) = V_w([y/x]y_2)$. By induction hypothesis, $w, V' \Vdash [y/x]B$ iff $w, V'^{[y/x]} \Vdash B$. Let $V^{[y/x]\prime}$ be the z-variant of $V^{[y/x]}$ on w with $V^{[y/x]\prime}_w(z) = V^{[y/x]}_w(y_2)$. Then $V^{[y/x]\prime}$ and $V'^{[y/x]}$ agree at w about all variables v in A: for $v = z$, $V^{[y/x]\prime}_w(z) = V_w([y/x]y_2) = V'_w([y/x]z) = V'^{[y/x]}_w(z)$; for $v = x \neq z$, $V^{[y/x]\prime}_w(x) = V^{[y/x]}_w(x) = V_w(y) = V'_w(y)$ (because $y \neq z$ and hence $[y/x]z \neq y$) $= V'_w([y/x]x) = V'^{[y/x]}_w(x)$; and for $v \notin \{x, y, z\}$, $V^{[y/x]\prime}_w(v) = V^{[y/x]}_w(v) = V_w(v) = V'_w(v) = V'^{[y/x]}_w(v)$. So by the coincidence lemma 2.10, $w, V'^{[y/x]} \Vdash B$ iff $w, V^{[y/x]\prime} \Vdash B$. And by definition 3.2, $w, V^{[y/x]\prime} \Vdash B$ iff $w, V^{[y/x]} \Vdash \langle y_2 : z\rangle B$.

6. A is $\Box B$. By definition 2.7, $w, V^{[y/x]} \Vdash \Box B$ iff $w', V^{[y/x]\prime} \Vdash B$ for all $w', V^{[y/x]\prime}$ with wRw' and $V^{[y/x]}_w \triangleright V^{[y/x]\prime}_{w'}$. On the other hand, $w, V \Vdash [y/x]\Box B$ iff $w, V \Vdash \Box[y/x]B$ (by definition 3.3), iff $w', V' \Vdash [y/x]B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$. By induction hypothesis, $w', V'^{[y/x]} \Vdash B$ iff $w', V' \Vdash [y/x]B$. So we have to show that

   (1) $\qquad w', V^{[y/x]\prime} \Vdash B$ for all $w', V^{[y/x]\prime}$ such that wRw' and $V^{[y/x]}_w \triangleright V^{[y/x]\prime}_{w'}$

iff

(2) $\quad w', V'^{[y/x]} \Vdash B$ for all w', V' such that wRw' and $V_w \triangleright V'_{w'}$.

(1) implies (2) because every interpretation $V'^{[y/x]}$ with $V_w \triangleright V'_{w'}$ is also an interpretation $V^{[y/x]\prime}$ with $V^{[y/x]}_w \triangleright V^{[y/x]\prime}_{w'}$. Assume $V_w \triangleright V'_{w'}$; i.e. there is a C such that for all z, $V'_{w'}(z)$ is a C-counterpart of $V_w(z)$. Then $V^{[y/x]}_w \triangleright V'^{[y/x]}_{w'}$ because for $z \neq x$, $V^{[y/x]}_w(z) = V_w(z) C V'_{w'}(z) = V'^{[y/x]}_{w'}(z)$, and $V^{[y/x]}_w(x) = V_w(y) C V'_{w'}(y) = V'^{[y/x]}_{w'}(x)$.

Assume y is not free in $\Box B$, and that (2) holds. In order to derive (1), consider any w'-image $V^{[y/x]\prime}$ of $V^{[y/x]}$ at w. I.e., there is a counterpart relation $C \in K_{w,w'}$ such that for all z, $V^{[y/x]\prime}_{w'}(z)$ is a C-counterpart of $V^{[y/x]}_w(z)$. Let V^* be like $V^{[y/x]\prime}$ except that $V^*_{w'}(y) = V^{[y/x]\prime}_{w'}(x)$. Let V' be like V^* except that $V'_{w'}(x)$ is some C-counterpart of $V_w(x)$, or undefined if there is none. Then $V_w \triangleright V'_{w'}$. (In particular, $V'_{w'}(y) = V^*_{w'}(y) = V^{[y/x]\prime}(x)$ is a C-counterpart of $V^{[y/x]}_w(x) = V_w(y)$, or undefined if there is none.) So by (2), $w', V'^{[y/x]} \Vdash B$. But $V'^{[y/x]} = V^*$ (since $V^*_{w'}(x) = V^*_{w'}(y)$). So $w', V^* \Vdash B$. And since y is not free in B and V^* is a y-variant of $V^{[y/x]\prime}$ on w', by the coincidence lemma 2.10, $w', V^* \Vdash B$ iff $w', V^{[y/x]\prime} \Vdash B$. ■

A stronger version of this will be proved as lemma 3.9 below. We only need this version to verify that renaming bound variables (a.k.a. α-conversion) does not affect truth-values.

Definition 3.7 (Alphabetic variant)
A formula A' of a language of quantified modal logic (with or without substitution) is an *alphabetic variant of* a formula A if one of the following conditions is satisfied.

1. $A = A'$.
2. $A = \neg B$, $A' = \neg B'$, and B' is an alphabetic variant of B.
3. $A = B \supset C$, $A' = B' \supset C'$, and B', C' are alphabetic variants of B, C, respectively.
4. $A = \forall x B$, $A' = \forall z [z/x] B'$, B' is an alphabetic variant of B, and either $z = x$ or $z \notin \mathit{Var}(B')$.
5. $A = \langle y : x \rangle B$, $A' = \langle y : z \rangle [z/x] B'$, B' is an alphabetic variant of B, and either $z = x$ or $z \notin \mathit{Var}(B')$.
6. $A = \Box B$, $A' = \Box B'$, and B' is an alphabetic variant of A'.

We could have weakened the condition that $z \notin \mathit{Var}(B')$ to the condition that z is not free in B'. But this would make it much harder to prove the following lemma. E.g. if we allow $\forall z [z/x] \forall z Fxz = \forall z \forall v Fzv$ as alphabetic variant of $\forall x \forall z Fxz$ (with $B = B' = \forall z Fxz$), then we can't apply to the preliminary restricted substitution lemma in the following proof. We would need to bootstrap ourselves even slower: first prove the alpha-conversion lemma for cases where $z \notin \mathit{Var}(B')$; then use this to prove the full restricted substitution lemma; then use this to prove the full alpha-conversion lemma.

Lemma 3.8 (alpha-conversion lemma)
If a formula A' is an alphabetic variant of a formula A, then for any world w in any structure $\mathcal{S}$ and any interpretation V on $\mathcal{S}$,

$$w, V \Vdash_{\mathcal{S}} A \text{ iff } w, V \Vdash_{\mathcal{S}} A'.$$

Proof by induction on A.

1. A is atomic. Then $A = A'$ and the claim is trivial.
2. A is $\neg B$. Then A' is $\neg B'$, where B' is an alphabetic variant of B. By induction hypothesis, $w, V \Vdash B$ iff $w, V \Vdash B'$. So $w, V \Vdash \neg B$ iff $w, V \Vdash \neg B'$ by definition 2.7.
3. A is $B \supset C$. Then A' is $B' \supset C'$, where B', C' are alphabetic variants of B, C, respectively. By induction hypothesis, $w, V \Vdash B$ iff $w, V \Vdash B'$ and $w, V \Vdash C$ iff $w, V \Vdash C'$. So $w, V \Vdash B \supset C$ iff $w, V \Vdash B' \supset C'$ by definition 2.7.
4. A is $\forall x B$. Then A' is either $\forall x B'$ or $\forall z[z/x]B'$, where B' is an alphabetic variant of B and $z \notin \mathit{Var}(B')$. If B is $\forall x B'$, then by definition 2.7, $w, V \Vdash \forall x B$ iff $w, V' \Vdash B$ for all existential x-variants V' of V on w, which, by induction hypothesis, holds iff $w, V' \Vdash B'$ for all such V', i.e. (by definition 2.7 again) iff $w, V \Vdash \forall x B'$.

   Consider then the case where B is $\forall z[z/x]B'$, with $z \notin \mathit{Var}(B')$. This means that z is not free in B, because alphabetic variants never differ in their free variables. Now if $w, V \nVdash \forall z[z/x]B'$, then by definition 2.7 there is an existential z-variant V^* of V on w such that $w, V^* \nVdash [z/x]B'$. And then $w, (V^*)^{[z/x]} \nVdash B'$ by lemma 3.6. By induction hypothesis, then $w, (V^*)^{[z/x]} \nVdash B$. Let V' be the x-variant of V on w with $V'(x) = (V^*)^{[z/x]}(x) = V^*(z)$. Since z is not free in B, V' and $(V^*)^{[z/x]}$ agree at w about all free variables in B. So by the coincidence lemma 2.10, $w, V' \nVdash B$. And so $w, V \nVdash \forall x B$ by definition 2.7.

   The converse, that if $w, V \Vdash \forall z[z/x]B'$, then $w, V \Vdash \forall x B$, follows from the fact that $\forall x B$ is $\forall x[x/z][z/x]B'$ and thus an alphabetic variant of $\forall z[z/x]B'$.
5. A is $\langle y : x \rangle B$. Then A' is either $\langle y : x \rangle B'$ or $\langle y : z \rangle [z/x]B'$, where B' is an alphabetic variant of B and $z \notin \mathit{Var}(B')$. Assume first that B is $\langle y : x \rangle B'$. By definition 3.2, $w, V \Vdash \langle y : x \rangle B$ iff $w, V' \Vdash B$ where V' is the x-variant of V on w with $V'_w(x) = V_w(y)$. By induction hypothesis, this holds iff $w, V' \Vdash B'$, i.e. (by definition 2.7 again) iff $w, V \Vdash \langle y : x \rangle B'$.

   Consider then the case where B is $\langle y : z \rangle [z/x]B'$, with $z \notin \mathit{Var}(B')$. This means that z is not free in B, because alphabetic variants never differ in their free variables. By definition 2.7, $w, V \Vdash \langle y : x\rangle B$ iff $w, V' \Vdash B$, where V' is the x-variant of V on w with $V'_w(x) = V_w(y)$. Let V^* be the z-variant of V on w with $V^*_w(z) = V'_w(x) = V_w(y)$. Since z is not free in B, V' and $(V^*)^{[z/x]}$ agree at w about all variables in B. So by the coincidence lemma 2.10, $w, V' \Vdash B$ iff $w, (V^*)^{[z/x]} \Vdash B$. By induction hypothesis, $w, (V^*)^{[z/x]} \Vdash B$ iff $w, (V^*)^{[z/x]} \Vdash B'$. By lemma 3.6, $w, (V^*)^{[z/x]} \Vdash B'$ iff $w, V^* \Vdash [z/x]B'$. And by definition 2.7, $w, V^* \Vdash [z/x]B'$ iff $w, V \Vdash \langle y : z\rangle[z/x]B'$.

6. A is $\Box B$. Then A' is $\Box B'$, where B' is an alphabetic variant of B. By definition 2.7, $w, V \Vdash \Box B$ iff $w', V' \Vdash B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$, and $w, V \Vdash \Box B'$ iff $w', V' \Vdash B'$ for all such w', V'. By induction hypothesis, $w', V' \Vdash B$ iff $w', V' \Vdash B'$. So $w, V \Vdash \Box B$ iff $w, V \Vdash \Box B'$ by definition 2.7. ■

Now for the more general version of lemma 3.6.

Lemma 3.9 (Restricted substitution lemma)
Let A be a sentence in a language $\mathcal{L}$ of quantified modal logic (with or without substitution), $\mathcal{S}$ a counterpart structure for $\mathcal{L}$, w a world in $\mathcal{S}$, V an interpretation on $\mathcal{S}$. Then

(i) $w, V^{[y/x]} \Vdash_{\mathcal{S}} A$ iff $w, V \Vdash_{\mathcal{S}} [y/x]A$, provided that either
   (a) y and x are modally separated in A, or
   (b) there is no world $w' \in W$ and counterpart relation $C \in K_{w,w'}$ such that $V_w(y)$ has multiple C-counterparts at w'.

(ii) if $w, V^{[y/x]} \Vdash_{\mathcal{S}} A$, then $w, V \Vdash_{\mathcal{S}} [y/x]A$, provided that y is modally free for x in A.

(Proviso (i).(b) is meant to include cases where $V_w(y)$ is undefined.)

The proviso in (i).(b) could obviously be weakened to "it is not the case that $V_w(y)$ has multiple counterparts at any world *reachable* from w", where reachability is the transitive closure of accessibility. Only considering worlds accessible from w is not enough: if y has only one counterpart at all w'. but two counterparts at some w'', then $\langle y : x\rangle\Diamond\Diamond x \neq y$ is true, but $[y/x]\Diamond\Diamond x \neq y = \Diamond\Diamond y \neq y$ is false.

Proof If y and x are the same variable, then $V^{[y/x]}$ is V, and $[y/x]A$ is A; so trivially $w, V^{[y/x]} \Vdash A$ iff $w, V \Vdash [y/x]A$. Assume then that y and x are different variables. The proof is by induction on A.

For the base case, the provisos can be ignored: $w, V^{[y/x]} \Vdash Px_1 \dots x_n$ iff $w, V \Vdash [y/x]Px_1 \dots x_n$ by the same reasoning as in lemma 3.6. For complex A, the induction hypothesis is that (i) and (ii) hold for formulas of lower complexity, in particular for subformulas of A. Note that if one of the provisos of (i) and (ii) applies to A, then it also applies to subformulas for A. (For example, if y is modally free for x in A, then y is modally free for x in every subformula of A.) Moreover, if A is not of the form $\Box B$, then the proviso of (ii) entails proviso (a) of (i), because y is modally free for x in $A \neq \Box B$ only if y and x do not occur together in the scope of a modal operator in A.

1. A is $\neg B$. By definition 2.7, $w, V^{[y/x]} \Vdash \neg B$ iff $w, V^{[y/x]} \nVdash B$. Since a proviso of (i) or (ii) applies to A and therefore a proviso of (i) applies to B, by induction hypothesis, $w, V^{[y/x]} \nVdash B$ iff $w, V \nVdash [y/x]B$. And the latter holds iff $w, V \Vdash [y/x]\neg B$ by definitions 2.7 and 3.3.

2. A is $B \supset C$. By definition 2.7, $w, V^{[y/x]} \Vdash B \supset C$ iff $w, V^{[y/x]} \nVdash B$ or $w, V^{[y/x]} \Vdash C$. Since a proviso of (i) or (ii) applies to A and therefore a proviso of (i) applies to B and C, by induction hypothesis, $w, V^{[y/x]} \nVdash B$ iff $w, V \nVdash [y/x]B$, and $w, V^{[y/x]} \Vdash C$ iff $w, V \Vdash [y/x]C$. So $w, V^{[y/x]} \Vdash B \supset C$ iff $w, V \Vdash [y/x](B \supset C)$ by definitions 2.7 and 3.3.

3. A is $\forall zB$. Assume first that $[y/x]\forall zB$ is $\forall[y/x]z[y/x]B$, i.e. (by definition 3.3) neither $z = y$ and $x \in \mathit{Varf}(B)$ nor $z = x$ and $y \in \mathit{Varf}(B)$. By definition 2.7, $w, V \Vdash \forall[y/x]z[y/x]B$ iff $w, V' \Vdash [y/x]B$ for all existential $[y/x]z$-variants V' of V on w. Since a proviso of (i) or (ii) applies to A and therefore a proviso of (i) applies to B, by induction hypothesis, $w, V' \Vdash [y/x]B$ iff $w, V'^{[y/x]} \Vdash B$.

   Now assume $z \notin \{x, y\}$. Then $V'^{[y/x]}_w(x) = V'^{[y/x]}_w(y) = V'_w(y) = V_w(y)$ and $V'^{[y/x]}_w(z) = V'^{[y/x]}_w([y/x]z)$ is some arbitrary member of D_w. So the interpretations $V'^{[y/x]}$ coincide with the existential z-variants $V^{[y/x]\prime}$ of $V^{[y/x]}$ on w. Alternatively, if $z = x$, and thus $y \notin \mathit{Varf}(B)$, then $V'^{[y/x]}_w(x)$ is some arbitrary member of D_w, as is $V^{[y/x]\prime}_w(x)$. $V'^{[y/x]}_w(y) = V'^{[y/x]}_w(x)$ may not equal $V^{[y/x]\prime}_w(y) = V_w(y)$, but $y \notin \mathit{Varf}(B)$. Similarly, if $z = y$ and thus $x \notin fvar(B)$, then $V'^{[y/x]}_w(y)$ is some arbitrary member of D_w, as is $V^{[y/x]\prime}_w(y)$. In either case, the interpretations $V'^{[y/x]}_w$ can be paired with the interpretations $V^{[y/x]\prime}$ such that the members of each pair agree at w about all free variables in B. So by the coincidence lemma 2.10, $w, V'^{[y/x]} \Vdash B$ for all existential $[y/x]z$-variants V' of V on w iff $w, V^{[y/x]\prime} \Vdash B$ for all existential z-variants $V^{[y/x]\prime}$ of $V^{[y/x]}$ on w, iff $w, V^{[y/x]} \Vdash \forall zB$ by definition 2.7. (The reasoning here is not very perspicuous.)

   Second, assume $[y/x]\forall zB$ is $\forall v[y/x][v/z]B$, for some new variable v. By the α-conversion lemma 3.8, $w, V^{[y/x]} \Vdash \forall zB$ iff $w, V^{[y/x]} \Vdash \forall v[v/z]B$. Since $v \notin \{x, y\}$, we can reason as above, with $[v/z]B$ in place of B, to show that $w, V \Vdash \forall v[y/x][v/z]B$ iff $w, V^{[y/x]} \Vdash \forall v[v/z]B$.

4. A is $\langle y_2 : z\rangle B$. This case is similar to the previous one. Assume first that $[y/x]\langle y_2 : z\rangle B$ is $\langle [y/x]y_2 : [y/x]z\rangle[y/x]B$, i.e. (by definition 3.3) neither $z = y$ and $x \in \mathit{Varf}(B)$ nor $z = x$ and $y \in \mathit{Varf}(B)$. By definition 2.7, $w, V \Vdash \langle [y/x]y_2 : [y/x]z\rangle[y/x]B$ iff $w, V' \Vdash [y/x]B$, where V' is the $[y/x]z$-variant of V on w with $V'_w([y/x]z) = V_w([y/x]y_2)$. Since a proviso of (i) or (ii) applies to A and therefore a proviso of (i) applies to B, by induction hypothesis, $w, V' \Vdash [y/x]B$ iff $w, V'^{[y/x]} \Vdash B$.

   Let V^* be the z-variant of $V^{[y/x]}$ on w with $V^*_w(z) = V_w([y/x]y_2)$. If $z \notin \{x, y\}$, then $V^*_w(x) = V^*_w(y) = V_w(y)$ and $V^*(z) = V_w([y/x]y_2)$. Moreover, $V'^{[y/x]}_w(x) = V'^{[y/x]}_w(y) = V_w(y)$ and $V'^{[y/x]}_w(z) = V'^{[y/x]}_w([y/x]z) = V_w([y/x]y_2)$. So $V'^{[y/x]}$ and V^* agree about all variables at w. Alternatively, if $z = x$, and thus $y \notin \mathit{Varf}(B)$, then $V'^{[y/x]}_w(x) = V_w([y/x]y_2) = V^*_w(x)$. Similarly, if $z = y$, and thus $x \notin \mathit{Varf}(B)$, then $V'^{[y/x]}_w(y) = V_w([y/x]y_2) = V^*_w(y)$. Either way, $V'^{[y/x]}$ and V^* agree at w about all free variables in B. By the coincidence lemma 2.10, $w, V'^{[y/x]} \Vdash B$ iff $w, V^* \Vdash B$, iff $w, V^{[y/x]} \Vdash \langle [y/x]y_2 : z\rangle B$ by definition 2.7.

5. A is $\Box B$. This is the interesting part. We have to go piecemeal.

(i). By definition 2.7, $w, V^{[y/x]} \Vdash \Box B$ iff $w', V^{[y/x]\prime} \Vdash B$ for all $w', V^{[y/x]\prime}$ with wRw' and $V_w^{[y/x]} \triangleright V_{w'}^{[y/x]\prime}$. On the other hand, $w, V \Vdash [y/x]\Box B$ iff $w, V \Vdash \Box[y/x]B$ (by definition 3.3), iff $w', V' \Vdash [y/x]B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$. Since the provisos of (i) carry over from $\Box B$ to B, by induction hypothesis, $w', V'^{[y/x]} \Vdash B$ iff $w', V' \Vdash [y/x]B$. So we have to show that

(1) $\qquad w', V^{[y/x]\prime} \Vdash B$ for all $w', V^{[y/x]\prime}$ such that wRw' and $V_w^{[y/x]} \triangleright V_{w'}^{[y/x]\prime}$

iff

(2) $\qquad w', V'^{[y/x]} \Vdash B$ for all w', V' such that wRw' and $V_w \triangleright V'_{w'}$.

(1) implies (2) because every interpretation $V'^{[y/x]}$ with $V_w \triangleright V'_{w'}$ is also an interpretation $V^{[y/x]\prime}$ with $V_w^{[y/x]} \triangleright V_{w'}^{[y/x]\prime}$. The converse, however, may fail: both $V'^{[y/x]}_{w'}$ and $V_{w'}^{[y/x]\prime}$ assign to x and y some counterpart of $V_w(y)$ (if there is any). But while $V'^{[y/x]}_{w'}$ assigns *the same* counterpart to x and y, $V_{w'}^{[y/x]\prime}$ may choose different counterparts for x and y relative to the same counterpart relation.

If there is no counterpart relation relative to which $V_w(y)$ has multiple counterparts, then this cannot happen. Thus under proviso (b), each $V^{[y/x]\prime}$ with $V_w^{[y/x]} \triangleright V_{w'}^{[y/x]\prime}$ is also a $V'^{[y/x]}$ with $V'_w \triangleright V'^{[y/x]}_{w'}$, and so (2) implies (1).

For proviso (a), assume x and y do not both occur in the scope of a modal operator in $\Box B$. Then either x or y does not occur at all in $\Box B$. Assume first that x does not occur in $\Box B$. Then $[y/x]\Box B$ is $\Box B$ (by definition 3.3), and $w, V^{[y/x]} \Vdash \Box B$ iff $w, V \Vdash [y/x]\Box B$ by the coincidence lemma 2.10. Alternatively, assume that y does not occur in $\Box B$, and that (2) holds. In order to derive (1), consider any w'-image $V^{[y/x]\prime}$ of $V^{[y/x]}$ at w; i.e. for some counterpart relation $C \in K_{w,w'}$ and all variables z, $V_{w'}^{[y/x]\prime}(z)$ is a C-counterpart of $V_w^{[y/x]}$ (or undefined if there is none). Let V^* be like $V^{[y/x]\prime}$ except that $V^*_{w'}(y) = V_{w'}^{[y/x]\prime}(x)$. Let V' be like V^* except that $V'_{w'}(x)$ is some C-counterpart of $V_w(x)$, or undefined if there is none. Then $V_w \triangleright V'_{w'}$. (In particular, $V'_{w'}(y) = V^*_{w'}(y) = V^{[y/x]\prime}(x)$ is some C-counterpart of $V_w^{[y/x]}(x) = V_w(y)$, or undefined if there is none.) So by (2), $w', V'^{[y/x]} \Vdash B$. But $V'^{[y/x]} = V^*$ (since $V^*_{w'}(x) = V^*_{w'}(y)$). So $w', V^* \Vdash B$. And since $y \notin \mathit{Var}(B)$ and V^* is a y-variant of $V^{[y/x]\prime}$ on w', by the coincidence lemma 2.10, $w', V^* \Vdash B$ iff $w', V^{[y/x]\prime} \Vdash B$.

(ii). Assume $w, V^{[y/x]} \Vdash \Box B$. By definition 2.7, then $w', V^{[y/x]\prime} \Vdash B$ for all $w', V^{[y/x]\prime}$ with wRw' and $V_w^{[y/x]} \triangleright V_{w'}^{[y/x]\prime}$. As before, every interpretation $V'^{[y/x]}$ with $V_w \triangleright V'_{w'}$ is also an interpretation $V^{[y/x]\prime}$ with $V_w^{[y/x]} \triangleright V_{w'}^{[y/x]\prime}$. So $w', V'^{[y/x]} \Vdash B$ for all $w', V'^{[y/x]}$ with wRw' and $V_w \triangleright V'_{w'}$.

If y is modally free for x in $\Box B$, then y is modally free for x in B. Then by induction hypothesis, $w', V' \Vdash [y/x]B$ *if* $w', V'^{[y/x]} \Vdash B$. So $w', V' \Vdash [y/x]B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$. By definition 2.7, this means that $w, V \Vdash \Box[y/x]B$, and so $w, V \Vdash [y/x]\Box B$ by definition 3.3. ■

The converse of (ii) is not true. E.g., $w, V \Vdash [y/x]\Box x{=}y$ does not imply $w, V^{[y/x]} \Vdash \Box x{=}y$. So the operation $[y/x]$, as defined in definition 3.3, does not always satisfy the "substitution lemma", not even when y is modally free for x.

Can we fix the definition? No – at least not if we allow for positive models. There is no operation Φ on sentences in standard languages of quantified modal logic such that in any (positive) model, $w, V^{[y/x]} \Vdash A$ iff $w, V \Vdash \Phi(A)$, and therefore no translation of $\langle y : x\rangle A$ into those languages. To prove this, we show that there are distinctions one can draw with $\langle y : x\rangle$ that cannot be drawn without it. In particular, the substitution quantifier allows us to say that an individual y has multiple counterparts at some accessible world (under the same counterpart relation): $\langle y : x\rangle\Diamond y{\neq}x$.

(In negative models, $\langle y : x\rangle A$ can be translated into $\exists x(x{=}y \wedge A) \vee (\neg Ey \wedge [y/x]A)$, which still has the downside of being very impractical, since $\Phi(A)$ can have much greater syntactic complexity than A.)

It is clear that $\Diamond y \neq y$ is not an adequate translation of $\langle y : x\rangle\Diamond x \neq y$. Before substituting y for x in $\Diamond x \neq y$, we would have to make x free for y by renaming the modally bound occurrence of y. However, the diamond, unlike the quantifier $\forall y$, binds y in such a way that the domain over which it ranges (the counterparts of y's original referent) depends on the previous reference of y. So we can't just replace y by some other variable z, translating $\langle y : x\rangle\Diamond x{\neq}y$ as $\Diamond y{\neq}z$. This only works if z happens to corefer with y. Since we can't presuppose that there is always another name available for any given individual, we would somehow have to introduce a name z that corefers with y. For instance, we could translate $\Diamond x{\neq}y$ into $\exists z(y{=}z \wedge \Diamond x{\neq}z)$. Now x has become free for y in the scope of the diamond, so we can translate $\langle y : x\rangle\Diamond x{\neq}y$ as $\exists z(y{=}z \wedge \Diamond x{\neq}y)$. The problem with this is that the quantifier $\exists$ ranges only over existing objects, while $\langle y : x\rangle$ bears no such restriction. In positive models, $V_w(y)$ can have multiple counterparts even if it lies outside D_w, so that $\exists z(y{=}z \wedge \Diamond x{\neq}y)$ is false. It wouldn't help to use $\forall z(y{=}z \supset A)$ instead, because that would translate *any* statement $\langle y : x\rangle A$ into something true in this scenario, even $\langle y : x\rangle(P \wedge \neg P)$. (One would need an "outer quantifier" in place of $\exists$.)

Here is the full proof.

Theorem 3.10 (Undefinability of substitution)
There is no operation Φ on formulas A in a standard language $\mathcal{L}$ of quantified modal logic such that for all worlds w in all positive counterpart models $\langle \mathcal{S}, V\rangle$, $w, V \Vdash_{\mathcal{S}} \Phi(A)$ iff $w, V^{[y/x]} \Vdash_{\mathcal{S}} A$.

Proof Let $\mathcal{M}_1 = \langle \mathcal{S}_1, V\rangle$ be a positive counterpart model with $W = \{w\}$, $R = \{\langle w, w\rangle\}$, $U_w = \{x, y, y^*\}$, $D_w = \{x\}$, $K_{w,w} = \{\{\langle d, d\rangle : d \in U_w\}\}, V_w(y) = y$, $V_w(z) = x$ for every variable $z \neq y$, and $V_w(P) = \emptyset$ for all non-logical predicates P. Let $\mathcal{M}_2 = \langle \mathcal{S}_2, V\rangle$ be like $\mathcal{M}_1$

except that y^* is also a counterpart of y, i.e. $K_{w,w'} = \{\{\langle x,x\rangle, \langle y,y\rangle, \langle y^*,y^*\rangle, \langle y,y^*\rangle\}\}$. Then $w, V^{[y/x]} \Vdash_{\mathcal{S}_2} \Diamond y \neq x$, but $w, V^{[y/x]} \nVdash_{\mathcal{S}_1} \Diamond y \neq x$.

On the other hand, every $\mathcal{L}$-sentence has the same truth-value at w under V in both models. We prove this by showing that for every $\mathcal{L}$-sentence A, $w, V \Vdash_{\mathcal{S}_1} A$ iff $w, V \Vdash_{\mathcal{S}_2} A$ iff $w, V^* \Vdash_{\mathcal{S}_2} A$, where V^* is the y-variant of V on w with $V^*_w(y) = V_w(y^*)$.

1. A is $Px_1 \ldots x_n$. It is clear that $w, V \Vdash_{\mathcal{S}_1} Px_1 \ldots x_n$ iff $w, V \Vdash_{\mathcal{S}_2} Px_1 \ldots x_n$ because counterpart relations do not figure in the evaluation of atomic formulas. Moreover, for non-logical P, $w, V \nVdash_{\mathcal{S}_2} Px_1 \ldots x_n$ and $w, V^* \nVdash_{\mathcal{S}_2} Px_1 \ldots x_n$, because $V_w(P) = V^*_w(P) = \emptyset$. For the identity predicate, observe that $w, V \nVdash_{\mathcal{S}_2} u = v$ iff exactly one of u, v is y, since $V_w(z) = x$ for all terms $z \neq y$. For the same reason, $w, V^* \nVdash_{\mathcal{S}_2} u = v$ iff exactly one of u, v is y. So $w, V \Vdash_{\mathcal{S}_2} u = v$ iff $w, V^* \Vdash_{\mathcal{S}_2} u = v$.

2. A is $\neg B$. $w, V \Vdash_{\mathcal{S}_1} \neg B$ iff $w, V \nVdash_{\mathcal{S}_1} B$ by definition 2.7, iff $w, V \nVdash_{\mathcal{S}_2} B$ by induction hypothesis, iff $w, V \Vdash_{\mathcal{S}_2} \neg B$ by definition 2.7. Moreover, $w, V \nVdash_{\mathcal{S}_2} B$ iff $w, V^* \nVdash_{\mathcal{S}_2} B$ by induction hypothesis, iff $w, V^* \Vdash_{\mathcal{S}_2} \neg B$ by definition 2.7.

3. A is $B \supset C$. $w, V \Vdash_{\mathcal{S}_1} B \supset C$ iff $w, V \nVdash_{\mathcal{S}_1} B$ or $w, V \Vdash_{\mathcal{S}_1} C$ by definition 2.7, iff $w, V \nVdash_{\mathcal{S}_2} B$ or $w, V \Vdash_{\mathcal{S}_2} C$ by induction hypothesis, iff $w, V \Vdash_{\mathcal{S}_2} B \supset C$ by definition 2.7. Moreover, $w, V \nVdash_{\mathcal{S}_2} B$ or $w, V \Vdash_{\mathcal{S}_2} C$, iff $w, V^* \nVdash_{\mathcal{S}_2} B$ or $w, V^* \Vdash_{\mathcal{S}_2} C$ by induction hypothesis, iff $w, V^* \Vdash_{\mathcal{S}_2} B \supset C$ by definition 2.7.

4. A is $\forall z B$. Let v be a variable not in $Var(B) \cup \{y\}$. By lemma 3.8, $w, V \Vdash_{\mathcal{S}_1} \forall z B$ iff $w, V \Vdash_{\mathcal{S}_1} \forall v [v/z] B$. By definition 2.7, $w, V \Vdash_{\mathcal{S}_1} \forall v [v/z] B$ iff $w, V' \Vdash_{\mathcal{S}_1} [v/z] B$ for all existential v-variants V' of V on w. As $D_w = \{x\}$ and $V(v) = x$, the only such v-variant is V itself. So $w, V \Vdash_{\mathcal{S}_1} \forall z B$ iff $w, V \Vdash_{\mathcal{S}_1} [v/z] B$. (Without the detour through v, the bound variable z could have been y, in which case V itself would not be the relevant z-variant V'.) By the same reasoning, $w, V \Vdash_{\mathcal{S}_2} \forall z B$ iff $w, V \Vdash_{\mathcal{S}_2} [v/z] B$. But by induction hypothesis, $w, V \Vdash_{\mathcal{S}_1} [v/z] B$ iff $w, V \Vdash_{\mathcal{S}_2} [v/z] B$. So $w, V \Vdash_{\mathcal{S}_1} \forall z B$ iff $w, V \Vdash_{\mathcal{S}_2} \forall z B$. Moreover, by induction hypothesis, $w, V \Vdash_{\mathcal{S}_2} [v/z] B$ iff $w, V^* \Vdash_{\mathcal{S}_2} [v/z] B$, iff $w, V^* \Vdash_{\mathcal{S}_2} \forall v [v/z] B$ because V^* is the only existential v-variant of V^* on w, iff $w, V^* \Vdash_{\mathcal{S}_2} \forall z B$ by lemma 3.8.

5. A is $\Box B$. In both structures, the only world accessible from w is w itself. Also in $\mathcal{S}_1$, V is the only w-image of V at w. So by definition 2.7, $w, V \Vdash_{\mathcal{S}_1} \Box B$ iff $w, V \Vdash_{\mathcal{S}_1} B$. In $\mathcal{S}_2$, there are two w-images of V at w: V and V^*. So $w, V \Vdash_{\mathcal{S}_2} \Box B$ iff both $w, V \Vdash_{\mathcal{S}_2} B$ and $w, V^* \Vdash_{\mathcal{S}_2} B$. By induction hypothesis, $w, V \Vdash_{\mathcal{S}_1} B$ iff both $w, V \Vdash_{\mathcal{S}_2} B$ and $w, V^* \Vdash_{\mathcal{S}_2} B$. So $w, V \Vdash_{\mathcal{S}_1} \Box B$ iff $w, V \Vdash_{\mathcal{S}_2} \Box B$. Moreover, in $\mathcal{S}_2$, V^* is the only w-image of V^* at w. So $w, V^* \Vdash_{\mathcal{S}_2} \Box B$ iff $w, V^* \Vdash_{\mathcal{S}_2} B$. By induction hypothesis, $w, V^* \Vdash_{\mathcal{S}_2} B$ iff $w, V \Vdash_{\mathcal{S}_2} B$. So $w, V^* \Vdash_{\mathcal{S}_2} \Box B$ iff both $w, V^* \Vdash_{\mathcal{S}_2} B$ and $w, V \Vdash_{\mathcal{S}_2} B$, which as we just saw holds iff $w, V \Vdash_{\mathcal{S}_2} \Box B$. ■

The proof of lemma 3.10 assumes positive semantics ($U \neq D$). In negative semantics, $V_w(y)$ can be undefined, but then it remains undefined under any image of V at w; so there can't be a non-existing individual with multiple counterparts. Moreover, if $V_w(y)$ does not have multiple counterparts at any world, then by lemma 3.9 we can simply define $\langle y : x \rangle A$ as $[y/x]A$. So the only case to take care of is that of an individual in D_w having multiple counterparts. But then we can say $\exists x(x = y \wedge \Diamond x \neq y)$ for $\langle y : x\rangle \Diamond x \neq y$.

To provide for all cases simultaneously, we might translate $\langle y : x\rangle A$ into $(Ey \wedge \exists x(x = y \wedge A)) \vee (\neg Ey \wedge [y/x]A) = \exists x(x = y \wedge A) \vee (\neg Ey \wedge [y/x]A)$.

$$
\begin{array}{ll}
& w, V \Vdash \exists x(x = y \wedge A) \vee (\neg Ey \wedge [y/x]A) \\
\text{iff} & w, V \Vdash \exists x(x = y \wedge A) \text{ or } w, V \Vdash \neg Ey \wedge [y/x]A \\
\text{iff} & w, V' \Vdash x = y \wedge A \text{ for some } V' \text{ with } V'_w(x) \in D_w \\
& \text{or } V_w(y) \notin D_w \text{ and } w, V \Vdash [y/x]A \\
\text{iff} & w, V' \Vdash A \text{ for some } V' \text{ with } V'_w(x) = V_w(y) \in D_w \\
& \text{or } V_w(y) \notin D_w \text{ and } w, V \Vdash [y/x]A \\
\text{iff} & w, V^{[y/x]} \Vdash A \text{ and } V_w(y) \in D_w \\
& \text{or } V_w(y) \notin D_w \text{ and } w, V \Vdash [y/x]A \\
\text{iff} & w, V^{[y/x]} \Vdash A \text{ and } V_w(y) \in D_w \\
& \text{or } V_w(y) \notin D_w \text{ and } w, V^{[y/x]} \Vdash A \qquad \text{(by lemma 3.9)} \\
\text{iff} & w, V^{[y/x]} \Vdash A
\end{array}
$$

Lemma 3.9 applies at the penultimate step because $V_w(y) \notin D_w$ entails in negative models that it is not the case that $V_w(y)$ has multiple counterparts at any world.

This definition of substitution is negative logics has the downside that $\langle y : x\rangle A$ is of significantly higher syntactic complexity than A. Nevertheless, we could use it in principles like (LL*) and (FUI*). I don't know whether the logics would thereby be complete: my proof of the canonical model lemma (lemma **??**) requires that $\langle y : x\rangle A$ is of lower complexity than $\forall x A$.

Among the less obvious ideas for defining $\langle y : x\rangle$ consider recursive definitions like

$$
\begin{aligned}
\langle y : x\rangle Fx &= \exists x(y = x \wedge Fx) \\
\langle y : x\rangle \neg A &= \neg \langle y : x\rangle A \\
\langle y : x\rangle (A \supset B) &= \langle y : x\rangle A \supset \langle y : x\rangle B \\
\langle y : x\rangle \forall z A &= \forall z' \langle y : x\rangle [z'/z]A \\
\langle y : x\rangle \Box A &= \exists x(y = x \wedge \Box A).
\end{aligned}
$$

Note that this sneakily breaks the link from Continuity to NE. Cont becomes

$$\exists x(x = y \wedge \Box A) \supset \Box \langle y : x\rangle A.$$

What this further amounts to depends on A. E.g. for $A = \neg\bot$, the RHS becomes $\Box\neg\exists x(x = y \wedge \bot)$, which is just right. There is, however, still the problem that $\langle y : x\rangle \Box\neg Fx$ entails Ey, which it doesn't on the primitive reading.

One might also always drive out the substitution to the front, so that

$$\Box\langle y : x\rangle A = \exists x(y = x \wedge \Box A).$$

But then inverse Continuity would also be valid. And how do we handle $\Diamond\exists y \langle y : x\rangle \Diamond Gxy$? Could we drive substitution out to wherever the variables(s??) were introduced? FWIW, we could also let $\langle y : x\rangle A$ be $[y/x]A$ whenever y does not occur modalised in A.

What we can do instead is introduce a new syntactic construction into the language that satisfies the substitution lemma by stipulation. This is what the substitution quantifier does. I have given its semantics in definition 3.2 by saying that $w, V \Vdash \langle y : x\rangle A$ iff $w, V' \Vdash A$, where V' is the x-variant of V on w with $V'_w(x) = V_w(y)$. By the locality lemma (corollary 2.11 of lemma 2.10), it immediately follows that

$$w, V \Vdash \langle y : x\rangle A \text{ iff } w, V^{[y/x]} \Vdash A.$$

In the following, I will consider both systems in extended languages that include the substitution quantifier $\langle y : x\rangle$ and systems in standard languages that exclude it. The advantage of having the substitution quantifier is that it not only adds welcome expressive resources to the language, but also makes the logic and model theory somewhat more streamlined, because the corresponding versions of principles like Leibniz's Law,

$$x = y \supset (A \supset \langle y : x\rangle A)$$

hold without restrictions.

It will be useful to have a notion of substitution that applies to several variables at once. To this end, let's generalise definition 3.3 (classical substitution).

> Definition 3.11 (Classical substitution, generalised)
> A *substitution* on a language $\mathcal{L}$ is a total function $\sigma : \mathit{Var}(\mathcal{L}) \to \mathit{Var}(\mathcal{L})$. If σ is injective, it is called a *transformation*. I write $[y_1, \ldots, y_n/x_1, \ldots, x_n]$ for the substitution that maps x_1 to y_1, ..., x_n to y_n, and every other variable to itself. (So $[y_1, \ldots, y_n/x_1, \ldots, x_n]$ is only meaningful (and non-redundant) if the $x_1, \ldots, x_n$ are pairwise distinct.)
>
> Application of a substitution σ to a formula A is defined as follows.
>
> $$\sigma(Px_1 \ldots x_n) = P\sigma(x_1)\ldots\sigma(x_n)$$
> $$\sigma(\neg A) = \neg\sigma(A);$$
> $$\sigma(A \supset B) = \sigma(A) \supset \sigma(B);$$
> $$\sigma(\forall zA) = \begin{cases} \forall v\sigma'([v/z]A) & \text{if there is an } x \text{ free in } \forall zA \text{ with } \sigma(x) = \sigma(z), \\ \forall\sigma(z)\sigma(A) & \text{otherwise,} \end{cases}$$
> where σ' is like σ except that $\sigma'(v) = v$, and v is the alphabetically first variable not in $\sigma(A)$;
>
> $$\sigma(\langle y_2 : z\rangle A) = \begin{cases} \langle\sigma'(y_2) : v\rangle\sigma([v/z]A) & \text{if there is an } x \neq v \text{ in } \mathit{Varf}(A) \text{ with } \sigma(x) = \sigma(z), \\ \langle\sigma(y_2) : \sigma(z)\rangle\sigma(A) & \text{otherwise,} \end{cases}$$
> where σ' is like σ except that $\sigma'(v) = v$, and v is the alphabetically first variable not in $\sigma(A)$;

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

I will also write σA or A^σ instead of $\sigma(A)$. If Γ is a set of formulas, I write $\sigma(\Gamma)$ or Γ^σ for $\{C^\tau : C \in \Gamma\}$.

Here is the corresponding generalisation of $V^{[y/x]}$.

Definition 3.12 (Interpretation under substitution, generalised)
For any interpretation V on a structure $\mathcal{S}$ and substitution σ, V^σ is the interpretation that is like V except that for any world w in $\mathcal{S}$ and variable x, $V^\sigma_w(x) = V_w(\sigma(x))$.

Substitutions can be composed. If σ and τ are substitutions, then $\tau\cdot\sigma$ is the substitution that maps each variable x to $\tau(\sigma(x))$. Observe that composition behaves differently in superscripts of formulas than in superscripts of interpretations: for formulas A,

$$(A^\sigma)^\tau = \tau(\sigma(A)) = A^{\tau\cdot\sigma},$$

but for interpretations V,

$$(V^\sigma)^\tau = V^{\sigma\cdot\tau}.$$

That's because $(V^\sigma)^\tau_w(x) = V^\sigma_w(\tau(x)) = V_w(\sigma(\tau(x)) = V_w(\sigma\cdot\tau(x)) = V^{\sigma\cdot\tau}_w(x)$.

Definition 3.11 draws attention to the class of injective substitutions, or transformations. A transformation never substitutes two distinct variables by the same variable. For instance, the identity substitution $[x/x]$ or the swapping operation $[x,y/y,x]$ are transformations. What's special about such substitutions is that they make capturing impossible: for the free variable y in $\forall xA(y)$ to be captured by the initial quantifier $\forall x$ after substitution, x and y have to be replaced by the same variable. Correspondingly, definition 3.11 entails that if σ is a transformation, then $\sigma(A)$ is simply A with all variables simultaneously replaced by their σ-value. Transformations satisfy the substitution lemma without any restrictions, even for modal formulas.

Lemma 3.13 (Transformation Lemma)
For any world w in any structure $\mathcal{S}$, any interpretation V on $\mathcal{S}$, any formula A and transformation τ, $w, V^\tau \Vdash A$ iff $w, V \Vdash A^\tau$.

Proof by induction on A.

1. $A = Px_1 \dots x_n$. $w, V^\tau \Vdash Px_1 \dots x_n$ iff $\langle V^\tau_w(x_1), \dots, V^\tau_w(x_n)\rangle \in V^\tau_w(P)$, iff $\langle V_w(x^\tau_1), \dots, V_w(x^\tau_n)\rangle \in V_w(P)$, iff $w, V \Vdash (Px_1 \dots x_n)^\tau$.
2. $A = \neg B$. $w, V^\tau \Vdash \neg B$ iff $w, V^\tau \nVdash B$, iff $w, V \nVdash B^\tau$ by induction hypothesis, iff $w, V \Vdash (\neg B)^\tau$.
3. $A = B \supset C$. $w, V^\tau \Vdash B \supset C$ iff $w, V^\tau \nVdash B$ or $w, V^\tau \Vdash C$, iff $w, V \nVdash B^\tau$ or $w, V \Vdash C^\tau$ by induction hypothesis, iff $w, V \Vdash (B \supset C)^\tau$.
4. $A = \langle y : x\rangle B$. By definition 3.2, $w, V^\tau \Vdash \langle y : x\rangle B$ iff $w, (V^\tau)^{[y/x]} \Vdash B$. Now $(V^\tau)^{[y/x]}_w(x) = V^\tau_w(y) = V_w(y^\tau) = V_w^{[y^\tau/x^\tau]}(x^\tau) = (V^{[y^\tau/x^\tau]})^\tau_w(x)$. And for any variable $z \neq x$, $(V^\tau)^{[y/x]}_w(z) = V^\tau_w(z) = V_w(z^\tau) = V_w^{[y^\tau/x^\tau]}(z^\tau)$ (because $z^\tau \neq x^\tau$, by injectivity of τ) $= (V^{[y^\tau/x^\tau]})^\tau_w(z)$. So $(V^\tau)^{[y/x]}$ coincides with $(V^{[y^\tau/x^\tau]})^\tau$ at w. By the locality lemma 2.11, $w, (V^\tau)^{[y/x]} \Vdash B$ iff $w, (V^{[y^\tau/x^\tau]})^\tau \Vdash B$. By induction hypothesis, the latter holds iff $w, V^{[y^\tau/x^\tau]} \Vdash B^\tau$, iff $w, V \Vdash \langle y^\tau : x^\tau\rangle B^\tau$ by definition 3.2, iff $w, V \Vdash (\langle y : x\rangle B)^\tau$ by definition 3.11.
5. $A = \forall x B$. Assume $w, V^\tau \nVdash \forall x B$. Then $w, V^* \nVdash B$ for some existential x-variant V^* of V^τ on w. Let V' be the (existential) x^τ-variant of V on w with $V'_w(x^\tau) = V^*_w(x)$. Then ${V'}^\tau_w(x) = V^*_w(x)$, and for any variable $z \neq x$, ${V'}^\tau_w(z) = V'_w(z^\tau) = V_w(z^\tau)$ (because $z^\tau \neq x^\tau$, by injectivity of τ) $= V^\tau_w(z) = V^*_w(z)$. So ${V'}^\tau$ coincides with V^* on w, and by locality (lemma 2.11), $w, {V'}^\tau \nVdash B$. By induction hypothesis, then $w, V' \nVdash B^\tau$. So there is an existential x^τ-variant V' of V on w such that $w, V' \nVdash B^\tau$. By definition 2.7, this means that $w, V \nVdash \forall x^\tau B^\tau$, and hence $w, V \nVdash (\forall x B)^\tau$ by definition 3.11.

   In the other direction, assume $w, V \nVdash (\forall x B)^\tau$, and thus $w, V \nVdash \forall x^\tau B^\tau$. Then $w, V' \nVdash B^\tau$ for some existential x^τ-variant V' of V on w, and by induction hypothesis $w, {V'}^\tau \nVdash B$. Let V^* be the (existential) x-variant of V^τ on w with $V^*_w(x) = V'_w(x^\tau)$. Then $V^*_w(x) = {V'}^\tau_w(x)$, and for any variable $z \neq x$, $V^*_w(z) = V^\tau_w(z) = V_w(z^\tau) = V'_w(z^\tau)$ (because $z^\tau \neq x^\tau$, by injectivity of τ) $= {V'}^\tau_w(z)$. So V^* coincides with ${V'}^\tau$ on w, and by locality (lemma 2.11), $w, V^* \nVdash B$. So there is an existential x-variant V^* of V^τ on w such that $w, V^* \nVdash B$. By definition 2.7, this means that $w, V^\tau \nVdash \forall x B$.
6. $A = \Box B$. Assume $w, V \nVdash \Box B^\tau$. Then $w', V' \nVdash B^\tau$ for some w', V' with wRw' and V' a w' image of V at w. This means that there is a counterpart relation $C \in K_{w,w'}$ such that for all variables x, $V'_{w'}(x)$ is some C-counterpart at w' of $V_w(x)$ at w (if any, else undefined). By induction hypothesis, $w', {V'}^\tau \nVdash B$. Since for all x, ${V'}^\tau_{w'}(x) = V'_{w'}(x^\tau)$ and $V^\tau_w(x) = V_w(x^\tau)$, it follows that ${V'}^\tau_{w'}(x)$ is a C-counterpart of $V^\tau_w(x)$ (if any, else undefined). So ${V'}^\tau$ is a w'-image of V^τ at w. Hence $w', {V'}^\tau \nVdash B$ for some $w', {V'}^\tau$ with wRw' and ${V'}^\tau$ a w'-image of V^τ at w. So $w, V^\tau \nVdash \Box B$.

   In the other direction, assume $w, V^\tau \nVdash \Box B$. Then $w', V^* \nVdash B$ for some w', V^* with wRw' and V^* a w' image of V^τ at w. This means that there is a counterpart relation $C \in K_{w,w'}$ such that for all variables x, $V^*_{w'}(x)$ is some C-counterpart at w' of $V^\tau_w(x)$ at w (if any, else undefined). Let V' be like V except that for all variables x, $V'_{w'}(x^\tau) = V^*_{w'}(x)$, and for all $x \notin \mathrm{Ran}(\tau)$, $V'_{w'}(x)$ is an arbitrary C-counterpart of $V_w(x)$, or undefined if there is none. V' is a w' image of V at w. Moreover, V^* is ${V'}^\tau$. By induction hypothesis, $w', V' \nVdash B^\tau$. So $w', V' \nVdash B^\tau$ for some w', V' with wRw' and V' a w' image of V at w. So $w, V \nVdash (\Box B)^\tau$. ∎

Here I use the fact that V^τ modifies V relative to all worlds.

For the substitution quantifier, we could introduce primitive polyadic quantifiers like $\langle y_1, y_2 : x_1, x_2\rangle$, which says '$y_1$ is an x_1 and y_2 an x_2 such that', and stipulate that

$$w, V \Vdash \langle y_1, y_2 : x_1, x_2\rangle A \text{ iff } w, V^{[y_1,y_2/x_1,x_2]} \Vdash A.$$

Geach's $\langle x : y, z\rangle$ is then equivalent to $\langle x, x : y, z\rangle$. But it turns out that $\langle y_1, y_2 : x_1, x_2\rangle$ is definable.

We can't simply say that $\langle y_1, y_2 : x_1, x_2\rangle A$ is $\langle y_1 : x_1\rangle\langle y_2 : x_2\rangle A$, since the bound variable x_1 might capture y_2, as in the swapping operator $\langle x, y : y, x\rangle$. We must store the original value of y_2 in a temporary variable z: $\langle y_2 : z\rangle\langle y_1 : x_1\rangle\langle z : x_2\rangle$.

Definition 3.14 (Substitution sequences)
For any $n > 1$, sentence A and variables $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ such that the $x_1, \ldots, x_n$ are pairwise distinct, let $\langle y_1, \ldots, y_n : x_1, \ldots, x_n\rangle A$ abbreviate $\langle y_n : z\rangle\langle y_1, \ldots, y_{n-1} : x_1, \ldots, x_{n-1}\rangle\langle z : x_n\rangle A$, where z is the alphabetically first variable not in A or $x_1, \ldots, x_n$.

Lemma 3.15 (Substitution sequence semantics)
For any world w in any structure $\mathcal{S}$, any interpretation V on $\mathcal{S}$,

$$w, V \Vdash_{\mathcal{S}} \langle y_1, \ldots, y_n : x_1, \ldots, x_n\rangle A \text{ iff } w, V^{[y_1,\ldots,y_n/x_1,\ldots,x_n]} \Vdash_{\mathcal{S}} A.$$

Proof By definition 3.14, $w, V \Vdash \langle y_1, \ldots, y_n : x_1, \ldots, x_n\rangle A$ iff $w, V \Vdash \langle y_n : z\rangle\langle y_1, \ldots, y_{n-1} : x_1, \ldots, x_{n-1}\rangle\langle z : x_n\rangle A$, for some z not in $x_1, \ldots, x_{n-1}, A$. By definition 3.2, $w, V \Vdash \langle y_n : z\rangle\langle y_1, \ldots, y_{n-1} : x_1, \ldots, x_{n-1}\rangle\langle z : x_n\rangle A$ iff $w, V^{[y_n/z]} \Vdash \langle y_1, \ldots, y_{n-1} : x_1, \ldots, x_{n-1}\rangle\langle z : x_n\rangle A$, which by induction hypothesis holds iff $w, V^{[y_n/z]\cdot[y_1,\ldots,y_{n-1}/x_1,\ldots,x_{n-1}]} \Vdash \langle z : x_n\rangle A$. By definition 3.2 again, $w, V^{[y_n/z]\cdot[y_1,\ldots,y_{n-1}/x_1,\ldots,x_{n-1}]} \Vdash \langle z : x_n\rangle A$ iff $w, V^{[y_n/z]\cdot[y_1,\ldots,y_{n-1}/x_1,\ldots,x_{n-1}]\cdot[z/x_n]} \Vdash A$. Now $[y_n/z] \cdot [y_1, \ldots, y_{n-1}/x_1, \ldots, x_{n-1}] \cdot [z/x_n]$ is the function $\sigma : Var \to Var$ such that

$$\sigma(x) = [y_n/z]([y_1, \ldots, y_{n-1}/x_1, \ldots, x_{n-1}]([z/x_n](x))).$$

Since $z \notin x_1, \ldots, x_{n-1}$, this means that

$$\sigma(x_n) = y_n,$$
$$\sigma(x_i) = y_i \text{ for } x_i \in \{x_1, \ldots, x_{n-1}\},$$
$$\sigma(z) = y_n,$$

and $\sigma(x) = x$ for every other variable x. Since $z \notin \mathit{Var}(A)$, V^σ agrees at w with $V^{[y_1,\ldots,y_n/x_1,\ldots,x_n]}$ about all variables in A. So by the coincidence lemma 2.10, $w, V^\sigma \Vdash A$ iff $w, V^{[y_1,\ldots,y_n/x_1,\ldots,x_n]} \Vdash A$. ∎

4 Logics

I now want to describe the minimal logics that are characterised by our semantics. Following tradition, a *logic* (or *system*) in this context is simply a set of formulas, and I will describe such sets by recursive clauses corresponding to the axioms and rules of a Hilbert-style calculus.

Recall that we have two kinds of models: positive models with two domains, and negative models with a single domain. Accordingly we have two kinds of logics.

The logic of all positive models is essentially the combination of standard positive free logic with the propositional modal logic K. The only place to be careful is with substitution principles like Leibniz' Law, which either have to be expressed with object-language substitution or restricted as explained in the previous section. (If we add the unrestricted principles, we get logics for functional structures, as we'll see later.)

For a given first-order language language $\mathcal{L}$, standard positive free logic can be characterized as the smallest set L of formulas that contains

(Taut) all propositional tautologies in $\mathcal{L}$,

as well as all $\mathcal{L}$-instances of

(UD) $\forall xA \supset (\forall x(A \supset B) \supset \forall xB)$,

(VQ) $A \supset \forall xA$, provided x is not free in A,

(FUI) $\forall xA \supset (Ey \supset [y/x]A)$,

($\forall$Ex) $\forall xEx$,

(=R) $x = x$,

(LL) $x = y \supset A \supset [y/x]A$,

and that is closed under modus ponens, universal generalisation, and variable substitution:

(MP) if $\vdash_L A$ and $\vdash_L A \supset B$, then $\vdash_L B$,

(UG) if $\vdash_L A$, then $\vdash_L \forall xA$,

(Sub) if $\vdash_L A$, then $\vdash_L [y/x]A$.

Here, as always, $\vdash_L A$ means $A \in L$.[4]

In the logic of positive counterpart structures (without object-language substitution), (LL), (FUI) and (Sub) are restricted to cases where y is modally free for x in A:

4 [Fitting and Mendelsohn 1998] use $\forall xA \leftrightarrow A$ in place of (VQ), which precludes empty inner domains (as [Kutz 2000: 38] points out). The systems presented here do not validate the claim that something exists. To rule out empty inner domains in positive and negative models, $\exists xEx$ would be needed as extra axiom.

(FUI*) $\forall x A \supset (Ey \supset [y/x]A)$, provided y is modally free for x in A,

(LL*) $x = y \supset A \supset [y/x]A$, provided y is modally free for x in A,

(Sub*) if $\vdash_L A$, then $\vdash_L [y/x]A$, provided y is modally free for x in A.

In addition, we include the modal schema

(K) $\Box A \supset (\Box(A \supset B) \supset \Box B)$,

and closure under necessitation,

(Nec) if $\vdash_L A$, then $\vdash_L \Box A$.

> Definition 4.1 (Minimal positive (quantified modal) logic)
> Given a standard language $\mathcal{L}$, the *minimal positive (quantified modal) logic* P in $\mathcal{L}$ is the smallest set $L \subseteq \mathcal{L}$ that contains all $\mathcal{L}$-instances of (Taut), (UD), (VQ), (FUI*), ($\forall$Ex), (=R), (LL*), (K), and that is closed under (MP), (UG), (Nec) and (Sub*).

Should I write PK instead of P, analogously NK, QK, PSK, NSK?

We shall also be interested in stronger logics adequate for various classes of counterpart structures. As a first stab, I will adopt the following definition.

> Definition 4.2 (Positive logics)
> Given a standard language $\mathcal{L}$, a *positive (quantified modal) logic* in $\mathcal{L}$ is a an extension $L \supseteq \mathsf{P}$ of the minimal positive logic P in $\mathcal{L}$ such that L is closed under (MP), (UG), (Nec) and (Sub*).

As it stands, definition 4.2 allows for "logics" in which (say) $F_1 x$ is a theorem but not $F_2 x$. This clashes with the idea that logical truths should be independent of the interpretation of non-logical terms. A more adequate definition would add a second-order closure condition to the effect that, roughly, whenever $\vdash_L A$ then $\vdash_L [B/Px_1 \ldots x_n]A$, where $[B/Px_1 \ldots x_n]A$ is A with all occurrences of the atomic formula $Px_1 \ldots x_n$ replaced by (the arbitrary formula) B. Making this precise requires some care, especially once we look at negative logics where $Fx \supset Ex$ is valid, but $\neg Fx \supset Ex$ is not. Definition 4.2 will do as long as we start off with a class of counterpart structures and look for a corresponding positive logic; that logic will always satisfy the definition 4.2.

The first-order closure condition (Sub*) at least excludes logics in which, for example, Fx is a theorem but not Fy. To see why (Sub*) needs the proviso, note that we could otherwise move from the (FUI*) instance $\vdash_L \forall x \Diamond Gxy \supset \Diamond Gzy$ to $\vdash_L \forall x \Diamond Gx \supset \Diamond Gyy$, which is invalid as long as individuals can have multiple counterparts.

You may also wonder why (Sub*) is needed in definition 4.1, since that the axioms are given by schemas: doesn't this mean that every substitution instance of an axiom is itself an axiom, and isn't this property of closure under substitution preserved by (MP), (UG) and (Nec)? Not quite. For example,

(6) $\quad v = y \supset (v = z \supset y = x)$

is an instance of (LL*), and

(7) $\quad x = y \supset (x = x \supset y = x)$

follows from (6) by (Sub*), but (7) is not itself an instance of (LL*). Of course it is possible to axiomatise P without (Sub*), and nothing really hangs on it. I have chosen the above axiomatisation just because I find it comparatively intuitive and convenient for the purposes of this paper. (For example, it frees me from having to prove that P is a positive logic in the sense of definition 4.2.)

> Theorem 4.3 (Soundness of P)
> Every member of P is valid in every positive counterpart model.

Proof We show that all P axioms are valid in every positive model, and that validity is closed under (MP), (UG), (Nec) and (Sub*).

1. (Taut). Propositional tautologies are valid in every model by the standard satisfaction rules for the connectives.
2. (UD). Assume $w, V \Vdash \forall x(A \supset B)$ and $w, V \Vdash \forall x A$ in some model. By definition 2.7, then $w, V' \Vdash A \supset B$ and $w, V' \Vdash A$ for every existential x-variant V' of V on w, and so $w, V' \Vdash B$ for every such V'. Hence $w, V \Vdash \forall x B$.
3. (VQ). Suppose $w, V \nVdash A \supset \forall x A$ in some model. Then $w, V \Vdash A$ and $w, V \nVdash \forall x A$. If x is not free in A, then by the coincidence lemma 2.10, $w, V' \Vdash A$ for every x-variant V' of V on D_w; so $w, V \Vdash \forall x A$. Contradiction. So if x is not free in A, then $A \supset \forall x A$ is valid in every model.
4. (FUI*). Assume $w, V \Vdash \forall x A$ and $w, V \Vdash Ey$ in some model. By definition 2.7, then $w, V' \Vdash A$ for all existential x-variants V' of V on w. So in particular, $w, V^{[y:x]} \Vdash A$. If y is modally free for x in A, then by lemma 3.9, $w, V \Vdash [y/x]A$.
5. ($\forall$Ex). By definition 2.7, $w, V \Vdash \forall x Ex$ iff $w, V' \Vdash Ex$ for all existential x-variants V' of V on w, iff for all existential x-variants V' of V on w there is an existential y-variant V'' of V' on w such that $w, V'' \Vdash x = y$. But this is always the case: for any V', let V'' be $V'^{[x/y]}$.

6. (=R). By definition 2.3, $V_w(=) = \{\langle d, d\rangle : d \in U_w\}$, and by definition 2.3, $V_w(x) \in U_w$ in every positive model. So $w, V \Vdash x{=}x$ in every such model, by definition 2.7.
7. (LL*). Assume $w, V \Vdash x = y$, $w, V \Vdash A$, and y is modally free for x in A. Since $V_w(x) = V_w(y)$, V coincides with $V^{[y/x]}$ at w. So $w, V^{[y/x]} \Vdash A$ by the coincidence lemma 2.10. By lemma 3.9, $w, V^{[y/x]} \Vdash A$ only if $w, V \Vdash [y/x]A$. So $w, V \Vdash [y/x]A$.
8. (K). Assume $w, V \Vdash \Box(A \supset B)$ and $w, V \Vdash \Box A$. Then $w', V' \Vdash A \supset B$ and $w', V' \Vdash A$ for every w', V' such that wRw' and V' is a w'-image of V at w. Then $w', V' \Vdash B$ for any such w', V', and so $w, V \Vdash \Box B$.
9. (MP). Assume $w, V \Vdash A \supset B$ and $w, V \Vdash A$ in some model. By definition 2.7, then $w, V \Vdash B$ as well. So for any world w in any model, (MP) preserves truth at w.
10. (UG). Assume $w, V \nVdash \forall x A$ in some model $\mathcal{M}$. Then $w, V' \nVdash A$ for some existential x-variant V' of V on w. So A is invalid in a model like $\mathcal{M}$ but with V' as the interpretation function in place of V. Hence if A is valid in all positive models, then so is $\forall x A$.
11. (Nec). Assume $w, V \nVdash_{\mathcal{M}} \Box A$ in some model $\mathcal{M}$. Then $w', V' \nVdash A$ for some w' with wRw' and V' some w'-image of V at w. Let $\mathcal{M}^*$ be like $\mathcal{M}$ except with V' in place of V. $\mathcal{M}^*$ is a positive model. Since A is not valid in $\mathcal{M}^*$, it follows contrapositively that whenever A is valid in all positive models, then so is $\Box A$.
12. (Sub*). Assume $w, V \nVdash [y/x]A$ in some model $\langle \mathcal{S}, V\rangle$, and y is modally free for x in A. By lemma 3.9, then $w, V^{[y/x]} \nVdash A$. So A is invalid in the model $\langle \mathcal{S}, V^{[y/x]}\rangle$. Hence if A is valid in all positive models, then so is $[y/x]A$. ■

It would be nice to proof something stronger here: that for every set of counterpart structures, the set of formulas valid in that set is a logic that extends P; cf. [Gabbay et al. 2009: 80]. Arguably that should go in the Correspondence section.

Let's move on to negative logics. Standard negative free logic replaces (=R) and (∀Ex) by

(∀=R) $\forall x(x{=}x)$,

(Neg) $Px_1 \ldots x_n \supset Ex_1 \wedge \ldots \wedge Ex_n$.

In our single-domain models, we need two further axioms, as mentioned on p. 5:

(NA) $\neg Ex \supset \Box\neg Ex$,

(TE) $x{=}y \supset \Box(Ex \supset Ey)$.

TE could be proved from $y = y$, but $y = y$ is not valid in negative logics. It could also be proved from $\Box(x{=}x \supset x{=}x)$ by Leibniz' Law, if that would licence substituting the last two x occurrences by y.

(In negative dual-domain semantics, (NA) and (TE) are invalid while the other axioms of N are valid. Negative dual-domain models are like positive models except that all predicates, including identity, are restricted to the inner domain. Thus (NA) is false at w if some individual in the outer domain of w has a counterpart in the inner domain of some accessible world. (TE) is false at w if $V_w(x) \in D_w$ has both an existing and a non-existing counterpart at some accessible world w'; then some w'-image V' of V at w assigns the non-existing individual to y and the existing individual to x, rendering $Ex \supset Ey$ false at w' under V', wherefore $\Box(Ex \supset Ey)$ is false at w under V.)

As I said in section 2, (NA) should not be confused with the claim that no individual exists at any world that isn't a counterpart of something at the centre, which would require something like the Barcan Formula,

(BF) $\forall x \Box A \supset \Box \forall x A$.

But this isn't valid in the class of negative models. For example, if $W = \{w, w'\}$, wRw', $D_w = \emptyset$ and $D_{w'} = \{0\}$, then $w, V \Vdash \forall x \Box x \neq x$, but $w, V \nVdash \Box \forall x\ x \neq x$.

Definition 4.4 (Minimal (strongly) negative (quantified modal) logic)
Given a standard language $\mathcal{L}$, the *minimal (strongly) negative (quantified modal) logic* N in $\mathcal{L}$ is the smallest set $L \subseteq \mathcal{L}$ that contains all $\mathcal{L}$-instances of (Taut), (UD), (VQ), (FUI*), (Neg), (LL*), ($\forall$=R), (K), (NA), (TE), and that is closed under (MP), (UG), (Nec) and (Sub*).

Definition 4.5 (Negative logics)
Given a standard language $\mathcal{L}$, a *negative (quantified modal) logic* in $\mathcal{L}$ is a an extension $L \supseteq \mathsf{N}$ of the minimal negative logic N in $\mathcal{L}$ such that L is closed under (MP), (UG), (Nec) and (Sub*).

Theorem 4.6 (Soundness of N)
Every member of N is valid in every negative counterpart model.

Proof We show that all N axioms are valid in every negative model, and that validity is closed under (MP), (UG), (Nec) and (Sub*). The proofs for (Taut), (UD), (VQ), (FUI*), (LL*), (K), (MP), (UG), (Nec) and (Sub*) are just as in theorem 4.3. The remaining cases are

1. (Neg). Assume $w, V \Vdash Px_1 \ldots x_n$ in some model. By definition 2.3, $V_w(P) \subseteq U_w^n$, and by definition 2.4, $U_w = D_w$ in negative models. So $V_w(P) \subseteq D_w^n$. By definition 2.7, $w, V \Vdash Px_1 \ldots x_n$ therefore entails that $V_w(x_i) \in D_w$ for all $x_i \in x_1, \ldots, x_n$, and that $w, V \Vdash Ex_i$ for all such x_i.

2. ($\forall$=R). $w, V \Vdash \forall x(x{=}x)$ iff $w, V' \Vdash x{=}x$ for all existential x-variants V' of V on w. This is always the case, since by definition 2.3 and 2.3, $V_w(=) = \{\langle d, d\rangle : d \in D_w\}$ in negative models.

3. (NA). Assume $w, V \Vdash \neg Ex$. By definition 2.7, this means that $V_w(x) \notin D_w$, and therefore that $V_w(x)$ is undefined if the model is negative. But if $V_w(x)$ is undefined, then there is no world w', individual d and counterpart relation $C \in K_{w,w'}$ such that $\langle V_w(x), w\rangle C\langle d, w'\rangle$. By definitions 2.7 and 2.6, it follows that there is no world w' and interpretation V' with wRw' and $V_w \triangleright V'_{w'}$ such that $w', V' \Vdash Ex$. So then $w, V \Vdash \Box\neg Ex$ by definition 2.7. Thus $w, V \Vdash \neg Ex \supset \Box\neg Ex$.

4. (TE). Assume $w, V \Vdash x{=}y$. Then $V_w(x) = V_w(y)$ by definitions 2.3 and 2.7. Let w', V' be such that wRw' and $V_w \triangleright V'_{w'}$, and $w', V' \Vdash Ex$. By definition 2.7, the latter means that $V'_{w'}(x)$ is some member of D_w. Moreover, $V_w \triangleright V'_{w'}$ means that there is a $C \in K_{w,w'}$ such that this $V'_{w'}(x) \in D_w$ is a C-counterpart of $V_w(x)$. It follows that $V_w(y) = V_w(x)$ has at least one C-counterpart at w', so $V'_{w'}(y)$ must be some such counterpart, which can only be in D_w. So $w', V' \Vdash Ey$. So if $w, V \Vdash x{=}y$, then $w, V \Vdash \Box(Ex \supset Ey)$, by definition 2.7, and so $w, V \Vdash x{=}y \supset \Box(Ex \supset Ey)$. ■

In the remainder of this section, I will prove a few properties derivable from the above axiomatisations. (Some of these will be needed later on in the completeness proof.) To this end, let $\mathcal{L}$ range over standard languages of quantified modal logic, and L over an arbitrary positive or negative logic in $\mathcal{L}$.

Lemma 4.7 (Closure under propositional consequence)
For all $\mathcal{L}$-formulas $A_1, \ldots, A_n, B$,

(PC) if $\vdash_L A_1, \ldots, \vdash_L A_n$, and B is a propositional consequence of $A_1, \ldots, A_n$, then $\vdash_L B$.

Proof If B is a propositional consequence of $A_1, \ldots, A_n$, then $A_1 \supset (\ldots \supset (A_n \supset B)\ldots)$ is a tautology. So by (Taut), $\vdash_L A_1 \supset (\ldots \supset (A_n \supset B)\ldots)$. If $\vdash_L A_1, \ldots, \vdash_L A_n$, then by n applications of (MP), $\vdash_L B$. ■

When giving proofs, I will often omit reference to (PC).

Lemma 4.8 (Redundant axioms)
For any $\mathcal{L}$-formulas A and variables x,

($\forall$Ex) $\vdash_L \forall x Ex$,

($\forall$=R) $\vdash_L \forall x(x{=}x)$.

Proof If L is positive, then ($\forall$Ex) is an axiom. In N, we have $\vdash_L x = x \supset Ex$ by (Neg); so by (UG) and (UD), $\vdash_L \forall x(x = x) \supset \forall x Ex$. Since $\vdash_L \forall x(x = x)$ by (=R), $\vdash_L \forall x Ex$.

If L is negative, then ($\forall$=R) is an axiom. In P, we have $\vdash_L x = x$ by (=R), and so ($\forall$=R) by (UG). ■

Lemma 4.9 (Existence and Self-Identity)
If L is *negative*, then for any $\mathcal{L}$-variable x,

(EI) $\vdash_L Ex \leftrightarrow x = x$;

Proof By (FUI*), $\vdash_L \forall x(x = x) \supset (Ex \supset x = x)$. By ($\forall$=R), $\vdash_L \forall x(x = x)$. So $\vdash_L Ex \supset x = x$. Conversely, $x = x \supset Ex$ by (Neg). ■

Lemma 4.10 (Symmetry and transitivity of identity)
For any $\mathcal{L}$-variables x, y, z,

(=S) $\vdash_L x = y \supset y = x$;

(=T) $\vdash_L x = y \supset y = z \supset x = z$.

Proof For (= S), let v be some variable $\notin \{x, y\}$. Then

1. $\vdash_L v = y \supset (v = x \supset y = x)$. (LL*)
2. $\vdash_L x = y \supset (x = x \supset y = x)$. (1, (Sub*))
3. $\vdash_L x = y \supset x = x$. ((=R), or (Neg) and ($\forall$=R))
4. $\vdash_L x = y \supset y = x$. (2, 3)

For (= T),

1. $\vdash_L x = y \supset y = x$. (=S)
2. $\vdash_L y = x \supset (y = z \supset x = z)$. (LL*)
3. $\vdash_L x = y \supset (y = z \supset x = z)$. (1, 2)

■

Next we have proof-theoretic analogues of lemmas 3.8 and 3.13:

Lemma 4.11 (Syntactic alpha-conversion)
If A, A' are $\mathcal{L}$-formulas, and A' is an alphabetic variant of A, then

(AC) $\vdash_L A \leftrightarrow A'$.

Proof by induction on A.

1. A is atomic. Then $A = A'$ and $A \leftrightarrow A'$ is a propositional tautology.
2. A is $\neg B$. Then A' is $\neg B'$, where B' is an alphabetic variant of A'. By induction hypothesis, $\vdash_L B \leftrightarrow B'$. So by (PC), $\vdash_L \neg B \leftrightarrow \neg B'$.
3. A is $B \supset C$. Then A' is $B' \supset C'$, where B', C' are alphabetic variants of B, C, respectively. By induction hypothesis, $\vdash_L B \leftrightarrow B'$ and $\vdash_L C \leftrightarrow C'$. So by (PC), $\vdash_L (B \supset C) \leftrightarrow (B' \supset C')$.
4. A is $\forall x B$. Then A' is either $\forall x B'$ or $\forall z[z/x]B'$, where B' is an alphabetic variant of B and $z \notin \mathit{Var}(B')$. Assume first that A' is $\forall x B'$. By induction hypothesis, $\vdash_L B \leftrightarrow B'$. So by (UG) and (UD), $\vdash_L \forall x B \leftrightarrow \forall x B'$.

   Alternatively, assume B is $\forall z[z/x]B'$ and $z \notin \mathit{Var}(B')$. Since B' differs from B at most in renaming bound variables, if z were free in B, then $z \in \mathit{Var}(B')$. So z is not free in B. Then

1.	$\vdash_L B \leftrightarrow B'$	(induction hypothesis)
2.	$\vdash_L [z/x]B \leftrightarrow [z/x]B'$	(1, (Sub*))
3.	$\vdash_L \forall x B \supset Ez \supset [z/x]B$	(FUI*)
4.	$\vdash_L \forall x B \supset Ez \supset [z/x]B'$	(2, 3)
5.	$\vdash_L \forall z \forall x B \supset \forall z Ez \supset \forall z[z/x]B'$	(4, (UG), (UD))
6.	$\vdash_L \forall z Ez$	($\forall$Ex)
7.	$\vdash_L \forall z \forall x B \supset \forall z[z/x]B'$	(5, 6)
8.	$\vdash_L \forall x B \supset \forall z \forall x B$	((VQ), z not free in B)
9.	$\vdash_L \forall x B \supset \forall z[z/x]B'$.	(7, 8)

   Conversely,

10.	$\vdash_L \forall z[z/x]B' \supset Ex \supset [x/z][z/x]B'$	(FUI*)
11.	$\vdash_L \forall z[z/x]B' \supset Ex \supset B$	(1, 10, $z \notin \mathit{Var}(B')$)
12.	$\vdash_L \forall x \forall z[z/x]B' \supset \forall x B$	(11, (UG), (UD), ($\forall$Ex))
13.	$\vdash_L \forall z[z/x]B' \supset \forall x \forall z[z/x]B'$	(VQ)
14.	$\vdash_L \forall z[z/x]B' \supset \forall x B$	(12, 13)

5. A is $\Box B$. Then A' is $\Box B'$, where B' is an alphabetic variant of B. By induction hypothesis, $\vdash_L B \leftrightarrow B'$. Then by (Nec), $\vdash_L \Box(B \leftrightarrow B')$, and by (K) and (PC), $\vdash_L \Box B \leftrightarrow \Box B'$. ■

Lemma 4.12 (Closure under transformations)
For any $\mathcal{L}$-formula A and transformation τ on $\mathcal{L}$,

(Sub$^\tau$) $\vdash_L A$ iff $\vdash_L A^\tau$.

Proof Assume $\vdash_L A$. Let $x_1, \ldots, x_n$ be the variables in A. If $n = 0$, then $A = A^\tau$ and the result is trivial. If $n = 1$, then A^τ is $[x_1^\tau/x_1]A$, and x_1^τ is either x_1 itself or does not occur in A. In the first case, $[x_1^\tau/x_1]A = A$ and the result is again trivial. In the second case, x_1^τ is modally free for x_1 in A, and thus $\vdash_L [x_1^\tau/x_1]A$ by (Sub*).

Assume then that $n > 1$. Note first that $A^\tau = [x_n^\tau/v_n] \ldots [x_2^\tau/v_2][x_1^\tau/x_1][v_2/x_2] \ldots [v_n/x_n]A$, where $v_2, \ldots, v_n$ are distinct variables not in A or A^τ. This is easily shown by induction on the subformulas B of A (ordered by complexity). To keep things short, let Σ abbreviate $[x_n^\tau/v_n] \ldots [x_2^\tau/v_2][x_1^\tau/x_1][v_2/x_2] \ldots [v_n/x_n]$.

1. If B is $Px_j \ldots x_k$, then $x_j, \ldots, x_k$ are variables from $x_1, \ldots, x_n$, and $\Sigma B = Px_j^\tau \ldots x_k^\tau = B^\tau$, by definitions 3.3 and 3.11.
2. If B is $\neg C$, then by induction hypothesis, $\Sigma C = C^\tau$, and hence $\neg\Sigma C = \neg C^\tau$. But $\Sigma\neg C$ is $\neg\Sigma C$ by definition 3.3, and $(\neg C)^\tau$ is $\neg C^\tau$ by definition 3.11.
3. The case for $C \supset D$ is analogous.
4. If B is $\forall zC$, then by induction hypothesis, $\Sigma C = C^\tau$. Since τ is injective, $\Sigma\forall zC$ is $\forall\Sigma z\Sigma C$ by definition 3.3, and $(\forall zC)^\tau$ is $\forall z^\tau C^\tau$ by definition 3.11. Moreover, since z is one of $x_1, \ldots, x_n$, $\Sigma z = z^\tau$. (Here things would get a lot more complicated if we had defined substitution differently, so that $[y/x]\forall xFx \neq \forall yFy$.)
5. If B is $\Box C$, then by induction hypothesis, ΣC is C^τ, and hence $\Box\Sigma C$ is $\Box C^\tau$. But $\Sigma\Box C$ is $\Box\Sigma C$ by definition 3.3, and $(\Box C)^\tau$ is $\Box C^\tau$ by definition 3.11.

Now we show that L contains all "segments" of $[x_n^\tau/v_n] \ldots [x_2^\tau/v_2][x_1^\tau/x_1][v_2/x_2] \ldots [v_n/x_n]A$, beginning with the rightmost substitution, $[v_n/x_n]A$. Since v_n is modally free for x_n in A, by (Sub*), $\vdash_L [v_n/x_n]A$. Likewise, for each $1 < i < n$, v_i is modally free for x_i in $[v_{i+1}/x_{i+1}] \ldots [v_n/x_n]A$. So $\vdash_L [v_2/x_2] \ldots [v_n/x_n]A$.

With respect to $[x_1^\tau/x_1]$, we distinguish three cases. First, if $x_1 = x_1^\tau$, then $\vdash_L [x_1^\tau/x_1][v_2/x_2] \ldots [v_n/x_n]A$, because $[x_1^\tau/x_1][v_2/x_2] \ldots [v_n/x_n]A$ is $[v_2/x_2] \ldots [v_n/x_n]A$. Second, if $x_1 \neq x_1^\tau$ and $x_1^\tau \notin \mathit{Var}(A)$, then $x_1^\tau \notin \mathit{Var}([v_2/x_2] \ldots [v_n/x_n]A)$, since the $v_1, \ldots, v_n$ are not in $\mathit{Var}(A)$ or $\mathit{Var}(A^\tau)$ (in particular, thus no new variables are introduced in $[v_2/x_2] \ldots [v_n/x_n]A$). So x_1^τ is modally free for x_1 in $[v_2/x_2] \ldots [v_n/x_n]A$, and by (Sub*), $\vdash_L [x_1^\tau/x_1][v_2/x_2] \ldots [v_n/x_n]A$. Third, if $x_1 \neq x_1^\tau$ and $x_1^\tau \in \mathit{Var}(A)$, then x_1^τ must be one of $x_2, \ldots, x_n$. Then again $x_1^\tau \notin \mathit{Var}([v_2/x_2] \ldots [v_n/x_n]A)$, and so $\vdash_L [x_1^\tau/x_1][v_2/x_2] \ldots [v_n/x_n]A$ by (Sub*).

Next, x_2^τ is modally free for v_2 in $[x_1^\tau/x_1][v_2/x_2] \ldots [v_n/x_n]A$, because τ is injective and hence $x_2^\tau \neq x_1^\tau$, so x_2^τ does not occur in $[x_1^\tau/x_1][v_2/x_2] \ldots [v_n/x_n]A$. Hence $\vdash_L [x_2^\tau/v_2][x_1^\tau/x_1]$ $[v_2/x_2] \ldots [v_n/x_n]A$. By the same reasoning, for each $2 < i \leq n$, x_i^τ is modally free for v_i in $[x_{i-1}^\tau/v_{i-1}] \ldots [x_2^\tau/v_2][x_1^\tau/x_1][v_2/x_2] \ldots [v_n/x_n]A$. So $\vdash_L [x_n^\tau/v_n] \ldots [x_2^\tau/v_2][x_1^\tau/x_1][v_2/x_2] \ldots$ $[v_n/x_n]A$, i.e. $\vdash_L A^\tau$.

This proves the left-to-right direction of (Sub$^\tau$). The other direction immediately follows. Let $x_1^\tau, \dots, x_n^\tau$ be the variables in A^τ, and let σ be an arbitrary transformation that maps each x_i^τ back to x_i (i.e., to $(x_i^\tau)^{\tau^{-1}}$). By the left-to-right direction of (Sub$^\tau$), $\vdash_L A^\tau$ entails $\vdash_L (A^\tau)^\sigma$, and $(A^\tau)^\sigma$ is simply A. ■

In a similar way, one can probably prove closure under generalised substitutions, provided all the $\mathit{Var}(A)^\sigma$ are somehow modally free for the variables in $\mathit{Var}(A)$. Lemma 4.11 should come in handy then.

LEMMA 4.13 (LEIBNIZ' LAW WITH PARTIAL SUBSTITUTION)
Let A be a formula of $\mathcal{L}$, and x, y variables of $\mathcal{L}$. Let $[y//x]A$ be A with *one or more* occurrences of x replaced by y.

(LL$_\text{p}^*$) $\vdash_L x\!=\!y \supset A \supset [y//x]A$, provided the following conditions are satisfied.

(i) $[y//x]A$ does not replace any occurrence of x in the scope of a quantifier binding x or y.

(ii) Either y is modally free for x in A, or $[y//x]A$ does not replace any occurrence of x in the scope of a modal operator in A that also contains y.

(iii) In the scope of any modal operator in A, $[y//x]A$ either replaces all or no occurrences of x by y.

PROOF Let $v \neq y$ be a variable not in $\mathit{Var}(A)$, and let $[v//x]A$ be like $[y//x]A$ except that all new occurrences of y are replaced by v: if $[y//x]A$ satisfies (i)–(iii), then so does $[y//x]A$ with all new occurrences of y replaced by v. Moreover, in the resulting formula $[v//x]A$ all occurrences of v are free and free for y, by clause (i); so $[y/v][v//x]A = [y//x]A$ by definition 3.3. By (LL*),

(1) $\vdash_L v\!=\!y \supset [v//x]A \supset [y/v][v//x]A$,

provided that y is modally free for v in $[v//x]A$, i.e. provided that either y is modally free for x in A, or $[v//x]A$ (and thus $[y//x]A$) does not replace any occurrence of x in the scope of a modal operator in A that also contains y. (The first clause is not redundant because A can be a box formula that contains x and y; in this case it is OK to replace x.) This is guaranteed by condition (ii). Since $[y/v][v//x]A$ is $[y//x]A$, (1) can be shortened to

(2) $\vdash_L v\!=\!y \supset [v//x]A \supset [y//x]A$.

By (Sub*), it follows that

(3) $\vdash_L [x/v](v\!=\!y \supset [v//x]A \supset [y//x]A)$,

provided that x is modally free for v in $v\!=\!y \supset [v//x]A \supset [y//x]A$. Since this isn't a formula of the form $\Box B$, x is modally free for v here iff no free occurrences of x and v lie in the scope of the same modal operator in $[v//x]A$. So whenever $[v//x]A$ (and thus $[y//x]A$) replaces some occurrences of x in the scope of a modal operator in A, then it must replace all occurrences of x in the scope of that operator. (It is not enough that x is modally free for v in $[v//x]A$: consider $x = y \supset \Box x = x \supset \Box x = y$, which the application of (Sub*) would derive from $x{=}v \supset \Box x{=}v \supset \Box x{=}y$.) This is guaranteed by condition (iii). By definition 3.3, (3) can be simplified to

(4) $\quad \vdash_L x{=}y \supset A \supset [y//x]A.$

■

There are other ways to define partial substitution, more along the lines of definition 3.3. But we have to be careful. For example, we would need to ensure that $[y//x]\forall xFx$ is not $\forall xFy$ or $\forall yFx$. To this end, we could stipulate that $[y//x]$ replaced either all or no variables in the scope of any quantifier binding x – paralleling clause (iii). Care is also required if we want to retain the existence of $[v//x]A$ with $[y/v][v//x]A = [y//x]A$. For example, simply following a partial version of definition 3.3 would allow $\forall zFx$ as an instance of $[y//x]\forall yFx$. The corresponding formula $[v//x]\forall yFx$ would then have to be $\forall zFv$, which is not licensed by the definition of partial substitution; rather, $[v//x]\forall yFx$ should be either $\forall yFv$ or $\forall yFx$, and so $[y/v][v//x]\forall yFx$ is either $\forall zFy$ or $\forall yFx$ – we don't get $\forall zFx$.

I will never actually use (LL^*_p). I mention it only because Leibniz' Law is often stated for partial substitutions, and you may have wondered how the corresponding version looks in the present systems. Now you know. We could indeed have used (LL^*_p) as basic axiom instead of (LL*); (LL*) would then be derivable, because every formula A has an alphabetic variant A' such that $[y/x]A$ is an instance of $[y//x]A'$ that satisfies (i)–(iii) iff y is modally free for x in A, and because (LL*) is not used in the proof of lemma 4.11. If there is no clash of bound variables, $A' = A$ and $[y/x]A$ is simply an instance of $[y//x]A$ replacing all rather than only some occurrences of x. Clause (iii) is trivial in this case, and clause (ii) reduces to y being modally free for x in A. If variables clash, $[y/x]A$ is not an instance of $[y//x]A$. For example, $[y/x]\forall yFx$ is $\forall vFy$, while the only $[y//x]A$ is $\forall yFx$, by clause (i). Here A' is $\forall vFx$. By lemma 4.11, $\vdash_L A \leftrightarrow A'$. So $\vdash_L A \land x{=}y \supset [y/x]A$ iff $\vdash_L A' \land x{=}y \supset [y/x]A$. If $[y/x]A$ is an instance of $[y//x]A'$, the latter is validated by (LL^*_p). In the example, $[y/x]A$ is $[y/x]\forall yFx = \forall vFy$, and $[y//x]A'$ is $[y//x]\forall vFx = \forall vFx$ or $\forall vFy$. In the case of $A = \forall xFx$, where $[y/x]A = \forall yFy$, A' is $\forall yFy$ as well. I have chosen (LL*) as basic due to its much greater simplicity.[5]

---

5 Kutz's system uses the following version of (LL^*_p) ([Kutz 2000: 43]):

($\text{LL}^\text{K}_\text{p}$) $\vdash x{=}y \supset A \supset [y//x]A$, provided that
(i) x is free in A and y is free for x in A,
(ii) y is not free in the scope of a modal operator in A, and
(iii) in the scope of any modal operator in A, $[y//x]A$ either replaces all or no occurrences of x by y.

Lemma 4.14 (Leibniz' Law with sequences)
For any $\mathcal{L}$-formula A and variables $x_1, \ldots, x_n, y_1, \ldots, y_n$ such that the $x_1, \ldots, x_n$ are pairwise distinct,

(LL$^*_\text{n}$) $\vdash_L x_1 = y_1 \wedge \ldots \wedge x_n = y_n \supset A \supset [y_1, \ldots, y_n/x_1, \ldots, x_n]A$, provided each y_i is modally free for x_i in $[y_1, \ldots, y_{i-1}/x_1, \ldots, x_{n-1}]A$.

For $i = 1$, the proviso is meant to say that y_1 is modally free for x_1 in A.
To explain the proviso, consider

$$x = y \wedge z = y \supset \Diamond x \neq z \supset [y, y/x, z]\Diamond x \neq z \equiv \Diamond y \neq y.$$

We don't want this to be valid, although y is free for x, and y is free for z, in $\Diamond x \neq z$. So it's not enough that each y_i is free for the corresponding x_i. The proviso above requires that y is free for x in $\Diamond x \neq z$ and y is free for z in $[y/x]\Diamond x \neq z \equiv \Diamond y \neq z$, which it isn't.

It is a bit odd that the proviso introduces an order into the x_i and y_i, although order matters neither in conjunction $x_1 = y_1 \wedge x_2 = y_2 \wedge \ldots$ nor in polyadic substitution $[y_1, y_2, \ldots/x_1, x_2, \ldots]$. What if we identify $z = y$ with $x_1 = y_1$ and $x = y$ with $x_2 = y_2$ in the example? The proviso then requires that y is free for z in $\Diamond x \neq z$ and that y is free for x in $[y/z]\Diamond x \neq z \equiv \Diamond x \neq y$. Is it always true that if the proviso fails on some ordering of variables, then it fails on all? If so, shouldn't there be an ordering-free statement of the proviso?

Proof By induction on n. For $n = 1$, (LL$^*_\text{n}$) is (LL*). Assume then that $n > 1$ and that each y_i in $y_1, \ldots, y_n$ is modally free for x_i in $[y_1, \ldots, y_{i-1}/x_1, \ldots, x_{n-1}]A$. Let z be some variable not in A, $x_1, \ldots, x_n$, $y_1, \ldots, y_n$. So z is modally free for x_n in A. By (LL*),

(1) $\vdash_L x_n = z \supset A \supset [z/x_n]A$.

By induction hypothesis,

(2) $\vdash_L x_1 = y_1 \wedge \ldots \wedge x_{n-1} = y_{n-1} \supset [z/x_n]A \supset [y_1, \ldots, y_{n-1}/x_1, \ldots, x_{n-1}][z/x_n]A$.

By assumption, y_n is modally free for x_n in $[y_1, \ldots, y_{n-1}/x_1, \ldots, x_{n-1}]A$. Then y_n is also modally free for z in $[y_1, \ldots, y_{n-1}/x_1, \ldots, x_{n-1}][z/x_n]A$. So by (LL*),

(3) $\vdash_L z = y_n \supset [y_1, \ldots, y_{n-1}/x_1, \ldots, x_{n-1}][z/x_n]A \supset$
$[y_n/z][y_1, \ldots, y_{n-1}/x_1, \ldots, x_{n-1}][z/x_n]A$.

Evidently, this is a lot more restrictive than (LL$^*_\text{p}$). For example, (LL$^*_\text{p}$) validates

$\vdash x = y \supset \Box Gxy \supset \Box Gyy$ and
$\vdash x = y \supset (Fx \vee \Diamond Gxy) \supset (Fy \vee \Diamond Gxy)$,

which can't be derived in Kutz's system.

But $[y_n/z][y_1,\dots,y_{n-1}/x_1,\dots,x_{n-1}][z/x_n]A$ is $[y_1,\dots,y_n/x_1,\dots,x_n]A$. Combining (1)–(3), we therefore have

(4) $\vdash_L x_1 = y_1 \wedge \dots \wedge x_{n-1} = y_{n-1} \supset x_n = z \wedge z = y_n \supset A \supset [y_1,\dots,y_n/x_1,\dots,x_n]A.$

So by (Sub*),

(5) $\vdash_L x_1 = y_1 \wedge \dots \wedge x_{n-1} = y_{n-1} \supset x_n = x_n \wedge x_n = y_n \supset A \supset [y_1,\dots,y_n/x_1,\dots,x_n]A.$

Since $\vdash_L x_n = y_n \supset x_n = x_n$ (by either (=R) or (Neg) and (∀=R)), it follows that

(6) $\vdash_L x_1 = y_1 \wedge \dots \wedge x_n = y_n \supset A \supset [y_1,\dots,y_n/x_1,\dots,x_n]A.$

■

LEMMA 4.15 (CROSS-SUBSTITUTION)
For any $\mathcal{L}$-formula A and variables x, y,

(CS) $\vdash_L x = y \supset \Box A \supset \Box(y = z \supset [z/x]A)$, provided z is not free in A.

More generally, for any variables $x_1,\dots,x_n, y_1,\dots,y_n$ such that the $x_1,\dots,x_n$ are pairwise distinct,

($\mathrm{CS_n}$) $\vdash_L x_1 = y_1 \wedge \dots \wedge x_n = y_n \supset \Box A \supset \Box(y_1 = z_1 \wedge \dots \wedge y_n = z_n \supset [z_1,\dots,z_n/x_1,\dots,x_n]A)$, provided none of $z_1,\dots,z_n$ is free in A.

PROOF For (CS), assume z is not free in A. Then

1. $\vdash_L x = z \supset A \supset [z/x]A.$ (LL*)
2. $\vdash_L A \supset (x = z \supset [z/x]A).$ (1)
3. $\vdash_L \Box A \supset \Box(x = z \supset [z/x]A).$ (2, (Nec), (K))
4. $\vdash_L x = y \supset \Box(x = z \supset [z/x]A) \supset \Box(y = z \supset [z/x]A).$ (LL*)
5. $\vdash_L x = y \supset \Box A \supset \Box(y = z \supset [z/x]A).$ (3, 4)

Step 4 is justified by the fact that x is not free in $[z/x]A$ and so x and y are modally separated in $x = z \supset [z/x]A$.

The proof for ($\mathrm{CS_n}$) is analogous. Assume none of $z_1,\dots,z_n$ is free in A. Then

1. $\vdash_L x_1 = z_1 \wedge \dots \wedge x_n = z_n \supset A \supset [z_1,\dots,z_n/x_1,\dots,x_n]A.$ ($\mathrm{LL^*_n}$)
2. $\vdash_L A \supset (x_1 = z_1 \wedge \dots \wedge x_n = z_n \supset [z_1,\dots,z_n/x_1,\dots,x_n]A).$ (1)
3. $\vdash_L \Box A \supset \Box(x_1 = z_1 \wedge \dots \wedge x_n = z_n \supset [z_1,\dots,z_n/x_1,\dots,x_n]A).$ (2, (Nec), (K))
4. $\vdash_L x_1 = y_1 \wedge \dots \wedge x_n = y_n \supset$
   $\Box(x_1 = z_1 \wedge \dots \wedge x_n = z_n \supset [z_1,\dots,z_n/x_1,\dots,x_n]A) \supset$
   $\Box(y_1 = z_1 \wedge \dots \wedge y_n = z_n \supset [z_1,\dots,z_n/x_1,\dots,x_n]A).$ ($\mathrm{LL^*_n}$)
5. $\vdash_L x_1 = y_1 \wedge \dots \wedge x_n = y_n \supset \Box A \supset$
   $\Box(x_1 = z_1 \wedge \dots \wedge x_n = z_n \supset [z_1,\dots,z_n/x_1,\dots,x_n]A).$ (3, 4)

Step 4 is justified by the fact that none of $x_1, \dots, x_n$ is free in $[z_1, \dots, z_n/x_1, \dots, x_n]A$, and each y_i is modally free for x_i in $[y_1, \dots, y_{i-1}/x_1, \dots, x_{i-1}]\Box(x_1 = z_1 \land \dots \land x_n = z_n \supset [z_1, \dots, z_n/x_1, \dots, x_n]A)$, i.e. in $\Box(y_1 = z_1 \land \dots \land y_{i-1} = z_{i-1} \land x_i = z_i \land \dots \land x_n = z_n \supset [z_1, \dots, z_n/x_1, \dots, x_n]A)$, because x_i and y_i are modally separated in $y_1 = z_1 \land \dots \land y_{i-1} = z_{i-1} \land x_i = z_i \land \dots \land x_n = z_n \supset [z_1, \dots, z_n/x_1, \dots, x_n]A$. ∎

Lemma 4.16 (Substitution-free Universal Instantiation)
For any $\mathcal{L}$-formula A and variables x, y,

(FUI**) $\vdash_L \forall xA \supset (Ey \supset \exists x(x=y \land A))$.

Proof Let z be a variable not in $Var(A), x, y$.

1.	$\vdash_L z=y \supset Ey \supset Ez$	(LL*)
2.	$\vdash_L \forall xA \supset Ez \supset [z/x]A$	((FUI*), $z \notin Var(A)$)
3.	$\vdash_L \forall xA \land Ey \supset z=y \supset [z/x]A$	(1, 2)
4.	$\vdash_L \forall x(x=z \supset \neg A) \supset Ez \supset (z=z \supset [z/x]\neg A)$	((FUI*), $z \notin Var(A)$)
5.	$\vdash_L Ez \supset z=z$	((=R), or (∀=R), (FUI*))
6.	$\vdash_L \forall x(x=z \supset \neg A) \supset Ez \supset [z/x]\neg A$	(4, 5)
7.	$\vdash_L Ez \supset [z/x]A \supset \exists x(x=z \land A)$	(6)
8.	$\vdash_L \forall xA \land Ey \supset z=y \supset \exists x(x=z \land A)$	(1, 3, 7)
9.	$\vdash_L z=y \supset \exists x(x=z \land A) \supset \exists x(x=y \land A)$	((LL*), $z \notin Var(A)$)
10.	$\vdash_L \forall xA \land Ey \supset z=y \supset \exists x(x=y \land A)$	(8, 9)
11.	$\vdash_L \forall z(\forall xA \land Ey) \supset \forall z(z=y \supset \exists x(x=y \land A))$	(10, (UG), (UD))
12.	$\vdash_L \forall xA \land Ey \supset \forall z(z=y \supset \exists x(x=y \land A))$	(11, (VQ))
13.	$\vdash_L \forall z(z=y \supset \exists x(x=y \land A)) \supset y=y \supset \exists x(x=y \land A)$	((FUI*), $z \notin Var(A)$)
14.	$\vdash_L Ey \supset y=y$	((=R), or (∀=R), (FUI*))
15.	$\vdash_L \forall z(z=y \supset \exists x(x=y \land A)) \supset Ey \supset \exists x(x=y \land A)$	(13, 14)
16.	$\vdash_L \forall xA \supset Ey \supset \exists x(x=y \land A)$	(12, 15)

∎

Incidentally, (FUI*) can also be derived from (FUI**), so we could just as well have used (FUI**) as basic axiom instead of (FUI*).

Proof

1. $\vdash_L x{=}y \supset A \supset [y/x]A$ (LL, given y is m.f. in A).
2. $\vdash_L \neg[y/x]A \supset x{=}y \supset \neg A$ (1, PC).
3. $\vdash_L \forall x(\neg[y/x]A \supset x{=}y \supset \neg A)$ (2, UG).
4. $\vdash_L \forall x(\neg[y/x]A) \supset \forall x(x{=}y \supset \neg A)$ (3, UD).
5. $\vdash_L \neg[y/x]A \supset \forall x(\neg[y/x]A)$ (VQ).
6. $\vdash_L \neg[y/x]A \supset \forall x(x{=}y \supset \neg A)$ (4, 5, PC).
7. $\vdash_L \exists x(x{=}y \wedge A) \supset [y/x]A$ (6, PC).
8. $\vdash_L \forall xA \supset Ey \supset \exists x(y{=}x \wedge A)$ (FUI*).
9. $\vdash_L \forall xA \supset Ey \supset [y/x]A$ (7,8).

■

Closure under first-order substitution may seem redundant for logics given by axiom schemas. The condition is still non-empty for extensions of the logic, e.g. by a further axiom Fy. Moreover, I don't think it actually is redundant. E.g. with (Sub), we can infer $\vdash x{=}y \supset Gxx \supset Gxy$ from the (LL) instance $\vdash v{=}y \supset Gxv \supset Gxy$, but the former isn't itself an (LL) instance. In classical logic, (Sub) is given by (UG) and (UI): we can go from $\vdash A$ to $\forall xA$ to $[y/x]A$. However, in free logics, the last step is restricted to existing y.

The restriction that y is m.f. in A is there to prevent e.g. the move from $\vdash v{=}y \supset \Box x{=}v \supset \Box x{=}y$ to $\vdash x{=}y \supset \Box x{=}x \supset \Box x{=}y$.

We don't have closure under second-order substitution. E.g. negative free logic is not closed under substitution of complex formulas for atomic predicates because $Fx \supset Ex$ is valid, but $\neg Fx \supset Ex$ is not.

5 Logics with explicit substitution

Let's move on to languages with substitution. We first have to lay down some axioms governing the substitution operator. An obvious suggestion would be the lambda-conversion principle

$$\langle y : x \rangle A \leftrightarrow [y/x]A,$$

which would allow us to move back and forth between e.g. $\langle y : x \rangle Fx$ and Fy. But we've seen in lemma 3.9 that if things can have multiple counterparts, then these transitions are sound only under certain conditions: the move from $\langle y : x \rangle A$ to $[y/x]A$ requires that y is modally free for x in A, the other direction requires that y and x are modally separated in A. So we have the following somewhat more complex principles:

(SC1) $\langle y : x \rangle A \leftrightarrow [y/x]A$, provided y and x are modally separated in A.

(SC2) $\langle y : x\rangle A \supset [y/x]A$, provided y is modally free for x in A.

But now we need further principles telling us how $\langle y : x\rangle$ behaves when y is not modally free for x. For example, $\langle y : x\rangle\neg A$ should always entail $\neg\langle y : x\rangle A$, even if y is not modally free for x in A. More generally, the substitution operator commutes with every non-modal operator as long as there is no clash of bound variables:

(S¬) $\langle y : x\rangle\neg A \leftrightarrow \neg\langle y : x\rangle A$,

(S⊃) $\langle y : x\rangle(A \supset B) \leftrightarrow (\langle y : x\rangle A \supset \langle y : x\rangle B)$,

(S∀) $\langle y : x\rangle\forall zA \leftrightarrow \forall z\langle y : x\rangle A$, provided $z \notin \{x, y\}$,

(SS1) $\langle y : x\rangle\langle y_2 : z\rangle A \leftrightarrow \langle y_2 : z\rangle\langle y : x\rangle A$, provided $z \notin \{x, y\}$ and $y_2 \neq x$.

Consider the different ways for x, y, z in (S∀) to not be distinct:

1. $x = y = z$. $\langle x : x\rangle\forall xA$ *is* equivalent to $\forall x\langle x : x\rangle A$: they both say the same as $\forall xA$. We don't need to allow for this instance here because it is provable from $\langle x : x\rangle A \leftrightarrow A$ ((SE1) below) via $\forall x\langle x : x\rangle A \leftrightarrow \forall xA$ ((UG), (UD)) and $\forall xA \leftrightarrow \langle x : x\rangle\forall xA$ ((VS)).
2. $x = y \neq z$. $\langle x : x\rangle\forall zA$ *is* equivalent to $\forall z\langle x : x\rangle A$.
3. $x = z \neq y$. $\langle y : x\rangle\forall xA$ *is not* equivalent to $\forall x\langle y : x\rangle A$. Rather, it is equivalent to $\forall xA$ (and provably so, by (VS)), while $\forall x\langle y : x\rangle A$ is equivalent to $\langle y : x\rangle A$ (and provably so, by (VQ), (FUI_s) and (VS)).
4. $x \neq y = z$. $\langle y : x\rangle\forall yA$ *is not* equivalent to $\forall y\langle y : x\rangle A$. The former says that A holds under all x, y-variants V' that map y into D_w and x to the original value $V(y)$ of y. The latter says that A holds under all x, y-variants V' that map x and y to the same member of D_w.

(SS) is a bit more complicated. $\langle y : x\rangle$ obviously commutes with $\langle y_2 : z\rangle$ if all four variables are distinct. Moreover, it wouldn't matter if $x = y$ or $y_2 = z$. Potential problems only arise if the two substitution terms share a variable. Then again, it wouldn't matter if the unbound variables y and y_2 were identical: $\langle y : x\rangle\langle y : z\rangle A \leftrightarrow \langle y : z\rangle\langle y : x\rangle A$. So the following condition is sufficient for two substitutions to commute: none of x, y is identical to any of z, y_2 except perhaps for y and y_2. Equivalently: $z \notin \{x, y\}$ and $y_2 \neq x$ (equivalently: $x \notin \{z, y_2\}$ and $y \neq z$). Note that under this condition, the LTR direction of the principle would be enough, since the RTL direction is just an instance of the LTR direction.

Among the remaining cases, it is clear that $\langle y : x\rangle\langle y_2 : z\rangle A \leftrightarrow \langle y : z\rangle\langle y_2 : x\rangle A$ is generally invalid if the two bound variable x and z are the same, since the second quantifier would trump the first concerning the interpretation of x/z. The only exceptions are if $y = y_2$, or either $y = x$ or $y_2 = x$. In the first case, the two sides of the biconditional are identical, so the principle is an instance of (Taut). In the second two cases, one of the two substitution terms is the trivial substitution $\langle x : x\rangle$, which does indeed commute with everything.

This leaves the case where the bound variable in one term is identical to the unbound variable in the other term, as in $\langle y : x\rangle\langle x : z\rangle$ or $\langle z : x\rangle\langle y_2 : z\rangle$. Clearly, we cannot go from $\langle y : x\rangle\langle x : z\rangle A$ (which says that A holds if x and z are mapped to y) to $\langle x : z\rangle\langle y : x\rangle A$ (which says that A holds if x is mapped to y but z is mapped to the original x), unless $x = y$. Nor can we go in the other direction from $\langle x : z\rangle\langle y : x\rangle A$ (which is of the form $\langle z : x\rangle\langle y_2 : z\rangle A$) to $\langle y : x\rangle\langle x : z\rangle A$ unless $x = y$. In the special case where $x = y$, one of the two substitution terms becomes trivial; the two directions reduce to $\langle x : x\rangle\langle x : z\rangle A \leftrightarrow \langle x : z\rangle\langle x : x\rangle A$.

Here then is the necessary and sufficient proviso for $\langle y : x\rangle\langle y_2 : z\rangle A \supset \langle y_2 : z\rangle\langle y : x\rangle A$ to be valid:

(i) $x \neq z$ (unless $y = y_2$, in which case the principle is a tautology, or $y = z$ or $y_2 = z$, in which case one of the two substitution terms is trivial), and

(ii) $y_2 \neq x$ (unless $y_2 = x = y$, in which case the first substitution term on the left is trivial), and

(iii) $y \neq z$ (unless $y = z = y_2$, in which case the other substitution term is trivial).

Unsurprisingly, these conditions are symmetrical: if they are fulfilled for $\langle y : x\rangle\langle y_2 : z\rangle A \supset \langle y_2 : z\rangle\langle y : x\rangle A$, then they are also fulfilled for its converse.

Ignoring the parentheses, we are back to the sufficient condition above: $z \notin \{x, y\}$ and $y_2 \neq x$ (equivalently: $x \notin \{z, y_2\}$ and $y \neq z$).

However, this principle can't be enough. In our language, free variables can occur at two places: as predicate arguments in atomic formulas, or in the first position of a substitution operator. (SA) tells us that $\langle y : x\rangle$ behaves like substitution on atomic formulas. We need something that tells us it behaves like this also on free variables in substitution operators.

So now consider $\langle x : z\rangle A$ as analogous to Fx. We want to say that $\langle y : x\rangle\langle x : z\rangle A$ says the same things as $\langle y : z\rangle A$ – or rather, because x may still occur freely in A, as $\langle y : z\rangle\langle y : x\rangle A$. This amounts to a special kind of substitution commutation, with exchange of free variables: $\langle y : x\rangle\langle y_2 : z\rangle A \leftrightarrow \langle [y/x]y_2 : z\rangle\langle y : x\rangle A$. This is an asymmetrical principle, so we might regard it as two principles, one from LTR, the other from RTL. If $y_2 \neq x$, then the two directions coincide, and we're back to the simple commutation principle above. Hence the version with substitution entails the simple commutation version. – Recall that $y_2 \neq x$ was one of the preconditions for the simple commutation principle, so our new principle merely adds something to the old one that covers this previously excluded case. (Strictly speaking, the precondition was that $y_2 \neq x$ unless $y_2 = x = y$; if $y_2 = x = y$, then $[y/x]y_2 = y_2$, so the new version won't contradict the old one here.)

We still need something like provisos (i) and (iii). If the two bound variables x and z are the same, we get $\langle y : x\rangle\langle y_2 : x\rangle A \leftrightarrow \langle [y/x]y_2 : x\rangle\langle y : x\rangle A$. The antecedent says that A is true if x is mapped to $[y/x]y_2$ (i.e. to y if $y_2 = x$, else to y_2). The consequent says that A is true if x is mapped to $[[y/x]y_2/x]y$ (i.e. to y if $y = x$ and $y_2 = x$, or to y_2 if $y = x$ and $y_2 \neq x$, or else to y if $y \neq x$ – more simply: to y_2 if $y = x$, else to y). The two are equivalent iff $[y/x]y_2 = [[y/x]y_2/x]y$, i.e. iff $(y_2 = x\ ?\ y : y_2) = (y = x\ ?\ y_2 : y)$, i.e. iff either $y_2 = x$ and $y = x$ (and, trivially, $y = y_2$), or $y_2 = x$ and $y \neq x$ (and $y_2 = y_2$), or $y_2 \neq x$ and $y = x$ (and $y_2 = y_2$), or $y_2 \neq x$ and $y \neq x$ and $y_2 = y$, i.e. iff $y_2 = x$ or $y_2 = y$ or $x = y$. These three cases are, respectively, $\langle y : x\rangle\langle x : x\rangle A \leftrightarrow \langle y : x\rangle\langle y : x\rangle A$, $\langle y : x\rangle\langle y : x\rangle A \leftrightarrow \langle y : x\rangle\langle y : x\rangle A$, and $\langle x : x\rangle\langle y_2 : x\rangle A \leftrightarrow \langle y_2 : x\rangle\langle x : x\rangle A$.

As to cases where the bound variable in one term is identical to the unbound variable in the other term, we have several cases, depending on whether we look at the terms on the left or on the right. The first case is where $x = y_2$. We then have $\langle y : x\rangle\langle x : z\rangle A \leftrightarrow \langle [y/x]x : z\rangle\langle y : x\rangle A$. Both sides say that A is true if x and z are mapped to y, so this is now valid. The next case is where $y = z$. We then have $\langle y : x\rangle\langle y_2 : y\rangle A \leftrightarrow \langle [y/x]y_2 : y\rangle\langle y : x\rangle A$. The left-hand side says that A is true if y is mapped to $[y/x]y_2$ and x to the original y (unless $x = y$). The right-hand side says that A is true if x and y are mapped to $[y/x]y_2$. The two coincide iff $y = [y/x]y_2$, i.e. iff $y_2 = y$ or $y_2 = x$. So we note: if $y = z$, then $y_2 = y$ or $y_2 = x$. Finally, there's the case where

$[y/x]y_2 = x$. Then $y_2 = x = y$. This is a subtype of the first case above.

So here's the complete proviso for (SS):

(i) $x \neq z$ (unless $y_2 = y$, in which case the principle is a tautology, or $y_2 = x$, in which case it says that $\langle y : x\rangle\langle x : x\rangle A \leftrightarrow \langle y : x\rangle\langle y : x\rangle A$, or $x = y$, in which case it says that $\langle x : x\rangle\langle y_2 : x\rangle A \leftrightarrow \langle y_2 : x\rangle\langle x : x\rangle A$), and

(ii) $y \neq z$ (unless $y_2 = y$, in which case the principle says that $\langle y : x\rangle\langle y : y\rangle A \leftrightarrow \langle y : y\rangle\langle y : x\rangle A$, or $y_2 = x$, in which case it says that $\langle y : x\rangle\langle x : y\rangle A \leftrightarrow \langle y : y\rangle\langle y : x\rangle A$).

Ignoring the parentheses, we have $z \notin \{x, y\}$. (This is asymmetrical, because the principle itself is asymmetrical.) But some of the parenthetical cases have to be added, as we'll see later. Note that the two parenthetical cases in (ii) also occur in (i). So we can pull them out:

1. $z \notin \{x, y\}$, *or*
2. $y_2 \in \{x, y\}$, *or*
3. $z = x = y$.

Equivalently,

1. $z \notin \{x, y\}$, *or*
2. $y_2 \in \{x, z\}$, *or*
3. $z = x = y$.

Substitution does not commute with the box. Roughly speaking, this is because $\langle y : x\rangle\Box A(x, y)$ says that at all accessible worlds, all counterparts x' and y' of y are $A(x', y')$, while $\Box\langle y : x\rangle A(x, y)$ says that at all accessible worlds, every counterpart $x' = y'$ of y is such that $A(x', y')$. In the first case, x' and y' may be different counterparts of y, while in the second case, they must be the same. Thus $\langle y : x\rangle\Box A$ entails $\Box\langle y : x\rangle A$, but the other direction holds only if either y does not have multiple counterparts at accessible worlds (relative to the same counterpart relation), or at most one of x and y occurs freely in A (including the special case where x and y are the same variable).

(S$\Box$) $\langle y : x\rangle\Box A \supset \Box\langle y : x\rangle A$,

(S$\Diamond$) $\langle y : x\rangle\Diamond A \supset \Diamond\langle y : x\rangle A$, provided at most one of x, y is free in A.

These principles largely make (SC1) and (SC2) redundant. We only need to add the special case for substituting free variables in atomic formulas and in substitution operators, as well as a principle for vacuous substitutions:

(SAt) $\langle y : x\rangle Px_1 \ldots x_n \leftrightarrow P[y/x]x_1 \ldots [y/x]x_n$.

(SS2) $\langle y : x\rangle\langle x : z\rangle A \leftrightarrow \langle y : z\rangle\langle y : x\rangle A$.

(VS) $A \leftrightarrow \langle y : x\rangle A$, provided x is not free in A.

Apart from the box axiom, my axioms are also valid for lambda substitution, but their necessitation isn't.

Lemma 5.1 (Soundness of the substitution axioms)
If $\mathcal{L}_s$ is a language of quantified modal logic with substitution, then every $\mathcal{L}_s$-instance of (S$\neg$), (S$\supset$), (S$\forall$), (SS1), (S$\Box$), (S$\Diamond$), (SAt), (SS2), and (VS) is valid in every (positive or negative) counterpart model.

Proof

1. (S$\neg$). $w, V \Vdash \langle y : x\rangle\neg A$ iff $w, V^{[y/x]} \Vdash \neg A$ by definition 3.2, iff $w, V^{[y/x]} \nVdash A$ by definition 2.7, iff $w, V \nVdash \langle y : x\rangle A$ by definition 3.2, iff $w, V \Vdash \neg\langle y : x\rangle A$ by definition 2.7.

2. (S$\supset$). $w, V \Vdash \langle y : x\rangle(A \supset B)$ iff $w, V^{[y/x]} \Vdash A \supset B$ by definition 3.2, iff $w, V^{[y/x]} \nVdash A$ or $w, V^{[y/x]} \Vdash B$ by definition 2.7, iff $w, V \nVdash \langle y : x\rangle A$ or $w, V \Vdash \langle y : x\rangle B$ by definition 3.2, iff $w, V \Vdash \langle y : x\rangle A \supset \langle y : x\rangle B$ by definition 2.7.

3. (S$\forall$). Assume $z \notin \{x, y\}$. Then the existential z-variants V' of $V^{[y/x]}$ on w coincide at w with the functions $(V^*)^{[y/x]}$ where V^* is an existential z-variant V^* of V on w. And so $w, V \Vdash \langle y : x\rangle\forall z A$ iff $w, V^{[y/x]} \Vdash \forall z A$ by definition 3.2, iff $w, V' \Vdash A$ for all existential z-variants V' of $V^{[y/x]}$ on w by definition 2.7, iff $w, (V^*)^{[y/x]} \Vdash A$ for all existential z-variants V^* of V on w, iff $w, V^* \Vdash \langle y : x\rangle A$ for all existential z-variants V^* of V on w by definition 3.2, iff $w, V \Vdash \forall z\langle y : x\rangle A$ by definition 2.7.

4. (SS1). Assume $z \notin \{x, y\}$ and $y_2 \neq x$. Then the function $[y/x] \cdot [y_2/z]$ is identical to the function $[y_2/z] \cdot [y/x]$. So $w, V \Vdash \langle y : x\rangle\langle y_2 : z\rangle A$ iff $w, V^{[y/x]\cdot[y_2/z]} \Vdash A$ by definition 3.2, iff $w, V^{[y_2/z]\cdot[y/x]} \Vdash A$, iff $w, V \Vdash \langle y_2 : z\rangle\langle y : x\rangle A$ by definition 3.2.

5. (S$\Box$). Assume $w, V \nVdash \Box\langle y : x\rangle A$. By definitions 2.7 and 3.2, this means that $w', V'^{[y/x]} \nVdash A$ for some w', V' such that wRw' and $V_w \triangleright V'_{w'}$, i.e. there is a $C \in K_{w,w'}$ for which $V'_{w'}$ assigns to every variable z a C-counterpart of its value under V_w (or nothing if there is none). Then for all z, $V'^{[y/x]}_{w'}(z)$ is a C-counterpart of $V^{[y/x]}_w(z)$ (or undefined if there is none), since $V'^{[y/x]}_{w'}(x) = V'_{w'}(y)$ is a C-counterpart of $V_w(y) = V^{[y/x]}_w(x)$ (or undefined if there is none). So $V^{[y/x]}_w \triangleright V'^{[y/x]}_{w'}$. And so $w', V^* \nVdash A$ for some w', V^* such that wRw' and $V^{[y/x]}_w \triangleright V^*_{w'}$. So $w, V \Vdash \langle y : x\rangle\Box A$ by definitions 2.7 and 3.2.

6. (S$\Diamond$). Assume $w, V \Vdash \langle y : x\rangle\Diamond A$ and at most one of x, y is free in A. By definitions 2.7 and 3.2, $w', V^* \Vdash A$ for some w', V^* such that wRw' and $V^{[y/x]}_w \triangleright V^*_{w'}$, i.e. there is a $C \in K_{w,w'}$ for which $V^*_{w'}$ assigns to every variable z a C-counterpart of its value under V_w (or nothing if there is none). We have to show that there is a w'-image V' of V at w such that $w, V'^{[y/x]} \Vdash A$, since then $w, V \Vdash \Diamond\langle y : x\rangle A$.

   If x is the same variable as y, then $V^*_{w'}(x) = V^*_{w'}(y)$ is a C-counterpart at w' of $V^{[y/x]}_w(x) = V^{[y/x]}_w(y) = V_w(x) = V_w(y)$ at w (or undefined if there is none), so we can choose V^* itself as V'. We then have $w, V'^{[y/x]} \Vdash A$ because $V'^{[y/x]} = V'$.

   Else if x is not free in A, let V' be some x-variant of V^* at w' such that $V^*_{w'}(x)$ is some C-counterpart at w' of $V_w(x)$ at w (or undefined if there is none). Since $V^*_{w'}(y)$ is a

C-counterpart at w' of $V_w^{[y/x]}(y) = V_w(y)$ at w (or undefined if there is none), V' is a w'-image of V at w. Moreover, $V'^{[y/x]}$ and V^* agree at w' about all variables other than x; so by the coincidence lemma 2.10, $w', V'^{[y/x]} \Vdash A$.

Else if y is not free in A, let V' be like V^* except that $V'_{w'}(y) = V^*_{w'}(x)$ and $V'_{w'}(x)$ is some C-counterpart at w' of $V_w(x)$ at w (or undefined if there is none). Since $V'_{w'}(y) = V^*_{w'}(x)$ is a C-counterpart at w' of $V_w^{[y/x]}(x) = V_w(y)$ at w (or undefined if there is none), V' is a w'-image of V at w. Moreover, $V'^{[y/x]}$ and V^* agree at w' about all variables other than y; in particular, $V'^{[y/x]}_{w'}(x) = V'_{w'}(y) = V^*_{w'}(x)$. So by the coincidence lemma 2.10, $w', V'^{[y/x]} \Vdash A$.

7. (SAt). $w, V \Vdash \langle y : x \rangle P x_1 \ldots x_n$ iff $w, V^{[y/x]} \Vdash P x_1 \ldots x_n$ by definition 3.2, iff $w, V \Vdash [y/x] P x_1 \ldots x_n$ by lemma 3.9.
8. (SS2). $w, V \Vdash \langle y : x \rangle \langle x : z \rangle A$ iff $w, V^{[y/x]\cdot[x/z]} \Vdash A$ by definition 3.2, iff $w, V^{[y/z]\cdot[y/x]} \Vdash A$ because $[y/x] \cdot [x/z] = [y/z] \cdot [y/x]$, iff $w, V \Vdash \langle y : z \rangle \langle y : x \rangle A$ by definition 3.2.
9. (VS). By definition 3.2, $w, V \Vdash \langle y : x \rangle A$ iff $w, V^{[y/x]} \Vdash A$. If x is not free in A, then $V^{[y/x]}$ agrees with V at w about all free variables in A. So by the coincidence lemma 2.10, $w, V^{[y/x]} \Vdash A$ iff $w, V \Vdash A$. So then $w, V \Vdash \langle y : x \rangle A$ iff $w, V \Vdash A$. ■

DEFINITION 5.2 (POSITIVE LOGICS WITH SUBSTITUTION)
Given a language $\mathcal{L}_s$ with substitution, a *positive (quantified modal) logic with substitution* in $\mathcal{L}_s$ is a set of formulas $L \subseteq \mathcal{L}_s$ that contains all $\mathcal{L}_s$-instances of the substitution axioms (S¬), (S⊃), (S∀), (SS1), (S□), (S◇), (SAt), (SS2), (VS), as well as (Taut), (UD), (VQ), (∀Ex), (=R), (K),

(FUI$_s$) $\forall x A \supset (E y \supset \langle y : x \rangle A)$,

(LL$_s$) $x = y \supset (A \supset \langle y : x \rangle A)$,

and that is closed under (MP), (UG), (Nec) and

(Sub$_s$) if $\vdash_L A$, then $\vdash_L \langle y : x \rangle A$.

The smallest such logic is called P_s.

DEFINITION 5.3 (NEGATIVE LOGICS WITH SUBSTITUTION)
Given a language $\mathcal{L}_s$ with substitution, a *negative (quantified modal) logic with substitution* in $\mathcal{L}_s$ is a set $L \subseteq \mathcal{L}_s$ that contains all $\mathcal{L}_s$-instances of the substitution axioms (S¬), (S⊃), (S∀), (SS1), (S□), (S◇), (SAt), (SS2), (VS), as well as (Taut), (UD), (VQ), (Neg), (NA), (∀=R), (K), (FUI$_s$), (LL$_s$), and that is closed under (MP), (UG), (Nec) and (Sub$_s$). The smallest such logic is called N_s.

Theorem 5.4 (Soundness of P_s)

Every member of P_s is valid in every positive counterpart model.

Proof We have to show that all P_s axioms are valid in every model, and that validity is closed under (MP), (UG), (Nec) and ($\mathrm{Sub_s}$). For (Taut), (UD), (VQ), (∀Ex), (=R), (K), (MP), (UG), (Nec), see the proof of theorem 4.3. For the substitution axioms, see lemma 5.1. The remaining cases are ($\mathrm{FUI_s}$), ($\mathrm{LL_s}$), and ($\mathrm{Sub_s}$).

1. ($\mathrm{FUI_s}$). Assume $w, V \Vdash \forall x A$ and $w, V \Vdash Ey$ in some model. By definition 2.7, the latter means that $V_w(y) \in D_w$, and the former means that $w, V' \Vdash A$ for all existential x-variants V' of V on w. So in particular, $w, V' \Vdash A$, where V' is the x-variant of V on w with $V_w(x) = V_w(y)$. So $w, V \Vdash \langle y : x \rangle A$ by definition 3.2.
2. ($\mathrm{LL_s}$). Assume $w, V \Vdash x = y$ and $w, V \Vdash A$. By definitions 2.7 and 2.3, then $V_w(x) = V_w(y)$. So $w, V \Vdash \langle y : x \rangle A$ by definition 3.2.
3. ($\mathrm{Sub_s}$). Assume $w, V \nVdash \langle y : x \rangle A$ in some model $\mathcal{M} = \langle \mathcal{S}, V \rangle$. By definition 3.2, then $w, V' \nVdash A$, where V' is the x-variant of V on w with $V'(x) = V(y)$. So A is invalid in the model $\langle \mathcal{S}, V' \rangle$. Hence if A is valid in all positive models, then so is $\langle y : x \rangle A$. ■

Theorem 5.5 (Soundness of N_s)

Every member of N_s is valid in every negative counterpart model.

Proof All the cases needed here are covered in the proofs of theorem 4.6 and 5.4. ■

To derive some further properties of these systems, let $\mathcal{L}$ range over languages of quantified modal logic with substitution, and L over positive or negative logics in $\mathcal{L}$.

Closure under propositional consequence and the validity of (∀Ex) and (∀=R) are proved just as for substitution-free logics (see lemmas 4.7 and 4.8). So we move on immediately to more interesting properties.

Lemma 5.6 (Substitution expansion)

If A is an $\mathcal{L}$-formula and x, y, z $\mathcal{L}$-variables, then

(SE1) $\vdash_L A \leftrightarrow \langle x : x \rangle A$;

(SE2) $\vdash_L \langle y : x \rangle A \leftrightarrow \langle y : z \rangle \langle z : x \rangle A$, provided z is not free in A.

Proof (SE1) is proved by induction on A.

1. A is atomic. Then $\vdash_L \langle x : x\rangle A \leftrightarrow [x/x]A$ by (SAt), and so $\vdash_L \langle x : x\rangle A \leftrightarrow A$ because $[x/x]A = A$.

2. A is $\neg B$. By induction hypothesis, $\vdash_L B \leftrightarrow \langle x : x\rangle B$. So by (PC), $\vdash_L \neg B \leftrightarrow \neg\langle x : x\rangle B$. And by $\langle S\neg\rangle$, $\vdash_L \langle x : x\rangle\neg B \leftrightarrow \neg\langle x : x\rangle B$.

3. A is $B \supset C$. By induction hypothesis, $\vdash_L B \leftrightarrow \langle x : x\rangle B$ and $\vdash_L C \leftrightarrow \langle x : x\rangle C$. So $\vdash_L (B \supset C) \leftrightarrow (\langle x : x\rangle B \supset \langle x : x\rangle C)$. And by $\langle S \supset\ \rangle$, $\vdash_L \langle x : x\rangle(B \supset C) \leftrightarrow (\langle x : x\rangle B \supset \langle x : x\rangle C)$.

4. A is $\forall zB$. If $z = x$, then $\vdash_L \forall xB \leftrightarrow \langle x : x\rangle\forall xB$ by (VS). If $z \neq x$, then by induction hypothesis, $\vdash_L B \leftrightarrow \langle x : x\rangle B$; by (UG) and (UD), $\vdash_L \forall zB \leftrightarrow \forall z\langle x : x\rangle B$; and $\vdash_L \langle x : x\rangle\forall zB \leftrightarrow \forall z\langle x : x\rangle B$ by (S$\forall$).

5. A is $\langle y : z\rangle B$. If $z = x$, then $\vdash_L \langle y : x\rangle B \leftrightarrow \langle x : x\rangle\langle y : x\rangle B$ by (VS). If $z \neq x$, then by induction hypothesis, $\vdash_L B \leftrightarrow \langle x : x\rangle B$; by (Sub$_s$) and (S$\supset$), $\vdash_L \langle y : z\rangle B \leftrightarrow \langle y : z\rangle\langle x : x\rangle B$; and $\vdash_L \langle x : x\rangle\langle y : z\rangle B \leftrightarrow \langle y : z\rangle\langle x : x\rangle B$ by (SS1) (if $y \neq x$) or (SS2) (if $y = x$).

6. A is $\Box B$. By (S$\Box$), $\vdash_L \langle x : x\rangle\Box B \supset \Box\langle x : x\rangle B$. Conversely, since at most one of x, x is free in $\neg B$, by (S$\Diamond$), $\vdash_L \langle x : x\rangle\Diamond\neg B \supset \Diamond\langle x : x\rangle\neg B$. Contraposing and unraveling the definition of the diamond, we have $\vdash_L \Box\neg\langle x : x\rangle\neg B \supset \neg\langle x : x\rangle\neg\Box\neg\neg B$. Since $\vdash_L \Box\neg\langle x : x\rangle\neg B \leftrightarrow \Box\langle x : x\rangle B$ and $\vdash_L \neg\langle x : x\rangle\neg\Box\neg\neg B \leftrightarrow \langle x : x\rangle B$ (by (S$\neg$), (Sub$_s$), (S$\supset$), (Nec) and (K)), this means that $\vdash_L \Box\langle x : x\rangle B \supset \langle x : x\rangle\Box B$.

As for (SE2): by (VQ), $\vdash_L \langle y : x\rangle A \leftrightarrow \langle y : z\rangle\langle y : x\rangle A$. And $\vdash_L \langle y : x\rangle\langle y : z\rangle A \leftrightarrow \langle y : z\rangle\langle y : x\rangle A$ by (SS1) (if $y \neq x$) or (SS2) (if $y = x$). Moreover, by (SS2), $\vdash_L \langle y : z\rangle\langle z : x\rangle A \leftrightarrow \langle y : x\rangle\langle y : z\rangle A$. So by (PC), $\vdash_L \langle y : x\rangle A \leftrightarrow \langle y : z\rangle\langle z : x\rangle A$. ∎

Lemma 5.7 (Substituting bound variables)
For any $\mathcal{L}$-sentence A and variables x, y,

(SBV) $\vdash_L \forall xA \leftrightarrow \forall y\langle y : x\rangle A$, provided y is not free in A.

Proof

1. $\vdash_L \forall y \langle y:x\rangle A \supset Ex \supset \langle x:y\rangle\langle y:x\rangle A.$ (FUI$_s$)
2. $\vdash_L \langle x:y\rangle\langle y:x\rangle A \leftrightarrow A.$ ((SE1), (SE2))
3. $\vdash_L \forall x\forall y\langle y:x\rangle A \supset \forall xEx \supset \forall xA.$ (1, 2, (UG), (UD))
4. $\vdash_L \forall x\forall y\langle y:x\rangle A \supset \forall xA.$ (3, (∀Ex))
5. $\vdash_L \forall y\langle y:x\rangle A \supset \forall x\forall y\langle y:x\rangle A.$ (VQ)
6. $\vdash_L \forall y\langle y:x\rangle A \supset \forall xA.$ (4, 5)
7. $\vdash_L \forall xA \supset Ey \supset \langle y:x\rangle A.$ (FUI$_s$)
8. $\vdash_L \forall y\forall xA \supset \forall y\langle y:x\rangle A.$ (7, (UG), (UD), (∀Ex))
9. $\vdash_L \forall xA \supset \forall y\forall xA.$ ((VQ), y not free in A)
10. $\vdash_L \forall xA \supset \forall y\langle y:x\rangle A.$ (8, 9)
11. $\vdash_L \forall xA \leftrightarrow \forall y\langle y:x\rangle A.$ (6, 10)

■

Lemma 5.8 (Substituting empty variables)
For any $\mathcal{L}$-sentence A and variables x, y,

(SEV) $\vdash_L x\neq x \wedge y\neq y \supset (A \leftrightarrow \langle y:x\rangle A).$

Proof (SEV) is trivial if L is positive, in which case $\vdash_L x=x$. For negative L, it is proved by induction on A.

1. A is atomic. If $x \notin Var(A)$, then $\vdash_L A \leftrightarrow \langle y:x\rangle A$ by (VS), and so $\vdash_L x\neq x \wedge y\neq y \supset (A \leftrightarrow \langle y:x\rangle A)$ by (PC). If $x \in Var(A)$, then by (Neg)

   (1) $\vdash_L x\neq x \wedge y\neq y \supset \neg A.$

   Also by (Neg), $\vdash_L x\neq x \wedge y\neq y \supset \neg[y/x]A$. By (SAt), $\vdash_L [y/x]A \leftrightarrow \langle y:x\rangle A$, and so $\vdash_L \neg[y/x]A \leftrightarrow \neg\langle y:x\rangle A$. So

   (2) $\vdash_L x\neq x \wedge y\neq y \supset \neg\langle y:x\rangle A.$

   Combining (1) and (2) yields $\vdash_L x\neq x \wedge y\neq y \supset (A \leftrightarrow \langle y:x\rangle A)$.

2. A is $\neg B$. By induction hypothesis, $\vdash_L x\neq x \wedge y\neq y \supset (B \leftrightarrow \langle y:x\rangle B)$. So by (PC), $\vdash_L x\neq x\wedge y\neq y \supset (\neg B \leftrightarrow \neg\langle y:x\rangle B)$, and by (S¬), $\vdash_L x\neq x\wedge y\neq y \supset (\neg B \leftrightarrow \langle y:x\rangle\neg B)$.

3. A is $B \supset C$. By induction hypothesis, $\vdash_L x\neq x \wedge y\neq y \supset (B \leftrightarrow \langle y:x\rangle B)$ and $\vdash_L x\neq x \wedge y\neq y \supset (C \leftrightarrow \langle y:x\rangle C)$. So by (PC), $\vdash_L x\neq x \wedge y\neq y \supset ((B \supset C) \leftrightarrow (\langle y:x\rangle B \supset \langle y:x\rangle C))$, and by (S⊃), $\vdash_L x\neq x \wedge y\neq y \supset ((B \supset C) \leftrightarrow \langle y:x\rangle(B \supset C))$.

4. A is $\forall zB$. We distinguish three cases.

a) $z \notin \{x, y\}$. Then

1. $\vdash_L x \neq x \wedge y \neq y \supset (B \leftrightarrow \langle y : x \rangle B)$ (ind. hyp.)
2. $\vdash_L \forall z\, x \neq x \wedge \forall z\, y \neq y \supset (\forall z B \leftrightarrow \forall z \langle y : x \rangle B)$ (1, UG, UD)
3. $\vdash_L x \neq x \wedge y \neq y \supset (\forall z B \leftrightarrow \forall z \langle y : x \rangle B)$ (2, VQ)
4. $\vdash_L x \neq x \wedge y \neq y \supset (\forall z B \leftrightarrow \langle y : x \rangle \forall z B)$. (3, (S$\forall$))

b) $z = x$. Then A is $\forall x B$, and $\vdash_L \forall x B \leftrightarrow \langle y : x \rangle \forall x B$ by (VS). So $\vdash_L x \neq x \wedge y \neq y \supset (\forall x B \leftrightarrow \langle y : x \rangle \forall x B)$ by (PC).

c) $z = y \neq x$. Then A is $\forall y B$. Let v be a variable not in $\mathit{Var}(A), x, y$.

1. $\vdash_L x \neq x \wedge v \neq v \supset (B \leftrightarrow \langle v : x \rangle B)$. (ind. hyp.)
2. $\vdash_L \forall y x \neq x \wedge \forall y v \neq v \supset (\forall y B \leftrightarrow \forall y \langle v : x \rangle B)$. (1, UG, UD)
3. $\vdash_L x \neq x \wedge v \neq v \supset (\forall y B \leftrightarrow \forall y \langle v : x \rangle B)$. (2, VQ)
4. $\vdash_L x \neq x \wedge v \neq v \supset (\forall y B \leftrightarrow \langle v : x \rangle \forall y B)$. (3, (S$\forall$))
5. $\vdash_L \langle y : v \rangle x \neq x \wedge \langle y : v \rangle v \neq v \supset (\langle y : v \rangle \forall y B \leftrightarrow \langle y : v \rangle \langle v : x \rangle \forall y B)$. (4, (Sub$_s$), (S$\supset$))
6. $\vdash_L x \neq x \wedge y \neq y \supset (\langle y : v \rangle \forall y B \leftrightarrow \langle y : v \rangle \langle v : x \rangle \forall y B)$. (5, (VS), (SAt))
7. $\vdash_L x \neq x \wedge y \neq y \supset (\forall y B \leftrightarrow \langle y : v \rangle \langle v : x \rangle \forall y B)$. (6, (VS))
8. $\vdash_L x \neq x \wedge y \neq y \supset (\forall y B \leftrightarrow \langle y : x \rangle \forall y B)$. (7, (SE2))

5. A is $\langle y_2 : z \rangle B$. We have four cases.

a) $z \notin \{x, y\}$ and $y_2 \neq x$. Then

1. $\vdash_L x \neq x \wedge y \neq y \supset (B \leftrightarrow \langle y : x \rangle B)$ (ind. hyp.)
2. $\vdash_L \langle y_2 : z \rangle x \neq x \wedge \langle y_2 : z \rangle y \neq y \supset (\langle y_2 : z \rangle B \leftrightarrow \langle y_2 : z \rangle \langle y : x \rangle B)$ (1, (Sub$_s$), (S$\supset$))
3. $\vdash_L x \neq x \wedge y \neq y \supset (\langle y_2 : z \rangle B \leftrightarrow \langle y_2 : z \rangle \langle y : x \rangle B)$ (2, (VS))
4. $\vdash_L x \neq x \wedge y \neq y \supset (\langle y_2 : z \rangle B \leftrightarrow \langle y : x \rangle \langle y_2 : z \rangle B)$. (3, (SS1))

b) $z \neq x$ and $y_2 = x$. Then A is $\langle x : z \rangle B$. (This case is surprisingly tough.)

1. $\vdash_L x \neq x \wedge z \neq z \supset (B \leftrightarrow \langle x : z \rangle B)$ (ind. hyp.)
2. $\vdash_L \langle y : z \rangle x \neq x \wedge \langle y : z \rangle z \neq z \supset (\langle y : z \rangle B \leftrightarrow \langle y : z \rangle \langle x : z \rangle B)$ (1, (Sub$_s$), (S$\supset$))
3. $\vdash_L x \neq x \wedge y \neq y \supset (\langle y : z \rangle B \leftrightarrow \langle y : z \rangle \langle x : z \rangle B)$ (2, (SAt), $z \neq x$)
4. $\vdash_L x \neq x \wedge y \neq y \supset (\langle y : z \rangle B \leftrightarrow \langle x : z \rangle B)$ (3, (VS), $z \neq x$)
5. $\vdash_L x \neq x \wedge y \neq y \supset (B \leftrightarrow \langle y : x \rangle B)$ (ind. hyp.)
6. $\vdash_L \langle y : z \rangle x \neq x \wedge \langle y : z \rangle y \neq y \supset (\langle y : z \rangle B \leftrightarrow \langle y : z \rangle \langle y : x \rangle B)$ (5, (Sub$_s$),(S$\supset$))
7. $\vdash_L x \neq x \wedge y \neq y \supset (\langle y : z \rangle B \leftrightarrow \langle y : z \rangle \langle y : x \rangle B)$ (6, (SAt), $z \neq x$)
8. $\vdash_L x \neq x \wedge y \neq y \supset (\langle x : z \rangle B \leftrightarrow \langle y : z \rangle \langle y : x \rangle B)$ (4, 7)
9. $\vdash_L x \neq x \wedge y \neq y \supset (\langle x : z \rangle B \leftrightarrow \langle y : x \rangle \langle x : z \rangle B)$. (8, (SS2))

c) $z = x$. Then A is $\langle y_2 : x \rangle B$, and $\vdash_L \langle y_2 : x \rangle B \leftrightarrow \langle y : x \rangle \langle y_2 : x \rangle B$ by (VS). So $\vdash_L x \neq x \wedge y \neq y \supset (\langle y_2 : x \rangle B \leftrightarrow \langle y : x \rangle \langle y_2 : x \rangle B)$ by (PC).

d) $z = y \neq x$ and $y_2 \neq x$. Then A is $\langle y_2 : y\rangle B$. Let v be a variable not in $Var(A), x, y, y_2$.

1. $\vdash_L x \neq x \wedge v \neq v \supset (B \leftrightarrow \langle v : x\rangle B)$. (ind. hyp.)
2. $\vdash_L \langle y_2 : y\rangle x \neq x \wedge \langle y_2 : y\rangle v \neq v \supset (\langle y_2 : y\rangle B \leftrightarrow \langle y_2 : y\rangle\langle v : x\rangle B)$. (1, (Sub$_s$), (S$\supset$))
3. $\vdash_L x \neq x \wedge v \neq v \supset (\langle y_2 : y\rangle B \leftrightarrow \langle y_2 : y\rangle\langle v : x\rangle B)$. (2, (VS))
4. $\vdash_L x \neq x \wedge v \neq v \supset (\langle y_2 : y\rangle B \leftrightarrow \langle v : x\rangle\langle y_2 : y\rangle B)$. (3, (SS1), $y_2 \neq x$)
5. $\vdash_L \langle y : v\rangle x \neq x \wedge \langle y : v\rangle v \neq v \supset (\langle y : v\rangle\langle y_2 : y\rangle B \leftrightarrow \langle y : v\rangle\langle v : x\rangle\langle y_2 : y\rangle B)$. (4, (Sub$_s$), (S$\supset$))
6. $\vdash_L x \neq x \wedge y \neq y \supset (\langle y : v\rangle\langle y_2 : y\rangle B \leftrightarrow \langle y : v\rangle\langle v : x\rangle\langle y_2 : y\rangle B)$. (5, (VS), (SAt))
7. $\vdash_L x \neq x \wedge y \neq y \supset (\langle y_2 : y\rangle B \leftrightarrow \langle y : v\rangle\langle v : x\rangle\langle y_2 : y\rangle B)$. (6, (VS))
8. $\vdash_L x \neq x \wedge y \neq y \supset (\langle y_2 : y\rangle B \leftrightarrow \langle y : x\rangle\langle y_2 : y\rangle B)$. (7, (SE2))

6. A is $\Box B$. Let v be a variable not in $Var(B)$.

1. $\vdash_L x \neq x \wedge v \neq v \supset (B \leftrightarrow \langle v : x\rangle B)$. (ind. hyp.)
2. $\vdash_L \Box(x \neq x \wedge v \neq v) \supset (\Box B \leftrightarrow \Box\langle v : x\rangle B)$. (1, (Nec), (K))
3. $\vdash_L x \neq x \wedge v \neq v \supset \Box(x \neq x \wedge v \neq v)$ ((NA), (EI), (Nec), (K))
4. $\vdash_L x \neq x \wedge v \neq v \supset (\Box B \leftrightarrow \Box\langle v : x\rangle B)$. (2, 3)
5. $\vdash_L x \neq x \wedge v \neq v \supset (\Box B \leftrightarrow \langle v : x\rangle\Box B)$. (4, (S$\Box$), (S$\Diamond$), $v \notin Var(B)$)
6. $\vdash_L \langle y : v\rangle x \neq x \wedge \langle y : v\rangle v \neq v \supset (\langle y : v\rangle\Box B \leftrightarrow \langle y : v\rangle\langle v : x\rangle\Box B)$. (5, (Sub$_s$), (S$\supset$))
7. $\vdash_L x \neq x \wedge y \neq y \supset (\langle y : v\rangle\Box B \leftrightarrow \langle y : v\rangle\langle v : x\rangle\Box B)$. (6, (SAt))
8. $\vdash_L x \neq x \wedge y \neq y \supset (\Box B \leftrightarrow \langle y : v\rangle\langle v : x\rangle\Box B)$. (7, (VS))
9. $\vdash_L x \neq x \wedge y \neq y \supset (\Box B \leftrightarrow \langle y : x\rangle\Box B)$. (8, (SE2))

Now we can prove (SC1) and (SC2). I will also prove that $\langle y : x\rangle A$ and $[y/x]A$ are provably equivalent conditional on $y \neq y$. Compare lemma 3.9 for a (slightly stronger) semantic version of this lemma. Condition (i).(b) of lemma 3.9 says that y does not have multiple counterparts at any accessible world. This can be expressed as $\neg\langle y : x\rangle\Diamond(x = x \wedge y = y \wedge y \neq x)$. To completely mirror lemma 3.9, we should replace the antecedent $y \neq y$ in (SCN) by this condition. (SCN) would then be derivable, because $y \neq y$ entails $\neg\langle y : x\rangle\Diamond(x = x \wedge y = y \wedge y \neq x)$ (interestingly in negative logic, trivially in positive logic.) Maybe I should prove the stronger version of (SCN) for neatness, even if I don't really use it.

LEMMA 5.9 (SUBSTITUTION CONVERSION)
For any $\mathcal{L}$-formula A and variables x, y,

(SC1) $\vdash_L \langle y : x\rangle A \leftrightarrow [y/x]A$, provided y and x are modally separated in A.

(SC2) $\vdash_L \langle y : x\rangle A \supset [y/x]A$, provided y is modally free for x in A.

(SCN) $\vdash_L y \neq y \supset (\langle y : x\rangle A \leftrightarrow [y/x]A)$.

PROOF If x and y are the same variable, then by (SE1), $\vdash_L \langle x : x\rangle A \leftrightarrow [x/x]A$. Assume then that x and y are different variables. We first prove (SC1) and (SC2), by induction on A. Observe that if A is not a box formula $\Box B$, then by definition 3.4, y is modally free for x in A iff y and x are modally separated in A, in which case y and x are also modally separated in any subformula of A.

1. A is atomic. By (SAt), $\vdash_L \langle y : x\rangle A \leftrightarrow [y/x]A$ holds without any restrictions.
2. A is $\neg B$. If y and x are modally separated in A, then by induction hypothesis, $\vdash_L \langle y : x\rangle B \leftrightarrow [y/x]B$. So by (PC), $\vdash_L \neg\langle y : x\rangle B \leftrightarrow \neg[y/x]B$. By (S$\neg$) and definition 3.3, it follows that $\vdash_L \langle y : x\rangle\neg B \leftrightarrow [y/x]\neg B$.
3. A is $B \supset C$. If y and x are modally separated in A, then by induction hypothesis, $\vdash_L \langle y : x\rangle B \leftrightarrow [y/x]B$ and $\vdash_L \langle y : x\rangle C \leftrightarrow [y/x]C$. By (S$\supset$), $\vdash_L \langle y : x\rangle(B \supset C) \leftrightarrow (\langle y : x\rangle B \supset \langle y : x\rangle C)$. So $\vdash_L \langle y : x\rangle(B \supset C) \leftrightarrow ([y/x]B \supset [y/x]C)$, and so $\vdash_L \langle y : x\rangle(B \supset C) \leftrightarrow [y/x](B \supset C)$ by definition 3.3.
4. A is $\forall zB$. We have to distinguish four cases, assuming each time that y and x are modally separated in A.
   a) $z \notin \{x, y\}$. By induction hypothesis, $\vdash_L \langle y : x\rangle B \leftrightarrow [y/x]B$. So by (UG) and (UD), $\vdash_L \forall z\langle y : x\rangle B \leftrightarrow \forall z[y/x]B$. Since $z \notin \{x, y\}$, $\vdash_L \langle y : x\rangle\forall zB \leftrightarrow \forall z\langle y : x\rangle B$ by (S$\forall$), and $\forall z[y/x]B$ is $[y/x]\forall zB$ by definition 3.3; so $\vdash_L \langle y : x\rangle\forall zB \leftrightarrow [y/x]\forall zB$.
   b) $z = y$ and $x \notin \mathit{Varf}(B)$. By definition 3.3, then $[y/x]\forall zB$ is $\forall y[y/x]B$.

1.	$\vdash_L \langle y : x\rangle B \leftrightarrow [y/x]B$.	(induction hypothesis)
2.	$\vdash_L \forall y\langle y : x\rangle B \leftrightarrow \forall y[y/x]B$.	(1, (UG), (UD))
3.	$\vdash_L B \leftrightarrow \langle y : x\rangle B$.	((VS), $x \notin \mathit{Varf}(B)$)
4.	$\vdash_L \forall yB \leftrightarrow \forall y\langle y : x\rangle B$.	(3, (UG), (UD))
5.	$\vdash_L \forall yB \leftrightarrow \langle y : x\rangle\forall yB$.	((VS), $x \notin \mathit{Varf}(B)$)
6.	$\vdash_L \langle y : x\rangle\forall yB \leftrightarrow \forall y[y/x]B$.	(2, 4, 5)

   c) $z = x$ and $y \notin \mathit{Varf}(B)$. By definition 3.3, then $[y/x]\forall zB$ is $\forall y[y/x]B$.

1.	$\vdash_L \langle y : x\rangle B \leftrightarrow [y/x]B$.	(induction hypothesis)
2.	$\vdash_L \forall y\langle y : x\rangle B \leftrightarrow \forall y[y/x]B$.	(1, (UG), (UD))
3.	$\vdash_L \forall xB \leftrightarrow \forall y\langle y : x\rangle B$.	((SBV), $y \notin \mathit{Varf}(B)$)
4.	$\vdash_L \forall xB \leftrightarrow \langle y : x\rangle\forall xB$.	(VS)
5.	$\vdash_L \langle y : x\rangle\forall xB \leftrightarrow \forall y[y/x]B$.	(2, 3, 4)

   d) $z = x$ and $y \in \mathit{Varf}(B)$, or $z = y$ and $x \in \mathit{Varf}(B)$. By definition 3.3, then $[y/x]\forall zB$ is $\forall v[y/x][v/z]B$ for some variable $v \notin \mathit{Var}(B) \cup \{x, y\}$. Since v and z are modally separated in B, by induction hypothesis $\vdash_L \langle v : z\rangle B \leftrightarrow [v/z]B$. So by (UG) and (UD), $\vdash_L \forall v\langle v : z\rangle B \leftrightarrow \forall v[v/z]B$. By (SBV), $\vdash_L \forall zB \leftrightarrow \forall v\langle v : z\rangle B$.

So $\vdash_L \forall zB \leftrightarrow \forall v[v/z]B$. Moreover, as $z \in \{x, y\}$, y and x are modally separated in $[v/z]B$. So by induction hypothesis, $\vdash_L \langle y : x\rangle[v/z]B \leftrightarrow [y/x][v/z]B$. Then

1.	$\vdash_L \forall zB \leftrightarrow \forall v[v/z]B$	(as just shown)
2.	$\vdash_L \langle y : x\rangle\forall zB \leftrightarrow \langle y : x\rangle\forall v[v/z]B$	(1, (Sub^s), (S$\neg$), (S$\supset$))
3.	$\vdash_L \langle y : x\rangle\forall v[v/z]B \leftrightarrow \forall v\langle y : x\rangle[v/z]B$.	(S$\forall$)
4.	$\vdash_L \langle y : x\rangle\forall zB \leftrightarrow \forall v\langle y : x\rangle[v/z]B$.	(2, 3)
5.	$\vdash_L \langle y : x\rangle[v/z]B \leftrightarrow [y/x][v/z]B$.	(induction hypothesis)
6.	$\vdash_L \forall v\langle y : x\rangle[v/z]B \leftrightarrow \forall v[y/x][v/z]B$.	(5, (UG), (UD))
7.	$\vdash_L \langle y : x\rangle\forall zB \leftrightarrow \forall v[y/x][v/z]B$.	(4, 6)

5. A is $\langle y_2 : z\rangle B$. Again we have four cases, assuming x and y are modally separated in A.

   a) $z \notin \{x, y\}$. By definition 3.3, then $[y/x]\langle y_2 : z\rangle B$ is $\langle [y/x]y_2 : z\rangle[y/x]B$.

1.	$\vdash \langle y : x\rangle\langle y_2 : z\rangle B \leftrightarrow \langle [y/x]y_2 : z\rangle\langle y : x\rangle B$	((SS1) or (SS2))
2.	$\vdash \langle y : x\rangle B \leftrightarrow [y/x]B$	(induction hypothesis)
3.	$\vdash \langle [y/x]y_2 : z\rangle(\langle y : x\rangle B \leftrightarrow [y/x]B)$	(2, (Sub$_s$))
4.	$\vdash \langle [y/x]y_2 : z\rangle\langle y : x\rangle B \leftrightarrow \langle [y/x]y_2 : z\rangle[y/x]B$	(3, (S$\supset$), (S$\neg$))
5.	$\vdash \langle y : x\rangle\langle y_2 : z\rangle B \leftrightarrow \langle [y/x]y_2 : z\rangle[y/x]B$.	(1, 4)

   b) $z = y$ and $x \notin Varf(B)$. By definition 3.3, then $[y/x]\langle y_2 : z\rangle B$ is $\langle [y/x]y_2 : y\rangle[y/x]B$. By induction hypothesis, $\vdash_L \langle y : x\rangle B \leftrightarrow [y/x]B$. So by (Sub$_s$) and (S$\supset$), $\vdash_L \langle [y/x]y_2 : y\rangle\langle y : x\rangle B \leftrightarrow \langle [y/x]y_2 : y\rangle[y/x]B$. If $y_2 = x$, then $\vdash_L \langle y : x\rangle\langle y_2 : y\rangle B \leftrightarrow \langle [y/x]y_2 : y\rangle\langle y : x\rangle B$ by (SS2). I.e. $\vdash_L \langle y : x\rangle\langle x : y\rangle B \leftrightarrow \langle y : y\rangle\langle y : x\rangle B$. If $y_2 \neq x$, then

1.	$\vdash_L \langle y_2 : y\rangle B \leftrightarrow \langle y : x\rangle\langle y_2 : y\rangle B$	((VS), $x \notin Varf(\langle y_2 : y\rangle B)$)
2.	$\vdash_L B \leftrightarrow \langle y : x\rangle B$	((VS), $x \notin Varf(B)$)
3.	$\vdash_L \langle y_2 : y\rangle B \leftrightarrow \langle y_2 : y\rangle\langle y : x\rangle B$	(1, (Sub$_s$), (S$\supset$))
4.	$\vdash_L \langle y : x\rangle\langle y_2 : y\rangle B \leftrightarrow \langle [y/x]y_2 : y\rangle\langle y : x\rangle B$	(1, 3)

   So either way $\vdash_L \langle y : x\rangle\langle y_2 : y\rangle B \leftrightarrow \langle [y/x]y_2 : y\rangle\langle y : x\rangle B$. So $\vdash_L \langle y : x\rangle\langle y_2 : y\rangle B \leftrightarrow \langle [y/x]y_2 : y\rangle[y/x]B$.

   c) $z = x$ and $y \notin Varf(B)$. By definition 3.3, then $[y/x]\langle y_2 : z\rangle B$ is ([y/x]y$_2$: y)$[y/x]B$. By induction hypothesis, $\vdash_L \langle y : x\rangle B \leftrightarrow [y/x]B$. So by (Sub$_s$) and (S$\supset$), $\vdash_L \langle [y/x]y_2 : y\rangle\langle y : x\rangle B \leftrightarrow \langle [y/x]y_2 : y\rangle[y/x]B$. Since $y \notin Varf(B)$, by (SE2), $\vdash_L \langle [y/x]y_2 : y\rangle\langle y : x\rangle B \leftrightarrow \langle [y/x]y_2 : x\rangle B$. Moreover, $\vdash_L \langle [y/x]y_2 : x\rangle B \leftrightarrow \langle y : x\rangle\langle y_2 : x\rangle B$ by either (VS) (if $x \neq y_2$) or by (SE1), (Sub$_s$) and (S$\supset$) (if $x = y_2$). So $\vdash_L \langle y : x\rangle\langle y_2 : x\rangle B \leftrightarrow \langle [y/x]y_2 : y\rangle\langle y : x\rangle B$.

d) $z = x$ and $y \in \mathit{Varf}(B)$, or $z = y$ and $x \in \mathit{Varf}(B)$. By definition 3.3, then $[y/x]\langle y_2 : z\rangle B$ is $\langle [y/x]y_2 : v\rangle [y/x][v/z]B$, where $v \notin \mathit{Var}(B) \cup \{x, y, y_2\}$.

1.	$\vdash \langle v : z\rangle B \leftrightarrow [v/z]B$	(induction hypothesis)
2.	$\vdash \langle y_2 : v\rangle\langle v : z\rangle B \leftrightarrow \langle y_2 : v\rangle [v/z]B$	(1, (Sub_s), (S$\supset$), (S$\neg$))
3.	$\vdash \langle y_2 : z\rangle B \leftrightarrow \langle y_2 : v\rangle\langle v : z\rangle B$	(SE2)
4.	$\vdash \langle y_2 : z\rangle B \leftrightarrow \langle y_2 : v\rangle [v/z]B$	(2, 3)

Since $z \in \{x, y\}$, x and y are modally separated in $[v/z]B$. So:

5.	$\vdash \langle y : x\rangle [v/z]B \leftrightarrow [y/x][v/z]B$	(ind. hyp.)
6.	$\vdash \langle [y/x]y_2 : v\rangle\langle y : x\rangle [v/z]B \leftrightarrow \langle [y/x]y_2 : v\rangle [y/x][v/z]B$	(5, (Sub_s), (S$\supset$))
7.	$\vdash \langle y : x\rangle\langle y_2 : z\rangle B \leftrightarrow \langle y : x\rangle\langle y_2 : v\rangle [v/z]B$	(4, (Sub_s), (S$\supset$))
8.	$\vdash \langle y : x\rangle\langle y_2 : v\rangle [v/z]B \leftrightarrow \langle [y/x]y_2 : v\rangle\langle y : x\rangle [v/z]B$	((SS1) or (SS2))
9.	$\vdash \langle y : x\rangle\langle y_2 : z\rangle B \leftrightarrow \langle [y/x]y_2 : v\rangle\langle y : x\rangle [v/z]B$	(7, 8)
10.	$\vdash \langle y : x\rangle\langle y_2 : z\rangle B \leftrightarrow \langle [y/x]y_2 : v\rangle [y/x][v/z]B$	(6, 9)

6. A is $\Box B$. For (SC1), assume x and y are modally separated in A. Then they are also modally separated in B, so by induction hypothesis, $\vdash_L \langle y : x\rangle B \leftrightarrow [y/x]B$. By (Nec) and (K), then $\vdash_L \Box\langle y : x\rangle B \leftrightarrow \Box[y/x]B$. By (S$\Box$), $\vdash_L \langle y : x\rangle\Box B \supset \Box\langle y : x\rangle B$. Since at most one of x, y is free in B, by (S$\Diamond$), $\vdash_L \langle y : x\rangle\Diamond\neg B \supset \Diamond\langle y : x\rangle\neg B$; so $\vdash_L \Box\langle y : x\rangle B \supset \langle y : x\rangle\Box B$ (by (S$\neg$), (Sub_s), (S$\supset$), (Nec), (K)). So $\vdash_L \langle y : x\rangle\Box B \leftrightarrow \Box[y/x]B$. Since $\Box[y/x]B$ is $[y/x]\Box B$ by definition 3.3, this means that $\vdash_L \langle y : x\rangle\Box B \leftrightarrow [y/x]\Box B$.

   For (SC2), assume y is modally free for x in $\Box B$. Then y is modally free for x in B, so by induction hypothesis, $\vdash \langle y : x\rangle B \supset [y/x]B$. By (Nec) and (K), then $\vdash \Box\langle y : x\rangle B \supset \Box[y/x]B$. By (S$\Box$), $\vdash \langle y : x\rangle\Box B \supset \Box\langle y : x\rangle B$. So $\vdash \langle y : x\rangle\Box B \supset \Box[y/x]B$.

Here is the proof for (SCN). The first three clauses are very similar.

1. A is atomic. Then $\vdash_L \langle y : x\rangle A \leftrightarrow [y/x]A$ as we've seen above, and so $\vdash_L y \neq y \supset (\langle y : x\rangle A \leftrightarrow [y/x]A)$ by (PC).

2. A is $\neg B$. By induction hypothesis, $\vdash_L y \neq y \supset (\langle y : x\rangle B \leftrightarrow [y/x]B)$. So by (PC), $\vdash_L y \neq y \supset (\neg\langle y : x\rangle B \leftrightarrow \neg[y/x]B)$. By (S$\neg$) and definition 3.3, it follows that $\vdash_L y \neq y \supset (\langle y : x\rangle\neg B \leftrightarrow [y/x]\neg B)$.

3. A is $B \supset C$. By induction hypothesis, $\vdash_L y \neq y \supset (\langle y : x\rangle B \leftrightarrow [y/x]B)$ and $\vdash_L y \neq y \supset (\langle y : x\rangle C \leftrightarrow [y/x]C)$. By (S$\supset$), $\vdash_L y \neq y \supset (\langle y : x\rangle(B \supset C) \leftrightarrow (\langle y : x\rangle B \supset \langle y : x\rangle C))$. So $\vdash_L y \neq y \supset (\langle y : x\rangle(B \supset C) \leftrightarrow ([y/x]B \supset [y/x]C))$, and so $\vdash_L y \neq y \supset (\langle y : x\rangle(B \supset C) \leftrightarrow [y/x](B \supset C))$ by definition 3.3.

4. A is $\forall zB$. If $z \notin \{x, y\}$, then by induction hypothesis, $\vdash_L y \neq y \supset (\langle y : x\rangle B \leftrightarrow [y/x]B)$. So by (UG) and (UD), $\vdash_L \forall z\, y \neq y \supset (\forall z\langle y : x\rangle B \leftrightarrow \forall z[y/x]B)$. Since $z \notin \{x, y\}$, $\vdash_L \langle y : x\rangle\forall zB \leftrightarrow \forall z\langle y : x\rangle B$ by (S$\forall$), and $\vdash_L y \neq y \supset \forall z\, y \neq y$ by (VQ), and $\forall z[y/x]B$ is $[y/x]\forall zB$ by definition 3.3; so $\vdash_L y \neq y \supset (\langle y : x\rangle\forall zB \leftrightarrow [y/x]\forall zB)$.

Alternatively, if $z \in \{x, y\}$, then either x or y is not free in A, and thus x and y are modally separated in A. By (SC2), then $\vdash_L \langle y : x \rangle \forall z B \leftrightarrow [y/x]\forall z B$, and so by (PC), $\vdash_L y \neq y \supset (\langle y : x \rangle \forall z B \leftrightarrow [y/x]\forall z B)$.

5. A is $\langle y_2 : z \rangle B$. If $z \notin \{x, y\}$, then by induction hypothesis, $\vdash_L y \neq y \supset (\langle y : x \rangle B \leftrightarrow [y/x]B)$. So by ($\text{Sub}_\text{s}$) and (S$\supset$), $\vdash_L \langle [y/x]y_2 : z \rangle y \neq y \supset (\langle [y/x]y_2 : z \rangle \langle y : x \rangle B \leftrightarrow \langle [y/x]y_2 : z \rangle [y/x]B)$. By (VS), $\langle [y/x]y_2 : z \rangle y \neq y \leftrightarrow y \neq y$. And by (SS1) or (SS2), $\langle y : x \rangle \langle y_2 : z \rangle B \leftrightarrow \langle [y/x]y_2 : z \rangle \langle y : x \rangle B$. So $\vdash_L y \neq y \supset (\langle y : x \rangle \langle y_2 : z \rangle B \leftrightarrow \langle [y/x]y_2 : z \rangle [y/x]B)$. But by definition 3.3, $[y/x]\langle y_2 : z \rangle B$ is $\langle [y/x]y_2 : y \rangle [y/x]B$.

   Alternatively, if $z \in \{x, y\}$, then either x or y is not free in A, and thus x and y are modally separated in A. By (SC2), then $\vdash_L \langle y : x \rangle \langle y_2 : z \rangle B \leftrightarrow [y/x]\langle y_2 : z \rangle B$, and so by (PC), $\vdash_L y \neq y \supset (\langle y : x \rangle \langle y_2 : z \rangle B \leftrightarrow [y/x]\langle y_2 : z \rangle B)$.

6. A is $\Box B$. Then

1.	$\vdash_L y \neq y \supset (\langle y : x \rangle B \leftrightarrow [y/x]B)$.	(ind. hyp.)
2.	$\vdash_L \Box y \neq y \supset (\Box \langle y : x \rangle B \leftrightarrow \Box [y/x]B)$.	(1, (Nec), (K))
3.	$\vdash_L y \neq y \supset \Box y \neq y$.	((=R) or (NA), (EI) and (Nec))
4.	$\vdash_L y \neq y \supset (\Box \langle y : x \rangle B \leftrightarrow \Box [y/x]B)$.	(2, 3)
5.	$\vdash_L y \neq y \supset \langle y : x \rangle (x \neq x \wedge y \neq y)$	((SAt), (S$\supset$), (S$\neg$))
6.	$\vdash_L (x \neq x \wedge y \neq y) \supset \Box (x \neq x \wedge y \neq y)$.	((=R) or (NA), (EI), (Nec) and
7.	$\vdash_L \Box (x \neq x \wedge y \neq y) \supset (\Box B \leftrightarrow \Box \langle y : x \rangle B)$.	((SEV), (Nec), (K))
8.	$\vdash_L (x \neq x \wedge y \neq y) \supset (\Box B \leftrightarrow \Box \langle y : x \rangle B)$.	(6, 7)
9.	$\vdash_L \langle y : x \rangle (x \neq x \wedge y \neq y) \supset (\langle y : x \rangle \Box B \leftrightarrow \langle y : x \rangle \Box \langle y : x \rangle B)$.	(8, (Sub_s), (S$\supset$))
10.	$\vdash_L \langle y : x \rangle (x \neq x \wedge y \neq y) \supset (\langle y : x \rangle \Box B \leftrightarrow \Box \langle y : x \rangle B)$.	(9, (VS))
11.	$\vdash_L y \neq y \supset (\langle y : x \rangle \Box B \leftrightarrow \Box \langle y : x \rangle B)$.	(7, 10)
12.	$\vdash_L y \neq y \supset (\langle y : x \rangle \Box B \leftrightarrow [y/x]\Box B)$.	(4, 13, def. 3.3)

■

LEMMA 5.10 (SYNTACTIC ALPHA-CONVERSION)
If A, A' are $\mathcal{L}$-formulas, and A' is an alphabetic variant of A, then

(AC) $\vdash_L A \leftrightarrow A'$.

PROOF by induction on A.

1. A is atomic. Then $A = A'$ and $\vdash_L A \leftrightarrow A'$ by (Taut).
2. A is $\neg B$. Then A' is $\neg B'$ with B' an alphabetic variant of B. By induction hypothesis, $\vdash_L B \leftrightarrow B'$. By (PC), $\vdash_L \neg B \leftrightarrow \neg B'$.

3. A is $B \supset C$. Then A' is $B' \supset C'$ with B', C' alphabetic variants of B, C, respectively. By induction hypothesis, $\vdash_L B \leftrightarrow B'$ and $\vdash_{sC} C \leftrightarrow C'$. By (PC), then $\vdash_L (B \supset C) \leftrightarrow (B' \supset C')$.

4. A is $\forall x B$. Then A' is either $\forall x B'$ or $\forall z[z/x]B'$, where B' is an alphabetic variant of B and $z \notin \mathit{Var}(B')$. Assume first that A' is $\forall x B'$. By induction hypothesis, $\vdash_L B \leftrightarrow B'$. So by (UG) and (UD), $\vdash_L \forall x B \leftrightarrow \forall x B'$.

   Alternatively, assume A' is $\forall z[z/x]B'$ and $z \notin \mathit{Var}(B')$. Since B' differs from B at most in renaming bound variables, if z were free in B, then $z \in \mathit{Var}(B')$. So z is not free in B. Then

1.	$\vdash_L B \leftrightarrow B'$.	induction hypothesis
2.	$\vdash_L \langle z:x\rangle B \leftrightarrow \langle z:x\rangle B'$.	(1, ($\mathrm{Sub_s}$), (S$\neg$))
3.	$\vdash_L \langle z:x\rangle B' \leftrightarrow [z/x]B'$.	((SC1), $z \notin \mathit{Var}(B')$)
4.	$\vdash_L \langle z:x\rangle B \leftrightarrow [z/x]B'$.	(2, 3)
5.	$\vdash_L \forall z\langle z:x\rangle B \leftrightarrow \forall z[z/x]B'$.	(4, (UG), (UD))
6.	$\vdash_L \forall x B \leftrightarrow \forall z\langle z:x\rangle B$.	((SBV), z not free in B)
7.	$\vdash_L \forall x B \leftrightarrow \forall z[z/x]B'$.	(5, 6)

5. A is $\langle y:x\rangle B$. Then A' is either $\langle y:x\rangle B'$ or $\langle y:z\rangle[z/x]B'$, where B' is an alphabetic variant of B and $z \notin \mathit{Var}(B)$. Assume first that A' is $\langle y:x\rangle B'$. By induction hypothesis, $\vdash_L B \leftrightarrow B'$. So by ($\mathrm{Sub_s}$) and (S$\supset$), $\vdash_L \langle y:x\rangle B \leftrightarrow \langle y:x\rangle B'$.

   Alternatively, assume A' is $\langle y:z\rangle[z/x]B'$ and $z \notin \mathit{Var}(B')$. Again, it follows that z is not free in B. So

1.	$\vdash_L B \leftrightarrow B'$.	induction hypothesis
2.	$\vdash_L \langle z:x\rangle B \leftrightarrow \langle z:x\rangle B'$.	(1, ($\mathrm{Sub_s}$), (S$\supset$))
3.	$\vdash_L \langle z:x\rangle B' \leftrightarrow [z/x]B'$.	((SC1), $z \notin \mathit{Var}(B')$)
4.	$\vdash_L \langle z:x\rangle B \leftrightarrow [z/x]B'$.	(2, 3)
5.	$\vdash_L \langle y:z\rangle\langle z:x\rangle B \leftrightarrow \langle y:z\rangle[z/x]B'$.	(4, ($\mathrm{Sub_s}$), (S$\supset$))
6.	$\vdash_L \langle y:z\rangle\langle z:x\rangle B \leftrightarrow \langle y:x\rangle B$.	((SE2), z not free in B)
7.	$\vdash_L \langle y:x\rangle B \leftrightarrow \langle y:z\rangle[z/x]B'$.	(5, 6)

6. A is $\Box A'$. Then B is $\Box B'$ with B' an alphabetic variant of A'. By induction hypothesis, $\vdash_L A' \leftrightarrow B'$. Then by (Nec), $\vdash_L \Box(A' \leftrightarrow B')$, and by (K), $\vdash_L \Box A' \leftrightarrow \Box B'$. ∎

THEOREM 5.11 (SUBSTITUTION AND NON-SUBSTITUTION LOGICS)
For any $\mathcal{L}$-formula A and variables x, y,

(FUI*) $\vdash_L \forall x A \supset (Ey \supset [y/x]A)$, provided y is modally free for x in A,

(LL*) $\vdash_L x{=}y \supset A \supset [y/x]A$, provided y is modally free for x in A, (Sub*) if $\vdash_L A$, then $\vdash_L [y/x]A$, provided y is modally free for x in A.

It follows that $\mathsf{P} \subseteq \mathsf{P}_s$ and $\mathsf{N} \subseteq \mathsf{N}_s$.

PROOF Assume y is modally free for x in A. Then by (SC2), $\vdash_L \langle y : x\rangle A \supset [y/x]A$. By ($\mathrm{FUI_s}$), $\vdash_L \forall x A \supset (Ey \supset \langle y : x\rangle A)$, so by (PC), $\vdash_L \forall x A \supset (Ey \supset [y/x]A)$. Similarly, by ($\mathrm{LL_s}$), $\vdash_L x{=}y \supset A \supset \langle y : x\rangle A$, so by (PC), $\vdash_L x{=}y \supset A \supset [y/x]A$. Finally, by ($\mathrm{Sub_s}$), if $\vdash_L A$, then $\vdash_L \langle y : x\rangle A$, so then $\vdash_L [y/x]A$ by (PC). ■

LEMMA 5.12 (SYMMETRY AND TRANSITIVITY OF IDENTITY)
For any $\mathcal{L}$-variables x, y, z,

(=S) $\vdash_L x{=}y \supset y{=}x$;

(=T) $\vdash_L x{=}y \supset y{=}z \supset x{=}z$.

PROOF Immediate from lemma 5.11 and lemma 4.10. ■

For (= S), let v be some variable $\notin \{x, y\}$. Then

1. $\vdash_L v{=}y \supset v{=}x \supset \langle y : v\rangle v{=}x$. ($\mathrm{LL_s}$)
2. $\vdash_L v{=}y \supset v{=}x \supset y{=}x$. (1, (SAt))
3. $\vdash_L \langle x : v\rangle(v{=}y \supset v{=}x \supset y{=}x)$. ($\mathrm{Sub_s}$)
4. $\vdash_L x{=}y \supset x{=}x \supset y{=}x$. (3, (S⊃), (SAt))
5. $\vdash_L x{=}y \supset x{=}x$. ((=R), or (Neg) and (∀=R))
6. $\vdash_L x{=}y \supset y{=}x$. (4, 5)

For (= T),

1. $\vdash_L x{=}y \supset y{=}x$. (=S)
2. $\vdash_L y{=}x \supset y{=}z \supset x{=}z$. (($\mathrm{LL_s}$), (SAt))
3. $\vdash_L x{=}y \supset y{=}z \supset x{=}z$. (1, 2)

■

LEMMA 5.13 (VARIATIONS ON LEIBNIZ' LAW)
If A is an $\mathcal{L}$-formula and x, y, y' are $\mathcal{L}$-variables, then

(LV1) $\vdash_L x{=}y \supset \langle y : x\rangle A \supset A$.

(LV2) $\vdash_L y{=}y' \supset \langle y : x\rangle A \supset [y'/x]A$, provided y' is modally free for x in A.

(LV1) (formerly (LL_i)) seems to be never used. Is it interesting enough to list on its own?

Proof (LV1). Let z be an $\mathcal{L}$-variable not in $Var(A)$. Then

1.	$\vdash_L x{=}z \supset \langle z:x\rangle A \supset \langle x:z\rangle\langle z:x\rangle A.$	(LL_s)
2.	$\vdash_L x{=}z \supset \langle z:x\rangle A \supset \langle x:x\rangle A.$	(1, (SE2), $z \notin Var(A)$)
3.	$\vdash_L x{=}z \supset \langle z:x\rangle A \supset A.$	(2, (SE1))
4.	$\vdash_L \langle y:z\rangle x{=}z \supset \langle y:z\rangle\langle z:x\rangle A \supset \langle y:z\rangle A.$	(3, (VS), (S$\supset$))
5.	$\vdash_L x{=}z \supset \langle y:z\rangle\langle z:x\rangle A \supset \langle y:z\rangle A.$	(4, (SAt))
6.	$\vdash_L x{=}z \supset \langle y:x\rangle A \supset \langle y:z\rangle A.$	(5, (SE2), $z \notin Var(A)$)
7.	$\vdash_L x{=}z \supset \langle y:x\rangle A \supset A.$	(6, (VS), $z \notin Var(A)$).

(LV2).

1.	$\vdash_L x{=}y \wedge y{=}y' \supset x{=}y'.$	(=T)
2.	$\vdash_L A \wedge x{=}y' \supset [y'/x]A.$	((LL*), y' m.f. in A)
3.	$\vdash_L A \wedge x{=}y \wedge y{=}y' \supset [y'/x]A.$	(1, 2)
4.	$\vdash_L \langle y:x\rangle A \wedge \langle y:x\rangle x{=}y \wedge \langle y:x\rangle y{=}y' \supset \langle y:x\rangle[y'/x]A.$	(3, (Sub_s), (S$\neg$), (S$\supset$))
5.	$\vdash_L y{=}y \supset \langle y:x\rangle x{=}y.$	(SAt)
6.	$\vdash_L y{=}y' \supset y{=}y.$	((LL*), (=S))
7.	$\vdash_L y{=}y' \supset \langle y:x\rangle y{=}y'.$	(VS)
8.	$\vdash_L \langle y:x\rangle A \wedge y{=}y' \supset \langle y:x\rangle[y'/x]A.$	(4, 5, 6, 7)
9.	$\vdash_L \langle y:x\rangle[y'/x]A \supset [y'/x]A.$	(VS)
10.	$\vdash_L \langle y:x\rangle A \wedge y{=}y' \supset [y'/x]A.$	(8, 9)

■

Lemma 5.14 (Leibniz' Law with sequences)
For any $\mathcal{L}$-formula A and variables $x_1,\dots,x_n,y_1,\dots,y_n$ such that the $x_1,\dots,x_n$ are pairwise distinct,

(LL_n) $\vdash_L x_1{=}y_1 \wedge \dots \wedge x_n{=}y_n \supset A \supset \langle y_1,\dots,y_n : x_1,\dots,x_n\rangle A.$

Proof For $n = 1$, (LL_n) is (LL_s). Assume then that $n > 1$. To keep formulas in the following proof at a managable length, let $\underline{\phi(i)}$ abbreviate the sequence $\phi(1),\dots,\phi(n-1)$. For example, $\langle \underline{y_i} : \underline{x_i}\rangle$ is $\langle y_1,\dots,y_{n-1} : x_1,\dots,x_{n-1}\rangle$. Let z be the alphabetically first variable not in A or

$x_1, \ldots, x_n$. Now

1.	$\vdash_L x_n = y_n \supset \langle \underline{y_i} : \underline{x_i} \rangle A \supset \langle y_n : x_n \rangle \langle \underline{y_i} : \underline{x_i} \rangle A.$	(LL$_\mathrm{s}$)
2.	$\vdash_L \langle y_n : x_n \rangle \langle \underline{y_i} : \underline{x_i} \rangle A \supset \langle y_n : z \rangle \langle z : x_n \rangle \langle \underline{y_i} : \underline{x_i} \rangle A.$	(SE1)
3.	$\vdash_L \langle z : x_n \rangle \langle \underline{y_i} : \underline{x_i} \rangle A \supset \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A.$	((SS1) or (SS2))
4.	$\vdash_L \langle y_n : z \rangle \langle z : x_n \rangle \langle \underline{y_i} : \underline{x_i} \rangle A$ $\supset \langle y_n : z \rangle \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A.$	(3, (Sub$_\mathrm{s}$), (S$\supset$))
5.	$\vdash_L x_n = y_n \supset \langle \underline{y_i} : \underline{x_i} \rangle A \supset \langle y_n : z \rangle \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A.$	(1, 2, 4)
6.	$\vdash_L x_n = z \supset \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A$ $\supset \langle z : x_n \rangle \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A.$	(LL$_\mathrm{s}$)
7.	$\vdash_L x_n = z \supset \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A$ $\supset \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle \langle z : x_n \rangle A.$	(6, (SS1))
8.	$\vdash_L z = x_n \supset \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle \langle z : x_n \rangle A$ $\supset \langle x_n : z \rangle \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle \langle z : x_n \rangle A.$	(LL$_\mathrm{s}$)
9.	$\vdash_L z = x_n \supset \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle \langle z : x_n \rangle A$ $\supset \langle \underline{y_i} : \underline{x_i} \rangle \langle x_n : z \rangle \langle z : x_n \rangle \langle z : x_n \rangle A.$	(8, (SS1), (SS2))
10.	$\vdash_L \langle x_n : z \rangle \langle z : x_n \rangle \langle z : x_n \rangle A \leftrightarrow \langle z : x_n \rangle A$	((SE1), (SE2))
11.	$\vdash_L \langle \underline{y_i} : \underline{x_i} \rangle \langle x_n : z \rangle \langle z : x_n \rangle \langle z : x_n \rangle A \supset \langle \underline{y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A$	(10, (Sub$_\mathrm{s}$), (S$\supset$))
12.	$\vdash_L z = x_n \supset x_n = z$	(=S)
13.	$\vdash_L z = x_n \supset \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A \supset \langle \underline{y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A.$	(7, 9, 11, 12)
14.	$\vdash_L x_n = y_n \supset \langle y_n : z \rangle z = x_n$	((=S), (SAt))
15.	$\vdash_L x_n = y_n \supset \langle y_n : z \rangle \langle \underline{[z/x_n] y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A$ $\supset \langle y_n : z \rangle \langle \underline{y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A.$	13, 14, (Sub$_\mathrm{s}$), (S$\supset$)
16.	$\vdash_L x_n = y_n \supset \langle \underline{y_i} : \underline{x_i} \rangle A \supset \langle y_n : z \rangle \langle \underline{y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A.$	5, 15
17.	$\vdash_L x_1 = y_1 \wedge \ldots \wedge x_{n-1} = y_{n-1} \supset A \supset \langle \underline{y_i} : \underline{x_i} \rangle A.$	(induction hypothesis)
18.	$\vdash_L x_1 = y_1 \wedge \ldots \wedge x_n = y_n \supset A \supset \langle y_n : z \rangle \langle \underline{y_i} : \underline{x_i} \rangle \langle z : x_n \rangle A.$	(16, 17)
19.	$\vdash_L x_1 = y_1 \wedge \ldots \wedge x_n = y_n \supset A \supset \langle y_1, \ldots, y_n : x_1, \ldots, x_n \rangle A.$	(18, def. 3.14)

■

Lemma 5.15 (Closure under transformations)
For any $\mathcal{L}$-formula A and transformation τ on $\mathcal{L}$,

(Sub$^\tau$) $\vdash_L A$ iff $\vdash_L A^\tau$.

Proof The proof is exactly as in lemma 4.12. Assume $\vdash_L A$. Let $x_1, \ldots, x_n$ be the variables in A. If $n = 0$, then $A = A^\tau$ and the result is trivial. If $n = 1$, then A^τ is $[x_1^\tau / x_1]A$, and x_1^τ is either x_1 itself or does not occur in A. In the first case, $[x_1^\tau/x_1]A = A$ and the result is again trivial. In the second case, x_1^τ is modally free for x_1 in A, and thus $\vdash_L [x_1^\tau/x_1]A$ by (Sub*). – More explicitly, we have $\vdash_L \langle x_1^\tau : x_1\rangle A$ by ($\mathrm{Sub_s}$), and so $\vdash_L [x_1^\tau/x_1]A$ by (SC2).

Assume then that $n > 1$. Note first that $A = [x_n^\tau/v_n]\ldots[x_2^\tau/v_2][x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]A$, where $v_2,\ldots,v_n$ are distinct variables not in A or A^τ (compare definition 3.14; this is easily shown by induction on the subformulas B of A): let Σ abbreviate $[x_n^\tau/v_n]\ldots[x_2^\tau/v_2][x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]$.

1. B is $Px_j\ldots x_k$. Then $x_j,\ldots,x_k$ are variables from $x_1,\ldots,x_n$, and $\Sigma B = Px_j^\tau\ldots x_k^\tau = B^\tau$.
2. B is $\neg C$. By induction hypothesis, $\Sigma C = C^\tau$, hence $\neg\Sigma C = \neg C^\tau$. By definitions **??** and 3.11, $\Sigma\neg C$ is $\neg\Sigma C$, and $(\neg C)^\tau$ is $\neg C^\tau$.
3. B is $C \supset D$. Similar.
4. B is $\forall zC$. By induction hypothesis, $\Sigma C = C^\tau$. Since τ is injective, by definitions 3.3 and 3.11, $\Sigma\forall zC$ is $\forall\Sigma z\Sigma C$, and $(\forall zC)^\tau$ is $\forall z^\tau C^\tau$. (Here things would get a lot more complicated if we had defined substitution differently, so that $[y/x]\forall xFx \neq \forall yFy$.) And it is easy to verify that $\Sigma z = z^\tau$.
5. B is $\langle y : z\rangle C$. By induction hypothesis, $\Sigma C = C^\tau$. Since τ is injective, by definitions 3.3 and 3.11, $\Sigma\langle y : z\rangle C$ is $\langle\Sigma y : \Sigma z\rangle\Sigma C$, and $(\langle y : z\rangle C)^\tau$ is $\langle y^\tau : z^\tau\rangle C^\tau$. And it is easy to verify that $\Sigma y = y^\tau$ and $\Sigma z = z^\tau$.
6. B is $\Box C$. By induction hypothesis, ΣC is C^τ, hence $\Box\Sigma C$ is $\Box C^\tau$. By definitions 3.3 and 3.11, $\Sigma\Box C$ is $\Box\Sigma C$, and $(\Box C)^\tau$ is $\Box C^\tau$.

Since v_n is modally free for x_n in A, by (Sub*), $\vdash_L [v_n/x_n]A$. – More precisely, by ($\mathrm{Sub_s}$), $\vdash_L \langle v_n : x_n\rangle A$, and by (SC2), $\vdash_L [v_n/x_n]A$. Likewise, for each $1 < i < n$, v_i is modally free for x_i in $[v_{i+1}/x_{i+1}]\ldots[v_n/x_n]A$. So $\vdash_L [v_2/x_2]\ldots[v_n/x_n]A$.

With respect to $[x_1^\tau/x_1]$, we distinguish three cases. First, if $x_1 = x_1^\tau$, then $\vdash_L [x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]A$, because $[x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]A$ is $[v_2/x_2]\ldots[v_n/x_n]A$. Second, if $x_1 \neq x_1^\tau$ and $x_1^\tau \notin \mathit{Var}(A)$, then $x_1^\tau \notin \mathit{Var}([v_2/x_2]\ldots[v_n/x_n]A)$, since the $v_1,\ldots,v_n$ are not in $\mathit{Var}(A)$ or $\mathit{Var}(A^\tau)$ (in particular, thus no new variables are introduced in $[v_2/x_2]\ldots[v_n/x_n]A$). So x_1^τ is modally free for x_1 in $[v_2/x_2]\ldots[v_n/x_n]A$, and by (Sub*), $\vdash_L [x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]A$. Third, if $x_1 \neq x_1^\tau$ and $x_1^\tau \in \mathit{Var}(A)$, then x_1^τ must be one of $x_2,\ldots,x_n$. Then again $x_1^\tau \notin \mathit{Var}([v_2/x_2]\ldots[v_n/x_n]A)$, and so $\vdash_L [x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]A$ by (Sub*).

Finally, x_2^τ is modally free for v_2 in $[x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]A$, because τ is injective and hence $x_2^\tau \neq x_1^\tau$, so x_2^τ does not occur in $[x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]A$. Hence $\vdash_L [x_2^\tau/v_2][x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]A$. By the same reasoning, for each $2 < i \leq n$, x_i^τ is modally free for v_i in $[x_{i-1}^\tau/v_{i-1}]\ldots[x_2^\tau/v_2][x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]$ So $\vdash_L [x_n^\tau/v_n]\ldots[x_2^\tau/v_2][x_1^\tau/x_1][v_2/x_2]\ldots[v_n/x_n]A$, i.e. $\vdash_L A^\tau$. ∎

6 Correspondence

A well-known feature of Kripke semantics for propositional modal logic is that various modal principles correspond to conditions on the accessibility relation, in the sense that the principle is valid in all and only those Kripke frames whose accessibility relation satisfies the condition: $\Box p \supset p$ *corresponds to* (or *defines*) the class of reflexive frames, $p \supset \Box\Diamond p$ the class of symmetrical frames, and so on.

In our counterpart semantics, all these facts are preserved if we translate the sentence letters of propositional modal logic into null-ary predicates (or other sentences without free variables). Things are slightly more complex if we substitute the sentence letters by formulas with free variables. For example, the general schema $\Box A \supset A$, where A may contain arbitrarily many free variables, is valid in a counterpart structure iff (i) every world can see itself, and (ii) every individual (and every sequence of individuals) at every world is a counterpart of itself (relative to some counterpart relation). To see why (i) is not enough, consider what is required for the validity of $\Box Fx \supset Fx$. Loosely speaking, the antecedent $\Box Fx$ is true at w iff all counterparts of x are F at all accessible worlds. This does not entail that x is F at w unless (i) w can see itself and (ii) x is its own counterpart at w.

Proof Assume first that some structure $\mathcal{S} = \langle W, R, U, D, K \rangle$ does not satisfy (i) and (ii). If some $w \in W$ is not R-related to itself, then $w, V \nVdash_{\mathcal{S}} \Box P_0 \supset P_0$, where P_0 is a zero-ary predicate with $V_w(P_0) = 0$ and $V_{w'}(P_0) = 1$ for all $w' \neq w$. If some individual $d \in U_w$ is not a counterpart of itself at w relative to some $C \in K_{w,w}$, then $w, V \nVdash_{\mathcal{S}} \Box\neg Fx \supset \neg Fx$, where $V_w(x) = d$, $V_w(F) = \{d\}$, and $V_{w'}(F) = \emptyset$ for all $w' \neq w$. In the other direction, assume some instance of $\Box A \supset A$ is false at some $w \in W$ in some structure $\mathcal{S} = \langle W, R, U, D, K \rangle$ under some interpretation V on $\mathcal{S}$. Then $w, V \Vdash_{\mathcal{S}} \Box A$ and $w, V \nVdash_{\mathcal{S}} A$. The former means that $w', V' \Vdash_{\mathcal{S}} A$ whenever wRw' and $V_w \triangleright V'_{w'}$. But if R is reflexive, then wRw. Moreover, if there is a $C \in K_{w,w}$ for which $V_w(x) C V_w(x)$ whenever $V_w(x)$ is defined, then $V_w \triangleright V'_{w'}$. Under these conditions, it therefore follows that $w, V \Vdash_{\mathcal{S}} A$, which refutes the assumption that $w, V \nVdash_{\mathcal{S}} \Box A \supset A$. ■

In a sense, formulas containing free variables don't just express properties of worlds, but relations between a world and a (finite) sequence of individuals. In any given counterpart model, the truth-value of Fx is fixed by choosing a world. In *this* sense, the formula expresses just a property of worlds. On the other hand, the truth-value of $\forall x Fx$ (or $\Box Fx$) at a world is not simply a function of the truth-value of Fx at that world or other worlds. Whether $\forall x Fx$ is true at w depends on whether Fx is true at w relative to every choice of an individual $d \in D_w$ as value of x. (Similarly, whether $\Box Fx$ is true at w depends on whether Fx is true at all accessible worlds w' relative to every choice of an x-counterpart as value of x.) In a Tarski-style semantics this is rendered explicit by the fact that formulas with free variables are evaluated not only relative to a world, but relative to a world w and an infinite sequence of individuals from w, representing different assignments of values to the variables. (Our Mates-style semantics incorporates the assignment function in the interpretation function V; hence our rules for quantifiers quantify over alternative interpretations rather than alternative assignments.) By distinguishing truth from satisfaction, we could still say that truth in a model is only relative to a world, since a formula A is *true* at a world w in a model $\langle \mathcal{S}, I, \Sigma \rangle$ iff $\mathcal{S}, I, w, \Sigma_w \Vdash A$, but for the present topic satisfaction is the more useful notion. That's the sense in which formulas express properties of sequences. Tarski's semantics misleadingly suggests that relevant sequences are infinite, although actual formulas of standard first-order logic only contain a finite number of variables and thus only ever constrain a finite initial segment of a Tarskian sequence.

It is important to look at *schemas*, rather than single formulas. For instance, it is not true in PML that for all A, $\Box A \supset A$ is valid in all and only reflexive frames. For example, with $A = p \supset p$, the formula is valid in all frames whatsoever, since the consequent is never false. – OTOH, in PML we can look at $\Box p \supset p$ instead: that's gonna be valid iff all instances of $\Box A \supset A$ are.

First a brief review of some definitions from propositional modal logic.

> Definition 6.1 (Languages of propositional modal logic)
> A set of formulas $\mathcal{L}_0$ is a *(unimodal) propositional language* if there is a denumerable set of symbols *Prop* (the *sentence letters* of $\mathcal{L}_0$) distinct from $\{\neg, \supset, \Box\}$ such that $\mathcal{L}_0$ is generated by the rule
>
> $$P \mid \neg A \mid (A \supset B) \mid \Box A,$$
>
> where $P \in$ *Prop*.

Note that any language $\mathcal{L}$ of quantified modal logic is also a unimodal propositional language, with sentence letters defined as all $\mathcal{L}$-formulas not of the form $\neg A, A \supset B$ or $\Box A$. (So $\forall x \Box (Fx \supset Gx)$, for example, is a sentence letter.) For future reference, let's call such formulas *quasi-atomic*.

> Definition 6.2 (Quasi-atomic formulas)
> A formula A of quantified modal logic is *quasi-atomic* if it is not of the form $\neg B, B \supset C$ or $\Box B$.

> Definition 6.3 (Frames and valuations)
> A *frame* is a pair consisting of a non-empty set W and a relation $R \subseteq W^2$.
>
> A *valuation of* a unimodal propositional language $\mathcal{L}_0$ *on* a frame $\mathcal{F} = \langle W, R \rangle$ is a function V that maps every sentence letter of $\mathcal{L}_0$ to a subset of W.

Definition 6.4 (Propositional truth)
For any frame $\mathcal{F} = \langle W, R \rangle$, point $w \in W$, and valuation V on $\mathcal{F}$,

$$\begin{aligned}
&\mathcal{F}, V, w \Vdash_K P && \text{for } P \in \mathit{Prop} \text{ iff } w \in V(P),\\
&\mathcal{F}, V, w \Vdash_K \neg B && \text{iff } \mathcal{F}, V, w \nVdash_K B,\\
&\mathcal{F}, V, w \Vdash_K B \supset C && \text{iff } \mathcal{F}, V, w \nVdash_K B \text{ or } \mathcal{F}, V, w \Vdash_K C,\\
&\mathcal{F}, V, w \Vdash_K \Box B && \text{iff } \mathcal{F}, V, w' \Vdash_K B \text{ for all } w' \text{ with } wRw'.
\end{aligned}$$

Definition 6.5 (Frame validity)
A formula A of a unimodal propositional language is *valid in a frame* $\mathcal{F} = \langle W, R \rangle$ if $\mathcal{F}, V, w \Vdash_K A$ for all $w \in W$ and valuation functions V of the language on $\mathcal{F}$. A set of formulas $\mathbb{A}$ is *valid in a frame* $\mathcal{F}$ if all members of $\mathbb{A}$ are valid in $\mathcal{F}$.

To keep things simple, I will focus on positive models for a moment.

Definition 6.6 (n-sequential accessibility relations)
Given a total counterpart structure $\mathcal{S} = \langle W, R, U, D, K \rangle$ and number $n \in \mathbb{N}$, the *n-sequential accessibility relation $R^n_\mathcal{S}$ of $\mathcal{S}$* is the binary relation such that $\langle w, d_1, \ldots, d_n \rangle R^n_\mathcal{S} \langle w', d'_1, \ldots, d'_n \rangle$ iff wRw' and for some $C \in K_{w,w'}$, $d_1 C d'_1, \ldots, d_n C d'_n$.

(Note that $R^0_\mathcal{S} = R$.) We can reformulate the semantics of section 2 in terms of $R^n_\mathcal{S}$.

Lemma 6.7 (Sequential semantics)
Let $A(x_1, \ldots, x_n)$ be a formula with free variables $x_1, \ldots, x_n$. Restricted to total structures $\mathcal{S}$ and interpretations V, the clause for the box in definition 2.7, viz.

$$w, V \Vdash_\mathcal{S} \Box A(x_1, \ldots, x_n) \text{ iff } w', V' \Vdash_\mathcal{S} A \text{ for all } w', V' \text{ such that } wRw' \text{ and } V_w \triangleright V'_{w'}.$$

is equivalent to

$$w, V \Vdash_\mathcal{S} \Box A(x_1, \ldots, x_n) \text{ iff } w', V' \Vdash_\mathcal{S} A(x_1, \ldots, x_n) \text{ for all } w', V' \text{ such that } \langle w, V_w(x_1), \ldots, V_w(x_n) \rangle R^n_\mathcal{S} \langle w', V'_{w'}(x_1,) \ldots, V'_{w'}(x_n) \rangle.$$

Proof By definitions 2.7 and 2.6, $w, V \Vdash_\mathcal{S} \Box A(x_1, \ldots, x_n)$ iff

(1) $w', V' \Vdash_{\mathcal{S}} A(x_1, \ldots, x_n)$ for all w', V' such that wRw' and for some $C \in K_{w,w'}$ and all variables x, $V_w(x)$ is C-related to $V'_{w'}(x)$.

By lemma 2.11, it doesn't matter what V' assigns to variables not in $A(x_1, \ldots, x_n)$. So (1) is equivalent to

(2) $w', V' \Vdash_{\mathcal{S}} A(x_1, \ldots, x_n)$ for all w', V' such that wRw' and for some $C \in K_{w,w'}$ and all variables $x_i \in x_1, \ldots, x_n$, $V_w(x_i)$ is C-related to $V'_{w'}(x_i)$.

By definition 6.6, $\langle w, V_w(x_1), \ldots, V_w(x_n)\rangle R^n_{\mathcal{S}} \langle w', V'_{w'}(x_1), \ldots, V_w(x_n)\rangle$ iff wRw' and for some $C \in K_{w,w'}$ and all variables $x_i \in x_1, \ldots, x_n$, $V_w(x_i)$ is C-related to $V'_{w'}(x_i)$. So (2) is equivalent to

(3) $w', V' \Vdash_{\mathcal{S}} A(x_1, \ldots, x_n)$ for all w', V' such that $\langle w, V_w(x_1), \ldots, V_w(x_n)\rangle R^n_{\mathcal{S}} \langle w', V'_{w'}(x_1), \ldots, V'_{w'}(x_n)\rangle$.

■

We could accommodate negative structures here, but then we'd have to define R^n to hold not between world-sequence pairs but between pairs of a world and a partial function from numbers to individuals. The proof then goes as follows.

Proof By definitions 2.7 and 2.6, $w, V \Vdash_{\mathcal{S}} \Box A(x_1, \ldots, x_n)$ iff

(1) $w', V' \Vdash_{\mathcal{S}} A(x_1, \ldots, x_n)$ for all w', V' such that wRw' and for some $C \in K_{w,w'}$ and all variables x, either $V_w(x)$ is C-related to $V'_{w'}(x)$ or there is no d to which $V_w(x)$ is C-related in which case $V'_{w'}(x)$ is undefined.

By lemma 2.11, it doesn't matter what V' assigns to variables not in $A(x_1, \ldots, x_n)$. So (1) is equivalent to

(2) $w', V' \Vdash_{\mathcal{S}} A(x_1, \ldots, x_n)$ for all w', V' such that wRw' and for some $C \in K_{w,w'}$ and all variables $x_i \in x_1, \ldots, x_n$, either $V_w(x_i)$ is C-related to $V'_{w'}(x_i)$ or there is no d to which $V_w(x_i)$ is C-related in which case $V'_{w'}(x_i)$ is undefined.

By definition 6.6, $\langle w, V_w(x_1, \ldots, x_n)\rangle R^* \langle w', V'_{w'}(x_1, \ldots, x_n)\rangle$ iff wRw' and for some $C \in K_{w,w'}$ and all variables $x_i \in x_1, \ldots, x_n$, either $V_w(x_i)$ is C-related to $V'_{w'}(x_i)$ or there is no d to which $V_w(x_i)$ is C-related in which case $V'_{w'}(x_i)$ is undefined. So (2) is equivalent to

(3) $w', V' \Vdash_{\mathcal{S}} A(x_1, \ldots, x_n)$ for all w', V' such that $\langle w, V_w(x_1, \ldots, x_n)\rangle R^* \langle w', V'_{w'}(x_1, \ldots, x_n)\rangle$.

■

By lemma 2.10, the truth-value of $A(x_1, \ldots, x_n)$ at a world w in a model $\mathcal{S}, V$ never depends on the values V assigns to variables other than $x_1, \ldots, x_n$, nor on the values V assigns to $x_1, \ldots, x_n$ relative to worlds other than w. Thus we can factor V into a predicate interpretation I and the sequence $V_w(x_1), \ldots, v_w(x_n)$ specifying the values assigned to $x_1, \ldots, x_n$ at w, and once again reformulate our semantics.

> Definition 6.8 (Rank)
> Let ρ be some fixed "alphabetical" order on the variables of $\mathcal{L}$, i.e. a bijection $Var \to \mathbb{N}^+$. I will use v to denote the inverse of ρ, so that v_1 is the alphabetically

first variable, v_2 the second, and so on. The *rank* of an $\mathcal{L}$-formula A is the smallest number $r \in \mathbb{N}$ such that all members of $Var(A)$ have a ρ-value less than r.

DEFINITION 6.9 (FINITARY SATISFACTION)
Let A be an $\mathcal{L}$-formula and r a number greater or equal to A's rank. Let $\mathcal{S} = \langle W, R, U, D, K \rangle$ be a total counterpart structure, I a predicate interpretation on $\mathcal{S}$, w a member of W, and $d_1, \ldots, d_r$ (not necessarily distinct) elements of U_w. Then

$$
\begin{array}{ll}
w, d_1, \ldots, d_r \Vdash_{\mathcal{S},I} Px_1 \ldots x_n & \text{iff } \langle d_{\rho(x_1)}, \ldots, d_{\rho(x_n)} \rangle \in I_w(P). \\
w, d_1, \ldots, d_r \Vdash_{\mathcal{S},I} \neg A & \text{iff } w, d_1, \ldots, d_r \nVdash_{\mathcal{S},I} A. \\
w, d_1, \ldots, d_r \Vdash_{\mathcal{S},I} A \supset B & \text{iff } w, d_1, \ldots, d_r \nVdash_{\mathcal{S},I} A \text{ or } w, d_1, \ldots, d_r \Vdash_{\mathcal{S},I} B. \\
w, d_1, \ldots, d_r \Vdash_{\mathcal{S},I} \forall x A & \text{iff } w, d'_1, \ldots, d'_r \Vdash_{\mathcal{S},I} A \text{ for all } d'_1, \ldots, d'_r \text{ such that} \\
& d'_{\rho(x)} \in D_w \text{ and } d'_i = d_i \text{ for all } i \neq \rho(x). \\
w, d_1, \ldots, d_r \Vdash_{\mathcal{S},I} \langle y : x \rangle A & \text{iff } w, d'_1, \ldots, d'_r \Vdash_{\mathcal{S},I} A \text{ for all } d'_1, \ldots, d'_r \text{ such that} \\
& d'_{\rho(x)} = d_{\rho(y)} \text{ and } d'_i = d_i \text{ for all } i \neq \rho(x). \\
w, d_1, \ldots, d_r \Vdash_{\mathcal{S},I} \Box A & \text{iff } w', d'_1, \ldots, d'_r \Vdash_{\mathcal{S},I} A \text{ for all } w, d'_1, \ldots, d'_r \text{ such that} \\
& \langle w, d_1, \ldots, d_r \rangle R^n_{\mathcal{S}} \langle w', d'_1, \ldots, d'_n \rangle,
\end{array}
$$

where $R^n_{\mathcal{S}}$ is the n-sequential counterpart relation of $\mathcal{S}$.

Notice that this looks just like standard Kripke semantics for a propositional modal language with multiple box operators $\Box$, $\forall v_1, \ldots, \forall v_r$, $\langle v_1 : v_1 \rangle, \langle v_1 : v_2 \rangle, \ldots, \langle v_r : v_r \rangle$, all governed by their own accessibility relation between points of the form $w, d_1, \ldots, d_r$. (As mentioned on p. 22, there is a bit of redundancy here: if we have substitution operators, a single box operator $\forall v_1$ would be enough.)

LEMMA 6.10 (TRUTH AND SATISFACTION)
For any total counterpart structure $\mathcal{S} = \langle W, R, U, D, K \rangle$, total interpretation V on $\mathcal{S}$, world $w \in W$, and formula A,

$$w, V \Vdash_{\mathcal{S}} A \text{ iff } w, d_1, \ldots, d_r \Vdash_{\mathcal{S},I} A,$$

where I is V restricted to predicates, r is a number greater than or equal to A's rank and $d_1 = V_w(v_1), \ldots, d_r = V_w(v_r)$.

PROOF by induction on A.

(i) A is $Px_1 \dots x_n$. $w, V \Vdash_{\mathcal{S}} Px_1 \dots x_n$ iff $\langle V_w(x_1), \dots, V_w(x_n) \rangle \in V_w(P)$ by definition 2.7, iff $\langle d_{\rho(x_1)}, \dots, d_{\rho(x_n)} \rangle \in I_w(P)$, iff $x, d_1, \dots, d_r \Vvdash_{\mathcal{S},I} Px_1 \dots x_n$ by definition 6.23.

(ii) A is $\neg B$. $w, V \Vdash_{\mathcal{S}} \neg B$ iff $w, V \nVdash_{\mathcal{S}} B$ by definition 2.7, iff $w, d_1, \dots, d_r \not\Vvdash_{\mathcal{S},I} B$ by induction hypothesis (since B has rank $\leq r$), iff $w, d_1, \dots, d_r \Vvdash_{\mathcal{S},I} \neg B$ by definition 6.23.

(iii) A is $B \supset C$. $w, V \Vdash_{\mathcal{S}} B \supset C$ iff $w, V \nVdash_{\mathcal{S}} B$ or $w, V \Vdash_{\mathcal{S}} C$ by definition 2.7, iff $w, d_1, \dots, d_r \not\Vvdash_{\mathcal{S},I} B$ or $w, d_1, \dots, d_r \Vvdash_{\mathcal{S},I} C$ by induction hypothesis (since B and C have rank $\leq r$), iff $w, d_1, \dots, d_r \Vvdash_{\mathcal{S},I} B \supset C$ by definition 6.23.

(iv) A is $\forall x B$. By definition 2.7, $w, V \Vdash_{\mathcal{S}} \forall x B$ iff $w, V' \Vdash_{\mathcal{S}} B$ for all existential x-variants V' of V on w. By definition 2.5, V' is an existential x-variant of V on w iff V and V' agree on all predicates, $V'_w(x) \in D_w$ and $V'_w(y) = V_w(y)$ for all $y \neq x$. Take any such V'. By induction hypothesis (since B has rank $\leq r$), $w, V' \Vdash_{\mathcal{S}} B$ iff $w, V'_w(v_1), \dots, V'_w(v_r) \Vdash_{\mathcal{S},I} B$. So $w, V \Vdash_{\mathcal{S}} \forall x B$ iff $w, V'_w(v_1), \dots, V'_w(v_r) \Vdash_{\mathcal{S},I} B$ for all V' such that $V'_w(x) \in D_w$ and $V'_w(y) = V_w(y)$ for all $y \neq x$. In other words, $w, V \Vdash_{\mathcal{S}} \forall x B$ iff $w, d'_1, \dots, d'_r \Vdash_{\mathcal{S},I} B$ for all $d'_1, \dots, d'_r$ such that $d'_{\rho(x)} \in D_w$ and $d'_i = d_i$ for all $i \neq \rho(x)$, iff $w, d_1, \dots, d_r \Vvdash_{\mathcal{S},I} \forall x B$ by definition 6.23.

(v) A is $\langle y : x \rangle B$. By definition 2.7, $w, V \Vdash_{\mathcal{S}} \langle y : x \rangle B$ iff $w, V' \Vdash_{\mathcal{S}} B$ where V' is the x-variant of V on w with $V'_w(x) = V_w(y)$. By induction hypothesis (since B has rank $\leq r$), $w, V' \Vdash_{\mathcal{S}} B$ iff $w, V'_w(v_1), \dots, V'_w(v_r) \Vvdash_{\mathcal{S},I} B$. So $w, V \Vdash_{\mathcal{S}} \langle y : x \rangle B$ iff $w, d'_1, \dots, d'_r \Vvdash_{\mathcal{S},I} B$ for all $d'_1, \dots, d'_r$ such that $d'_{\rho(x)} = d_{\rho(y)}$ and $d'_i = d_i$ for all $i \neq \rho(x)$, iff $w, d_1, \dots, d_r \Vvdash_{\mathcal{S},I} \langle y : x \rangle B$ by definition 6.23.

(vi) A is $\Box B$. By definition 2.7, $w, V \Vdash_{\mathcal{S}} \Box B$ iff $w', V' \Vdash_{\mathcal{S}} B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$. By definition 2.6 and totality of $\mathcal{S}$ and V, the latter holds iff V' and V agree on all predicates and for some $C \in K_{w,w'}$ and all variables x, $V_w(x) C V'_{w'}(x)$. Take any such w', V'. By induction hypothesis (since B has rank $\leq r$), $w', V' \Vdash_{\mathcal{S}} B$ iff $w', V'_w(v_1), \dots, V'_w(v_r) \Vvdash_{\mathcal{S},I} B$. So $w, V \Vdash_{\mathcal{S}} \Box B$ iff $w', V'_w(v_1), \dots, V'_w(v_r) \Vdash_{\mathcal{S},I} B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$, iff $w', d'_1, \dots, d'_r \Vdash_{\mathcal{S},I} B$ for all $w', d'_1, \dots, d'_r$ such that $\langle w, d_1, \dots, d_r \rangle R^n_{\mathcal{S}} \langle w', d'_1, \dots, d'_n \rangle$ by definition 6.6, iff $w, d_1, \dots, d_r \Vvdash_{\mathcal{S},I} \Box B$ by definition 6.23. ■

Clearly the evaluation of a formula whose only box operator is $\Box$ does not depend on the accessibility relations associated with the quantificational box operators ($\forall x$, $\langle y : x \rangle$). As we will see, the same is true if we consider the evaluation of modal *schemas* like $\Box A \supset A$, i.e. the set of formulas that result from $\Box p \supset p$ by uniformly substituting arbitrary $\mathcal{L}$-formulas for p. In this way, every purely modal schema corresponds to a constraint on sequential accessibility relations in counterpart structures.

To make this connection between counterpart models and (unimodal) Kripke models even more explicit, we use the following terminology.

Definition 6.11 (Opaque Propositional Guise)
The *n-ary opaque propositional guise of a total counterpart structure* $\mathcal{S} = \langle W, R, U, D, K \rangle$ is the Kripke frame $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}} \rangle$ where $W^n_{\mathcal{S}}$ is the set of points $\langle w, d_1, \ldots, d_n \rangle$ such that $w \in W, d_1 \in U_w, \ldots, d_n \in U_w$, and $R^n_{\mathcal{S}}$ is the n-accessibility relation for $\mathcal{S}$. The *n-ary opaque propositional guise of a predicate interpretation I* for a language $\mathcal{L}$ on $\mathcal{S}$ is the valuation function V^n on $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}} \rangle$ such that for every quasi-atomic formula $A \in \mathcal{L}$, $V^n(A) = \{\langle w, d_1, \ldots, d_n \rangle : w, d_1, \ldots, d_n \Vdash\!\vdash_{\mathcal{S},I} A\}$.

Lemma 6.12 (Truth-preservation under opaque guises)
For any total counterpart structure $\mathcal{S} = \langle W, R, U, D, K \rangle$, predicate interpretation I on $\mathcal{S}$, world $w \in W$, individuals $d_1, \ldots, d_n \in U_w$, and $\mathcal{L}$-formula A with rank $\leq n$,

$$w, d_1, \ldots, d_n \Vdash\!\vdash_{\mathcal{S},I} A \text{ iff } \mathcal{S}^n, V^n, \langle w, d_1, \ldots, d_n \rangle \Vdash_K A,$$

where $\mathcal{S}^n$ and V^n are the n-ary opaque propositional guises of $\mathcal{S}$ and I respectively.

Proof by induction on A, where quasi-atomic formulas all have complexity zero.

(i) A is quasi-atomic. By definition 6.25, $V^n(A) = \{\langle w, d_1, \ldots, d_n \rangle : w, d_1, \ldots, d_n \Vdash\!\vdash_{\mathcal{S},I} A\}$. So $w, d_1, \ldots, d_n \Vdash\!\vdash_{\mathcal{S},I} A$ iff $\langle w, d_1, \ldots, d_n \rangle \in V^n(A)$, iff $\mathcal{S}^n, V^n, \langle w, d_1, \ldots, d_n \rangle \Vdash_K A$ by definition 6.4.

(ii) A is $\neg B$. $w, d_1, \ldots, d_n \Vdash_{\mathcal{S},I} \neg B$ iff $w, d_1, \ldots, d_n \nVdash_{\mathcal{S},I} B$ by definition 6.23, iff $\mathcal{S}^n, V^n, \langle w, d_1, \ldots, d_n \rangle \nVdash_K B$ by induction hypothesis, iff $\mathcal{S}^n, V^n, \langle w, d_1, \ldots, d_n \rangle \Vdash_K \neg B$ by definition 6.4.

(iii) A is $B \supset C$. $w, d_1, \ldots, d_n \Vdash_{\mathcal{S},I} B \supset C$ iff $w, d_1, \ldots, d_n \nVdash_{\mathcal{S},I} B$ or $w, d_1, \ldots, d_n \Vdash_{\mathcal{S},I} C$ by definition 6.23, iff $\mathcal{S}^n, V^n, \langle w, d_1, \ldots, d_n \rangle \nVdash_K B$ or $\mathcal{S}^n, V^n, \langle w, d_1, \ldots, d_n \rangle \Vdash_K C$ by induction hypothesis, iff $\mathcal{S}^n, V^n, \langle w, d_1, \ldots, d_n \rangle \Vdash_K B \supset C$ by definition 6.4.

(iv) A is $\Box B$. $w, d_1, \ldots, d_n \Vdash_{\mathcal{S},I} \Box B$ iff $w', d'_1, \ldots, d'_n \Vdash\!\vdash_{\mathcal{S},I} B$ for all $w, d'_1, \ldots, d'_r$ such that $\langle w, d_1, \ldots, d_r \rangle R^n_{\mathcal{S}} \langle w', d'_1, \ldots, d'_n \rangle$ by definition 6.23, iff $\mathcal{S}^*, V^*, \langle w', d'_1, \ldots, d'_n \rangle \Vdash_K B$ for all such $w, d'_1, \ldots, d'_r$ by induction hypothesis, iff $\mathcal{S}^*, V^*, \langle w, d_1, \ldots, d_n \rangle \Vdash_K \Box B$ by definition 6.4. ■

Lemma 6.13 (Finite correspondence transfer)
If A is a formula of (unimodal) propositional modal logic that is valid in all and only the Kripke frames in some class F, and $n \in \mathbb{N}$, then the set of $\mathcal{L}$-formulas that result from A by uniformly substituting sentence letters by $\mathcal{L}$-formulas of rank $\leq n$ is positively valid in all and only the total counterpart structures $\mathcal{S} = \langle W, R, U, D, K \rangle$ whose n-ary opaque propositional guise is in F.

Proof Assume A is valid in all and only the Kripke frames in F, and let $p_1, \ldots, p_k$ be the sentence letters in A. Let $\mathcal{S} = \langle W, R, U, D, K\rangle$ be a total counterpart structure whose n-ary opaque propositional guise $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle$ is in F. Suppose for reductio that some formula A' is not (positively) valid in $\mathcal{S}$ that results from A by uniformly substituting the sentence letters p_i in A by $\mathcal{L}$-formulas $p_i^{\mathcal{L}}$ of rank $\leq n$. Then there is an interpretation V on $\mathcal{S}$ and a world $w \in W$ such that $w, V \nVdash_{\mathcal{S}} A'$. By lemma 6.24, this means that $w, d_1, \ldots, d_r \nVdash_{\mathcal{S},I} A'$, where I is V restricted to predicates and $d_1 = V_w(v_1), \ldots, d_r = V_w(v_r)$. By lemma 6.26, it follows that $\mathcal{S}^n, V^n, \langle w, d_1, \ldots, d_n\rangle \nVdash_K A'$. But then $\mathcal{S}^n, V^{n\prime}, \langle w, d_1, \ldots, d_n\rangle \nVdash_K A$, where $V^{n\prime}$ is such that for all sentence letters p_i in A, $V^{n\prime}(p_i) = V^n(p_i^{\mathcal{L}})$. This contradicts the assumption that A is valid in $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle$.

We also have to show that the relevant $\mathcal{L}$-formulas are valid *only* in structures $\mathcal{S}$ whose n-ary opaque propositional guise is in F. So let $\mathcal{S}$ be a structure whose guise $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle$ is not in F. Since A is valid only in frames in F, we know that there is some valuation V on $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle$ and some $\langle w, d_1, \ldots, d_n\rangle \in W^n_{\mathcal{S}}$ such that $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w, d_1, \ldots, d_n\rangle \nVdash_K A$. Let A' result from A by uniformly substituting each sentence letter p_i in A by an n-ary predicate P_i followed by the variables $v_1 \ldots v_n$, with distinct predicates for distinct sentence letters. Let I be a predicate interpretation such that for all P_i and $w' \in W$, $I_{w'}(P_i) = \{\langle d'_1, \ldots, d'_n\rangle : \langle w', d'_1, \ldots, d'_n\rangle \in V(p_i)\}$. A simple induction on subformulas B of A shows that for all $\langle w', d'_1, \ldots, d'_n\rangle \in W^n_{\mathcal{S}}$, $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w', d'_1, \ldots, d'_n\rangle \Vdash_K B$ iff $w', d'_1, \ldots, d'_n \Vvdash B'$, where B' is B with all p_i replaced by $P_i v_1 \ldots v_n$. Given that $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w, d_1, \ldots, d_n\rangle \nVdash_K A$ it follows that $w, d_1, \ldots, d_n \nVvdash_{\mathcal{S},I} A'$.

(i) B is a sentence letter p_i. Then B' is $P_i v_1 \ldots v_n$. $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w', d'_1, \ldots, d'_n\rangle \Vdash_K p_i$ iff $\langle w', d'_1, \ldots, d'_n\rangle \in V(p_i)$ by definition 6.4, iff $\langle d'_1, \ldots, d'_n\rangle \in I_w(P_i)$ by construction of I, iff $w', d'_1, \ldots, d'_n \Vvdash_{\mathcal{S},I} P_i v_1 \ldots v_n$ by definition 6.23.

(ii) B is $\neg C$. Then B' is $\neg C'$, where C' is C with all p_i replaced by $P_i v_1 \ldots v_n$. $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w', d'_1, \ldots, d'_n\rangle \Vdash_K \neg C$ iff $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w', d'_1, \ldots, d'_n\rangle \nVdash_K C$ by definition 6.4, iff $w', d'_1, \ldots, d'_n \nVvdash_{\mathcal{S},I} C'$ by induction hypothesis, iff $w', d'_1, \ldots, d'_n \Vvdash_{\mathcal{S},I} \neg C'$ by definition 6.23.

(iii) B is $C \supset D$. Then B' is $C' \supset D'$, where C' and D' are C and D respectively with all p_i replaced by $P_i v_1 \ldots v_n$. $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w', d'_1, \ldots, d'_n\rangle \Vdash_K C \supset D$ iff $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w', d'_1, \ldots, d'_n\rangle \nVdash_K C$ or $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w', d'_1, \ldots, d'_n\rangle \Vdash_K D$ by definition 6.4, iff $w', d'_1, \ldots, d'_n \nVvdash_{\mathcal{S},I} C'$ or $w', d'_1, \ldots, d'_n \Vvdash_{\mathcal{S},I} D'$ by induction hypothesis, iff $w', d'_1, \ldots, d'_n \Vvdash_{\mathcal{S},I} C' \supset D'$ by definition 6.23.

(iv) B is $\Box C$. Then B' is $\Box C'$, where C' is C with all p_i replaced by $P_i v_1 \ldots v_n$. $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w', d'_1, \ldots, d'_n\rangle \Vdash_K \Box C$ iff $\langle W^n_{\mathcal{S}}, R^n_{\mathcal{S}}\rangle, V, \langle w'', d''_1, \ldots, d''_n\rangle \Vdash_K C$ for all $\langle w'', d''_1, \ldots, d''_n\rangle$ with $\langle w', d'_1, \ldots, d'_n\rangle R^n_{\mathcal{S}} \langle w'', d''_1, \ldots, d''_n\rangle$ by definition 6.4, iff $w'', d''_1, \ldots, d''_n \nVvdash_{\mathcal{S},I} C'$ for all such $\langle w'', d''_1, \ldots, d''_n\rangle$ by induction hypothesis, iff $w', d'_1, \ldots, d'_n \Vvdash_{\mathcal{S},I} \Box C'$ by definition 6.23. ∎

Given that a modal schema restricted to the variables $v_1, \ldots, v_n$ defines a constraint on the n-sequential accessibility relation of a counterpart structure $\mathcal{S}$, the unrestricted schema defines a constraint on all sequential accessibility relations. Let's fold these into a single entity.

Definition 6.14 (Sequential accessibility relation)
The *sequential accessibility relation* $R^*_{\mathcal{S}}$ of a total counterpart structure $\mathcal{S}$ is the union of the n-sequential accessibility relations $R^n_{\mathcal{S}}$ of $\mathcal{S}$, i.e. $R^*_{\mathcal{S}} = \bigcup_{n\in\mathbb{N}} R^n_{\mathcal{S}}$.

Definition 6.15 (Opaque Propositional Guise (part 1))
The *opaque propositional guise of a total counterpart structure* $\mathcal{S} = \langle W, R, U, D, K\rangle$ is the disjoint union of the n-ary opaque propositional guises of $\mathcal{S}$, i.e. the Kripke frame $\langle W^*_{\mathcal{S}}, R^*_{\mathcal{S}}\rangle$ such that $R^*_{\mathcal{S}}$ is the sequential accessibility relation of $\mathcal{S}$ and $W^*_{\mathcal{S}}$ is the set of points w^* such that for some $n \in \mathbb{N}$, world $w \in W$ and individuals $d_1, \ldots, d_n \in U_w$, $w^* = \langle w, d_1, \ldots, d_n\rangle$.

Theorem 6.16 ((Positive) correspondence transfer)
If A is a formula of (unimodal) propositional modal logic that is valid in all and only the Kripke frames in some class F, then the set of $\mathcal{L}$-formulas that result from A by uniformly substituting sentence letters by $\mathcal{L}$-formulas is positively valid in all and only the total counterpart structures $\mathcal{S}$ whose opaque propositional guise is in F.

Proof Since validity in propositional modal logic is preserved under disjoint unions, A is valid in the opaque propositional guise of a structure $\mathcal{S}$ iff A is valid in each n-ary opaque propositional guise of $\mathcal{S}$, with $n \in \mathbb{N}$. (See e.g. [Blackburn et al. 2001], p.140, Theorem 3.14.(i).) Thus the opaque propositional guise of $\mathcal{S}$ is in F iff all n-ary opaque propositional guises of $\mathcal{S}$ are in F.

Assume A is valid in all and only the Kripke frames in F. Let A' be an $\mathcal{L}$-formula that results from A by uniformly substituting sentence letters by $\mathcal{L}$-formulas. By lemma 6.13, A' is (positively) valid in all total counterpart structures $\mathcal{S}$ whose n-ary propositional guise is in F, where n is the rank of A'. Any total structure whose propositional guise is in F satisfies this condition.

To show that the A schema is valid *only* in structures $\mathcal{S}$ whose guise is in F, let $\mathcal{S}$ be a structure whose guise is not in F. Then there is some n such that the n-ary guise of $\mathcal{S}$ is not in F. By lemma 6.13, there is an $\mathcal{L}$-substitution instance A' of A with rank n that is not valid in $\mathcal{S}$. ■

As a union of relations of different arity, R^* is a somewhat gerrymandered entity. It may help to understand statements about R^* as universal statements about its members R^n. For example, the schema $\Box A \supset A$ is valid iff (0) every world can see itself and

(1) every individual at every world is its own counterpart (relative to some counterpart relation), (2) every pair of individuals at every world is its own counterpart, and so on. Each clause (i) covers instances of the schema with i free variables.

In Tarski's semantics for first-order logic, every formula is evaluated relative to an infinite sequence of individuals. Accordingly, the points of evaluation in a Tarski-style first-order modal logic would consist of a world together with an infinite sequence of individuals. Quantifiers and modal operators shift these points of evaluation. Thus one might expect that modal schemas constrain the infinitary accessibility relation R^{∞} that holds between a pair $\langle w, s \rangle$ of a world w and an infinite sequence s on U_w and another such pair $\langle w', s' \rangle$ iff for some $C \in K_{w,w'}$, each element s_i of s has the corresponding element s'_i of s' as C-counterpart. But this is false. For example, $\Box\bot$ is valid in a Kripke frame iff no point can see itself. Suppose in some structure $\mathcal{S}$, no point $\langle w, s \rangle$ stands in R^{∞} to any point $\langle w', s' \rangle$, although for some w, w', wRw'. Since $\neg\langle w, s \rangle R^{\infty} \langle w', s' \rangle$, we know that there is no $C \in K_{w,w'}$ such that $s_1 C s'_1, s_2 C s'_2, \ldots$. In other words, for all $C \in K_{w,w'}$ there is some i such that $\neg s_i C s'_i$. But (as long as $K_{w,w'}$ is infinite) this is compatible with the fact that for all i there is a $C \in K_{w,w'}$ such that $s_1 C s'_1, \ldots, s_i C s'_i$.[6] The general point is that in a Kripke model with points of the form $\langle w, s \rangle$, formulas can express arbitrary properties of world and infinite sequences of individuals; by contrast, formulas of $\mathcal{L}$ can only talk about a finite initial segment of a sequences.

Thus far, I have set aside negative counterpart models. In negative models, variables can be undefined, so sequential accessibility relations must be redefined to hold between pairs of a world w and a possibly *gappy* sequence of individuals from D_w, i.e. a partial function from numbers to members of D_w. Now we could re-run the above arguments, but we can also cut all this short by using lemma 2.16.

Definition 6.17 (Opaque Propositional Guise (part 2))
The *opaque propositional guise of a partial counterpart structure* $\mathcal{S}$ is the opaque propositional guise of its positive transpose $\mathcal{S}^+$. Accordingly, the *sequential accessibility relation* $R^*_{\mathcal{S}}$ *of* $\mathcal{S}$ is the sequential accessibility relation of $\mathcal{S}^+$.

Corollary 6.18 (Negative correspondence transfer)
If A is a formula of (unimodal) propositional modal logic that is valid in all and only the Kripke frames in some class F, then the set of $\mathcal{L}$-formulas that result from A by uniformly substituting sentence letters by $\mathcal{L}$-formulas is negatively valid in all and only the single-domain counterpart structures $\mathcal{S}$ whose opaque propositional guise is in F.

6 Thanks here to Joel David Hamkins and Sam Roberts.

Proof Assume A is valid in all and only the Kripke frames in F. Let A' be an $\mathcal{L}$-formula that results from A by uniformly substituting sentence letters by $\mathcal{L}$-formulas. Let $\mathcal{S}$ be a single-domain counterpart structure whose guise is in F, and let w, V be a world from $\mathcal{S}$ and a partial interpretation on $\mathcal{S}$. By lemma 2.16, $w, V \Vdash_{\mathcal{S}} A'$ iff $w, V^+ \Vdash_{\mathcal{S}^+} A'$. By theorem 6.16, $w, V^+ \Vdash_{\mathcal{S}^+} A'$. So A' is negatively valid in $\mathcal{S}$.

Now let $\mathcal{S}$ be a structure whose guise is not in F. By theorem 6.16, there is a substitution instance A' of A, a world w and a total interpretation V' on $\mathcal{S}^+$ such that $w, V' \nVdash_{\mathcal{S}^+} A'$. Define V so that V and V' agree on all predicates and for all worlds w' and variables x, $V_{w'}(x)$ is $V'_{w'}(x)$ if $V'_{w'}(x) \in D_w$, otherwise $V_{w'}(x)$ is undefined. Then V' is the positive transpose of V. By lemma 2.16, $w, V \Vdash_{\mathcal{S}} A'$ iff $w, V' \Vdash_{\mathcal{S}^+} A'$. So $w, V \nVdash_{\mathcal{S}} A'$. So A' is not negatively valid in $\mathcal{S}$. ■

Here are some applications of theorems 6.16 and 6.18.

Corollary 6.19 (Correspondence facts)

1. The schema $\Box A \supset A$ is valid in a counterpart structure $\mathcal{S}$ iff $R^*_{\mathcal{S}}$ is reflexive.
2. The schema $A \supset \Box\Diamond A$ is valid in a counterpart structure $\mathcal{S}$ iff $R^*_{\mathcal{S}}$ is symmetrical.
3. The schema $\Box A \supset \Box\Box A$ is valid in a counterpart structure $\mathcal{S}$ iff $R^*_{\mathcal{S}}$ is transitive.
4. The schema $\Diamond A \supset \Box\Diamond A$ is valid in a counterpart structure $\mathcal{S}$ iff $R^*_{\mathcal{S}}$ is euclidean.

Let's reformulate the Mates-style semantics of section 2 as a Tarski-style semantics. Interpretation functions V are now split into a *predicate interpretation* I and a *variable interpretation* Σ. For every world w in the relevant structure, I_w is a function mapping predicates P to subsets of U_w^n (with $I_w(=) = \{\langle d, d\rangle : d \in U_w\}$); Σ_w is partial function from Var into U_w.

I assume that the variables (and individual constants) Var in any language of quantified modal logic have a fixed *alphabetical order*, captured by a one-one map $\nu : \mathbb{N} \to Var$. This means that we can effectively identify the variables with the numbers. In particular, I usually won't distinguish between an assignment function $\Sigma_w : Var \to U_w$ and the corresponding function $s : \mathbb{N} \to U_w$ with $s(i) = \Sigma_w(\nu(i))$. Since a partial function from $\mathbb{N}$ to U_w is a (possibly gappy) sequence of members of U_w, I will also call Σ_w a *sequence on U_w*.

Definition 6.20 (Sequence)

For any set X, a *sequence on* X is partial function $s : \mathbb{N} \to X$. If a sequence is undefined for some $i \in \mathbb{N}$, it is called *gappy*, otherwise *non-gappy*.

By lemma 2.11, when we evaluate a formula at a world w, all that matters about the variable interpretation Σ is Σ_w. Hence the satisfaction relation can be defined to hold between formulas on the one hand and a counterpart structure $\mathcal{S}$, a predicate interpretation I on $\mathcal{S}$, a world w in $\mathcal{S}$ and a sequence s on U_w on the other. To define this relation, we first need to redefine *variant* and *image* in terms of sequences (compare definitions 2.5 and 2.6).

Definition 6.21 (Variant)
Let w be a world in a counterpart structure, and s, s' sequences on w. s' is an *x-variant of s on w* iff $s'_y = s_y$ for all $y \neq x$. s' is an *existential x-variant* of s if in addition $s_x \in D_w$.

Definition 6.22 (Image)
Let w, w' be worlds in a counterpart structure $\langle W, R, U, D, K\rangle$, and s, s' sequences on w and w' respectively. s' is a *w'-image of s at w* iff there is a $C \in K_{w,w'}$ such that for every variable x, either $s_x C s'_x$ or there is no d with $s_x C d$ in which case s'_x is undefined.

Definition 6.23 (Satisfaction)
Let $\mathcal{S} = \langle W, R, U, D, K\rangle$ be a counterpart structure, I a predicate interpretation on $\mathcal{S}$, w a member of W, and s any sequence on U_w. Then

$\mathcal{S}, I, w, s \Vdash Px_1 \ldots x_n$	iff $\langle s(x_1), \ldots, s(x_n)\rangle \in I_w(P)$.
$\mathcal{S}, I, w, s \Vdash \neg A$	iff $\mathcal{S}, I, w, s \nVdash A$.
$\mathcal{S}, I, w, s \Vdash A \supset B$	iff $\mathcal{S}, I, w, s \nVdash A$ or $\mathcal{S}, I, w, s \Vdash B$.
$\mathcal{S}, I, w, s \Vdash \forall x A$	iff $\mathcal{S}, I, w, s' \Vdash A$ for all existential x-variants s' of s.
$\mathcal{S}, I, w, s \Vdash \langle y : x\rangle A$	iff $\mathcal{S}, I, w, s' \Vdash A$ for all x-variants s' of s with $s'(x) = s(y)$.
$\mathcal{S}, I, w, s \Vdash \Box A$	iff $\mathcal{S}, I, w', s' \Vdash A$ for all w, s' such that wRw' and s' is a w'-image of s at w.

In the following lemma and afterwards, I will identify interpretations V in the sense of definition 2.3 with pairs of a predicate interpretation I and a variable interpretation Σ.

Lemma 6.24 (Truth and satisfaction)
For any counterpart structure $\mathcal{S} = \langle W, R, U, D, K\rangle$, interpretation $V = \langle I, \Sigma\rangle$ on $\mathcal{S}$, world $w \in W$, and formula A,

$$w, V \Vdash_{\mathcal{S}} A \text{ iff } \mathcal{S}, I, w, \Sigma_w \Vdash A.$$

Proof by induction on A.

(i) A is $Px_1 \ldots x_n$. $w, \langle I, \Sigma\rangle \Vdash_{\mathcal{S}} Px_1 \ldots x_n$ iff $\langle \Sigma_w(x_1), \ldots, \Sigma_w(x_n)\rangle \in I_w(P)$ by definition 2.7, iff $I, w, \Sigma_w \Vdash Px_1 \ldots x_n$ by definition 6.23.

(ii) A is $\neg B$. $w, V \Vdash_{\mathcal{S}} \neg B$ iff $w, V \nVdash_{\mathcal{S}} B$ by definition 2.7, iff $I, w, \Sigma_w \not\Vvdash B$ by induction hypothesis, iff $I, w, \Sigma_w \Vvdash \neg B$ by definition 6.23.

(iii) A is $B \supset C$. $w, V \Vdash_{\mathcal{S}} B \supset C$ iff $w, V \nVdash_{\mathcal{S}} B$ or $w, V \Vdash_{\mathcal{S}} C$ by definition 2.7, iff $I, w, \Sigma_w \not\Vvdash B$ or $I, w, \Sigma_w \Vvdash C$ by induction hypothesis, iff $I, w, \Sigma_w \Vvdash B \supset C$ by definition 6.23.

(iv) A is $\forall x B$. By definition 2.7, $w, V \Vdash_{\mathcal{S}} \forall x B$ iff $w, V' \Vdash_{\mathcal{S}} B$ for all existential x-variants V' of V on w. By definition 2.5, $V' = \langle I', \Sigma' \rangle$ is an existential x-variant V' of $V = \langle I, \Sigma \rangle$ on w iff $I' = I$, Σ'_w is an existential x-variant of Σ_w on w in the sense of definition 6.21, and $\Sigma'_{w'} = \Sigma_{w'}$ for all $w' \neq w$. So by induction hypothesis, $w, I, \Sigma' \Vdash_{\mathcal{S}} B$ for all existential x-variants $\langle I, \Sigma' \rangle$ of $\langle I, \Sigma \rangle$ on w iff $I, w, s' \Vvdash B$ for all existential x-variants s' of Σ_w. So $I, w, \Sigma_w \Vvdash \forall x B$ by definition 6.23.

(v) A is $\langle y : x \rangle B$. Let s' be such that $s'(x) = \Sigma_w(y)$ and for all $z \neq x$, $s'(z) = s(z)$. Let Σ' be such that $\Sigma'_w = s'$ and for all $w \neq w'$, $\Sigma'_{w'} = \Sigma_{w'}$. By induction hypothesis, $w, I, \Sigma' \Vdash_{\mathcal{S}} B$ iff $I, w, s' \Vvdash B$. By definition 3.2, $w, I, \Sigma \Vdash_{\mathcal{S}} \langle y : x \rangle B$ iff $w, I, \Sigma' \Vdash_{\mathcal{S}} B$, and by definition 6.23, $I, w, \Sigma_w \Vvdash \langle y : x \rangle B$ iff $I, w, s' \Vvdash B$. So $w, I, \Sigma \Vdash_{\mathcal{S}} \langle y : x \rangle B$ iff $I, w, \Sigma_w \Vvdash \langle y : x \rangle B$.

(vi) A is $\Box B$. By definition 2.7, $w, V \Vdash_{\mathcal{S}} \Box B$ iff $w', V' \Vdash_{\mathcal{S}} B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$. By definition 2.6, the latter holds iff $V' = \langle I', \Sigma' \rangle$ is such that $I = I'$ and $\Sigma'_{w'}$ is a sequence on $U_{w'}$ such that for some $C \in K_{w,w'}$ and all variables x, either $\Sigma_w(x) C \Sigma'_{w'}(x)$ or there is no d with $\Sigma_w(x) C d$ and $\Sigma'_{w'}(x)$ is undefined – in other words, iff $\Sigma'_{w'}$ is a w'-image of Σ_w at w in the sense of definition 6.22. By induction hypothesis, $w', I, \Sigma' \Vdash_{\mathcal{S}} B$ iff $I, w', \Sigma'_{w'} \Vvdash B$ for all such w', Σ'. So $w', V' \Vdash_{\mathcal{S}} B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$ iff $I, w', s' \Vvdash B$ for all w', s' with wRw' and s' a w'-image of Σ_w at w, iff $I, w, \Sigma_w \Vvdash \Box B$ by definition 6.23. ∎

In a counterpart model, every sentence expresses a property of worlds, in the sense that its truth-value is fixed by choosing a world in the model. On the other hand, the truth-value of $\forall x F x$ (or $\Box F x$) at a world w is not simply a function of the truth-value of Fx at w and other worlds. Rather, whether $\forall x F x$ is true at w depends on whether Fx is true at w relative to every choice of an individual $d \in D_w$ as value of x. (Similarly, whether $\Box F x$ is true at w depends on whether Fx is true at all accessible worlds w' relative to every choice of an x-counterpart as value of x.) In the recursive semantics, formulas are therefore evaluated not only relative to a world, but relative to a world w and a sequence of individuals from w, representing different assignments of values to the variables. The reformulated, Tarski-style semantics makes this explicit. By distinguishing truth from satisfaction, we can still say that truth in a model is only relative to a world: a formula A is *true* at a world w in a model $\langle \mathcal{S}, I, \Sigma \rangle$ iff $\mathcal{S}, I, w, \Sigma_w \Vvdash A$.

It is now very natural to define opaque propositional guises in terms of R^∞:

Definition 6.25 (Opaque Propositional Guise)
The *opaque propositional guise of a counterpart structure* $\mathcal{S} = \langle W, R, U, D, K \rangle$ is the frame $\langle W^*, R^* \rangle$ with

(i) $W^* = \{\langle w, s \rangle : w \in W, s \text{ a sequence on } U_w\}$,

(ii) $R^* = \{\langle\langle w, s\rangle, \langle w', s'\rangle\rangle : wRw'$ and s' is a w'-image of s at w (i.e., for some $C \in K_{w,w'}$ and all $x \in \mathbb{N}$, either $s_x C s'_x$ or there is no d with $s_x C d$ and s'_x is undefined) $\}$.

The *opaque propositional guise of a predicate interpretation* I for a language $\mathcal{L}$ on $\mathcal{S}$ is the valuation function V^* such that for every quasi-atomic formula $A \in \mathcal{L}$,

(iii) $V^*(A) = \{\langle w, s\rangle : \mathcal{S}, I, w, s \Vdash A\}$.

Lemma 6.26 (Truth-preservation under opaque guises)
For any counterpart structure $\mathcal{S} = \langle W, R, U, D, K\rangle$, interpretation $V = \langle I, \Sigma\rangle$ on $\mathcal{S}$, world $w \in W$, sequence s on U_w, and formula A of quantified modal logic (with or without substitution),

$$\mathcal{S}, I, w, s \Vdash A \text{ iff } \mathcal{S}^*, V^*, w^* \Vdash_K A,$$

where $\mathcal{S}^*$ and V^* are the opaque propositional guises of $\mathcal{S}$ and I respectively, and w^* is $\langle w, s\rangle$. Thus in particular, $\mathcal{S}, I, w, \Sigma_w \Vdash A$ iff $\mathcal{S}^*, V^*, \langle w, \Sigma_w\rangle \Vdash_K A$.

Proof by induction on A, where quasi-atomic formulas all have complexity zero.

(i) A is quasi-atomic. By definition 6.25, $V^*(A) = \{\langle w, s\rangle : \mathcal{S}, I, w, s \Vdash A\}$. So $\mathcal{S}, I, w, s \Vdash A$ iff $\langle w, s\rangle \in V^*(A)$, iff $\mathcal{S}^*, V^*, w^* \Vdash_K A$ by definition 6.4.

(ii) A is $\neg B$. $\mathcal{S}, w, I, s \Vdash \neg B$ iff $\mathcal{S}, w, I, s \nVdash B$ by definition 6.23, iff $\mathcal{S}^*, V^*, w^* \nVdash_K B$ by induction hypothesis, iff $\mathcal{S}^*, V^*, w^* \Vdash_K \neg B$ by definition 6.4.

(iii) A is $B \supset C$. $\mathcal{S}, w, I, s \Vdash B \supset C$ iff $\mathcal{S}, w, I, s \nVdash B$ or $\mathcal{S}, w, I, s \Vdash C$ by definition 6.23, iff $\mathcal{S}^*, V^*, w^* \nVdash_K B$ or $\mathcal{S}^*, V^*, w^* \Vdash_K C$ by induction hypothesis, iff $\mathcal{S}^*, V^*, w^* \Vdash_K B \supset C$ by definition 6.4.

(iv) A is $\Box B$. $\mathcal{S}, I, w, s \Vdash \Box B$ iff $\mathcal{S}, I, w', s' \Vdash B$ for all w', s' such that wRw' and s' a w'-image of s on w by definition 6.23, iff $\mathcal{S}, I, w', s' \Vdash B$ for all w' with wRw' and sequences s' on U_w for which there is a $C \in K_{w,w'}$ such that for every variable x, either $s_x C s'_x$ or there is no d with $s_x C d$ in which case s'_x is undefined, by definition 6.22, iff $\mathcal{S}^*, V^*, \langle w', s'\rangle \Vdash_K B$ for all such w', s' by induction hypothesis, iff $\mathcal{S}^*, V^*, w^{*\prime} \Vdash_K B$ for all $w^{*\prime}$ with $w^* R^* w^{*\prime}$ by definition 6.25, iff $\mathcal{S}^*, V^*, w^* \Vdash_K \Box B$ by definition 6.4. ∎

But now we run into the problem I mentioned. We'd like to show that whenever some schema of propositional (uni)modal logic is valid in all and only the Kripke frames in some class $\mathbb{F}$, then the schema is valid in all and only the counterpart structures $\mathcal{S}$ whose opaque propositional guise is in $\mathbb{F}$. Given lemma 6.26, it is easy to show that the schema is valid in *all* structures $\mathcal{S}$ whose opaque propositional guise is in $\mathbb{F}$. (Suppose some instance A of the schema is not valid in $\mathcal{S}$. Then there is an interpretation $\langle I, S\rangle$ on $\mathcal{S}$ such that $\mathcal{S}, I, w, \Sigma_w \nVdash A$. By lemma 6.26, it follows that $\mathcal{S}^*, V^*, \langle w, \Sigma_w\rangle \nVdash_K A$, where $\mathcal{S}^* = \langle W^*, R^*\rangle$ and V^* are the opaque propositional guises of

$\mathcal{S}$ and I respectively. So A is not valid in $\mathcal{S}^*$.) However, we can't show that the schema is valid in *only* structures $\mathcal{S}$ with $\mathcal{S}^* \in \mathbb{F}$.

If $\mathcal{S}^* \notin \mathbb{F}$, we know that there is some instance $A \in \mathcal{L}_0$ of the schema, some valuation V^* on $\mathcal{S}^*$ and some $\langle w, s \rangle \in W^*$ such that $\mathcal{S}^*, V^*, \langle w, s \rangle \nVdash_K A$. Can we use these ingredients to construct a matching formula $A' \in \mathcal{L}$ and interpretation $V = \langle I, \Sigma \rangle$ on $\mathcal{S}$ such that $\mathcal{S}, I, w, \Sigma_w \nVdash A'$? Ideally, we'd find a mapping from the sentence letters P in A to atomic $\mathcal{L}$-sentences $Px_1 \ldots x_n$ such that $\mathcal{S}^*, V^*, \langle w, s \rangle \Vdash_K P$ iff $\mathcal{S}, I, w, s \Vdash Px \ldots x$. Unfortunately, there is no guarantee that the truth-value of P at points $\langle w, s \rangle$ of $\mathcal{S}^*, V^*$ is fixed by the first n elements of s, for any n: P could be true at $\langle w, s \rangle$ and false at $\langle w, s' \rangle$, where s and s' differ only after position n. (In addition, P could be true at $\langle w, s \rangle$ even if s_1 is undefined, in which case $Px_1 \ldots x_n$ can't be satisfied at $\mathcal{S}, I, w, s$.)

More abstractly, the main problem is that sentence letters in $\mathcal{S}^*$ can express arbitrary properties of pairs of a world and an *infinite* (and possibly gappy) sequence of individuals. By contrast, sentences of $\mathcal{L}$ can only take about a finite initial segment of the sequences.

The conjecture could be proved if we could show that whenever a formula is valid in all finite truncations of a propositional guise, then it is valid in the whole guise. But this isn't true. Consider the formula $\Box\bot$, which is valid in $\mathcal{S}^*$ iff no point $\langle w, s \rangle$ in $\mathcal{S}^*$ can see any point $\langle w', s' \rangle$. Assume for some w, w', wRw'. Since $\langle w, s \rangle \not R^* \langle w', s' \rangle$, we know that there is no $C \in K_{w,w'}$ such that $s_1 C s'_1, s_2 C s'_2, \ldots$. In other words, for all $C \in K_{w,w'}$ there is some i such that $s_i \not C s'_i$. But (as long as $K_{w,w'}$ is infinite) this is compatible with the fact that for all i there is a $C \in K_{w,w'}$ such that $s_1 C s'_1, \ldots, s_i C s'_i$. Thus the validity of $\Box\bot$ in $\mathcal{S}^*$ is compatible with its invalidity in all truncated fragments of $\mathcal{S}^*$. Accordingly, there is no transfer from validity in the full propositional guise of $\mathcal{S}$ to validity in $\mathcal{S}$.

There are of course further aspects of structures that can only be captured in quantified modal logic.

Definition 6.27 (Types of structures)
A structure $\mathcal{S} = \langle W, R, U, D, K \rangle$ is

total if every individual at every world has at least one counterpart at every accessible world: whenever wRw' and $d \in U_w$ then there is a $d' \in U_{w'}$ with $\langle d, w \rangle C \langle d', w' \rangle$;

functional if every individual at every world has at most one counterpart at every accessible world;

inversely functional if no two individuals at any world have a common counterpart at some accessible world;

injective if it is both functional and inversely functional.

To get clearer on exactly what it means for a given formula to be valid in a class of structures, let's construct a non-modal language $\mathcal{L}_c$ (the *correspondence language*) to talk about counterpart structures and models. The language is built in stages. Here is the first part.

Definition 6.28 (Simple correspondence language)
Given a language $\mathcal{L}$ of quantified modal logic, the *simple correspondence language* $\mathcal{L}_c$ is a non-modal first-order language with two (distinct) sorts of variables, *Var* for individuals and Var_ω for worlds, both denumerable. Every non-logical predicate P of $\mathcal{L}$ is also a predicate of $\mathcal{L}_c$, where it has a further argument place for a world. In addition, $\mathcal{L}_c$ has the standard binary identity predicate for individuals, a binary predicate E expressing existence of an individual at a world, and, for each $n \geq 0$, a $(2n)+2$-ary predicates R^n applying to a pair of worlds and $2n$ individuals: $\omega R^0 \omega'$ says that ω' is accessible from ω, $\omega x R^1 \omega' x'$ that x' at ω' is a counterpart of x at ω, and so on.

From now on, let $\mathcal{L}$ be any fixed language of quantified modal logic (with or without substitution).

Definition 6.29 (Standard translation)
The *standard translation* $(\cdot)^c$ maps formulas of $\mathcal{L}$ to formulas of $\mathcal{L}_c$ such that

$$\begin{aligned}
(Px_1 \dots x_n)^c &= P\omega x_1 \dots x_n \\
(x{=}y)^c &= x{=}y \\
(\neg A)^c &= \neg(A)^c \\
(A \supset B)^c &= (A)^c \supset (B)^c \\
(\forall x A)^c &= \forall x(E\omega x \supset (A)^c) \\
(\langle y : x\rangle A)^c &= \forall x(y{=}x \supset (A)^c) \\
(\Box A)^c &= \forall \omega' \forall x'_1 \dots x'_n(\omega x_1 \dots x_n R^n \omega' x'_1 \dots x'_n \supset [\omega', x'_1, \dots, x'_n / \omega, x_1, \dots, x_n](A)^c),
\end{aligned}$$

where $x_1, \dots, x_n$ are the free individual variables in $(A)^c$ and $\omega', x'_1, \dots, x'_n$ are (the alphabetically first) distinct variables not in $Var((A)^c)$.

A few examples:

$$\begin{aligned}
(\Box Fx)^c &= \forall \omega' \forall x'(\omega x R \omega' x' \supset F\omega' x'). \\
(\Box x{=}y)^c &= \forall \omega' \forall x' \forall y'(\omega xy R \omega' x' y' \supset x'{=}y'). \\
(\Box\Box Fx)^c &= \forall \omega' \forall x'(\omega x R \omega' x' \supset (\forall \omega'' \forall x''(\omega' x' R \omega'' x'' \supset F\omega' x''))). \\
(\Box Fx \supset Fx)^c &= \forall \omega' \forall x'(\omega x R \omega' x' \supset F\omega' x') \supset F\omega x.
\end{aligned}$$

A few notes on existence and identity. In a counterpart structure, contingent identity is only "superficial" in the sense that an individual d in the domain of a world w never occurs as two individuals at another world w'; if d has two counterparts at w', these two members of D_w are not both d. The correspondence language looks at such a structure from God's point of view;

its quantifiers range over all individuals in all domains, and its statements are true or false not relative to a world but relative to the whole model. Since individuals can have different properties at different worlds, predicates carry an additional argument place for worlds. What about the identity predicate? Suppose $x = y$ is true in $\mathcal{L}$ at w. Does it follow that $x = y$ is true in $\mathcal{L}$ fullstop, or do we have to say $\omega x = y$ to indicate at which world the identity holds (as we do for Gxy which turns into $G\omega xy$)? Well, if $x = y$ is true in $\mathcal{L}$ at w, then 'x' and 'y' denote the same individual in D_w, both in the pointed model of $\mathcal{L}$ centred on w and in the corresponding model of $\mathcal{L}_c$. So $x = y$ is true absolutely. Of course, $x = y$ can be false at another world w' considered as counterfactual, but this is captured by the fact that some counterparts of x and y are non-identical, at most one of which is identical to the original individual d at w. (By contrast, Gxy can be false at w' and true at w even though the only counterpart of d at w' is d itself; so here we must add an additional world variable.)

In $\mathcal{L}_c$, existence of x at w is obviously not expressed by $\exists y(x\!=\!y)$, which is trivially true. We need a basic predicate E to express existence and translate the inner quantifiers of $\mathcal{L}$. Note that $Ex = \exists y(x\!=\!y)$ translates into $\exists y(E\omega y \wedge x\!=\!y)$, which is equivalent to $E\omega x$.

Note that $\mathcal{L}$-statements of the form Ex, which abbreviates $\exists y(x\!=\!y)$, translate into $\exists y(E\omega y \wedge x\!=\!y)$, which is equivalent in first-order logic to $E\omega x$. Since the $\mathcal{L}_c$-quantifiers range over all individuals in all domains, world-relative existence has to be taken as primitive in $\mathcal{L}_c$.

A standard first-order model for $\mathcal{L}_c$ consists of

(i) two non-empty sets W, U,

(ii) a predicate interpretation I mapping each $P \in \mathit{Pred}$ to a subset of $W \times U^n$, E to the set of pairs $\langle w, d \in D_w \rangle$, and each R^n to a subset of $W \times U^n \times W \times U^n$,

(iii) a variable interpretation Σ mapping each world variable $\omega \in \mathit{Var}_\omega$ to a member of W and each individual variable $x \in \mathit{Var}$ to a member of U.

A variable interpretation Σ can be identified with a pair $\langle v, s \rangle$ of non-gappy sequences, one on W and one on U. Every positive counterpart model obviously provides a model of this kind. (In negative models, variables can be empty, which isn't possible in standard first-order logic, so for now we have to restrict ourselves to positive models. I will use W, U, I, v, sA to say that the $\mathcal{L}_c$-formula A is true relative to W, U, I, v, s.

Definition 6.30 (Induced Correspondence Models)
Let $\mathcal{S} = \langle W, R, U, D, K \rangle$ be a total counterpart structure, V a total interpretation $V = \langle I, \Sigma \rangle$ for $\mathcal{L}$ on $\mathcal{S}$, and w^* a world in W. Then the *$\mathcal{L}_c$-model $\langle W', U', I', \Sigma' \rangle$ induced by* $\mathcal{S}, V, w^*$ is defined as follows:

1. $W' = W$,
2. $U' = \bigcup_w U_w$,
3. for all $P \in \mathit{Pred}$, $I'(P) = \{\langle w, d_1, \ldots, d_n \rangle : \langle d_1, \ldots, d_n \rangle \in I_w(P)\}$,
4. for all $n \geq 0$, $I'(R^n) = \{\langle w, d_1, \ldots, d_n, w', d'_1, \ldots, d'_n \rangle : wRw'$ and for some $C \in K_{w,w'}$, $d_1 C d'_1, \ldots, d_n C d'_n \}$,
5. for all $x \in \mathit{Var}$, $\Sigma'(x) = V_{w^*}(x)$,

6. for all $\omega \in Var_\omega$, $\Sigma'(\omega) = w^*$.

Now the crucial fact about the standard translation is that it preserves truth:

THEOREM 6.31 (LOCAL MODEL CORRESPONDENCE)
For any total counterpart structure $\mathcal{S} = \langle W, R, U, D, K \rangle$, total interpretation $V = \langle I, \Sigma \rangle$ of $\mathcal{L}$ on $\mathcal{S}$, world $w \in W$, non-gappy sequence s on U_w and formula A of $\mathcal{L}$,

$$\mathcal{S}, I, w, s \Vdash A \text{ iff } W, U', I', w^\infty, s \Vdash (A)^c,$$

where $\langle W, U', I', \Sigma' \rangle$ is the $\mathcal{L}_c$-model induced by $\mathcal{S}, V, w$ and w^∞ is the sequence on W consisting entirely of infinitely many copies of w. Thus in particular $\mathcal{S}, I, w, \Sigma_w \Vdash A$ iff $W, U', I', \Sigma' \Vdash (A)^c$.

PROOF by induction on A.

(i) A is $Px_1 \dots x_n$. $\mathcal{S}, I, w, s \Vdash Px_1 \dots x_n$ iff $\langle s(x_1), \dots, s(x_n) \rangle \in I_w(P)$ by definition 6.23, iff $\langle w, s(x_1), \dots, s(x_n) \rangle \in I'(P)$ by definition 6.30, iff $W, U', I', w^\infty, s \Vdash P\omega x_1 \dots x_n$ by the semantics of $\mathcal{L}_c$.

(ii) A is $x = y$. $\mathcal{S}, I, w, s \Vdash x = y$ iff $s(x_1) = s(x_n)$ by definition 6.23 and the fact that $I_w(=) = \{\langle d, d \rangle : d \in U_w\}$, iff $W, U', I', w^\infty, s \Vdash x = y$ by the semantics of $\mathcal{L}_c$.

(iii) A is $\neg B$. $\mathcal{S}, I, w, s \Vdash \neg B$ iff $\mathcal{S}, I, w, s \nVdash B$ by definition 6.23, iff $W, U', I', w^\infty, s \nVdash B$ by induction hypothesis, iff $W, U', I', w^\infty, s \Vdash \neg B$ by the semantics of $\mathcal{L}_c$.

(iv) A is $B \supset C$. $\mathcal{S}, I, w, s \Vdash B \supset C$ iff $\mathcal{S}, I, w, s \nVdash B$ or $\mathcal{S}, I, w, s \Vdash C$ by definition 6.23, iff $W, U', I', w^\infty, s \nVdash B$ or $W, U', I', w^\infty, s \Vdash C$ by induction hypothesis, iff $W, U', I', w^\infty, s \Vdash B \supset C$ by the semantics of $\mathcal{L}_c$.

(v) A is $\forall x B$, so $(A)^c = \forall x(E\omega x \supset (B)^c)$. By definition 6.23, $\mathcal{S}, I, w, s \Vdash \forall x B$ iff $\mathcal{S}, I, w, s' \Vdash B$ for all existential x-variants s' of s on w, i.e. for all sequence s' on w such that $s'_x \in D_w$ and $s'_y = s_y$ for all variables $y \neq x$. By induction hypothesis, $\mathcal{S}, I, w, s' \Vdash B$ iff $W, U', I', w^\infty, s' \Vdash (B)^c$, for all such s'. By the semantics of $\mathcal{L}_c$, $W, U', I', w^\infty, s \Vdash \forall x(E\omega x \supset (B)^c)$ iff for all x-variants s' of s, $W, U', I', w^\infty, s' \Vdash E\omega x \supset (B)^c$, i.e. either $W, U', I', w^\infty, s' \nVdash E\omega x$ or $W, U', I', w^\infty, s' \Vdash (B)^c$. Since $W, U', I', w^\infty, s' \Vdash E\omega x$ iff $s'_x \in D_w$, this means that $W, U', I', w^\infty, s \Vdash \forall x(E\omega x \supset (B)^c)$ iff $W, U', I', w^\infty, s' \Vdash (B)^c$ for all existential x-variants of s on w.

(vi) A is $\langle y : x \rangle B$, so $(A)^c = \forall y(y = x \supset (B)^c)$. $\mathcal{S}, I, w, s \Vdash \langle y : x \rangle B$ iff $\mathcal{S}, I, w, s' \Vdash B$ for all x-variants s' of s with $s'_x = s_y$ by definition 6.23, iff $W, U', I', w^\infty, s' \Vdash (B)^c$ for all such s' by induction hypothesis, iff $W, U', I', w^\infty, s \Vdash \forall y(y = x \supset (B)^c)$ by the fact that $W, U', I', w^\infty, s \Vdash \forall y(y = x \supset (B)^c)$ iff $W, U', I', w^\infty, s' \Vdash y = x \supset (B)^c$ for all y-variants s' of s.

(vii) A is $\Box B$. Let $x_1, \ldots, x_n$ be the (distinct) free individual variables in $(B)^c$, so that $(A)^c = \forall\omega'\forall x'_1 \ldots x'_n(\omega x_1 \ldots x_n R\omega' x'_1 \ldots x'_n \supset [\omega', x'_1, \ldots, x'_n/\omega, x_1, \ldots, x_n](B)^c)$, where $\omega', x'_1, \ldots, x'_n$ do not occur in $(B)^c$.

By definitions 6.23 and 6.22 and the fact that $\mathcal{S}$ is total,

(1) $\mathcal{S}, I, w, s \Vdash \Box B$

iff

(2) $\mathcal{S}, I, w', s' \Vdash B$ for all w', s' such that wRw' and for some $C \in K_{w,w'}$, each s_i is C-related to s'_i.

By induction hypothesis, this in turn holds iff

(3) $W, U', I', w'^{\infty}, s' \Vdash (B)^c$ for all w', s' such that wRw' and for some $C \in K_{w,w'}$, each s_i is C-related to s'_i.

By the substitution lemma for first-order logic, $W, U', I', w'^{\infty}, s' \Vdash (B)^c$ iff $W, U', I', v, s'' \Vdash [\omega', x'_1, \ldots, x'_n/\omega, x_1, \ldots, x_n](B)^c$, where s'' is the $x'_1, \ldots, x'_n$-variant of s' with $s''(x'_i) = s'(x_i)$ and v is the ω'-variant of w'^{∞} with $v(\omega') = w'^{\infty}(\omega') = w'$. (Evidently, $v = w'^{\infty}$.) Moreover, in first-order logic, the truth-value of a formula X relative to a variable interpretation Σ does not depend on what Σ assigns to variables that aren't free in X. The only free variables in $[\omega', x'_1, \ldots, x'_n/\omega, x_1, \ldots, x_n](B)^c$ are $\omega', x'_1, \ldots, x'_n$. Thus (3) obtains iff

(4) $W, U', I', v, s' \Vdash [\omega', x'_1, \ldots, x'_n/\omega, x_1, \ldots, x_n](B)^c$ for all ω'-variants v of w^{∞} and $x'_1, \ldots, x'_n$-variants s' of s such that $v_\omega R v_{\omega'}$ and for some $C \in K_{v_\omega, v_{\omega'}}$, $s'(x_1)Cs'(x'_1), \ldots, s'(x_n)Cs'(x'_n)$.

Coming from the other direction, by the semantics of $\mathcal{L}_c$,

(5) $W, U', I', w^{\infty}, s \Vdash \forall\omega'\forall x'_1 \ldots x'_n(\omega x_1 \ldots x_n R^n \omega' x'_1 \ldots x'_n \supset [\omega', x'_1, \ldots, x'_n/\omega, x_1, \ldots, x_n](B)^c)$

iff

(6) $W, U', I', v, s' \Vdash \omega x_1 \ldots x_n R^n \omega' x'_1 \ldots x'_n \supset [\omega', x'_1, \ldots, x'_n/\omega, x_1, \ldots, x_n](B)^c$ for all ω'-variants v of w^{∞} and $x'_1, \ldots, x'_n$-variants s' of s.

By definition 6.30, $W, U', I', v, s' \Vdash \omega x_1 \ldots x_n R^n \omega' x'_1 \ldots x'_n$ iff $v_\omega R v_{\omega'}$ and for some $C \in K_{v_\omega, v_{\omega'}}$, $s'(x_1)Cs'(x'_1), \ldots, s'(x_n)Cs'(x'_n)$. So (6) holds iff

(4) $W, U', I', v, s' \Vdash [\omega', x'_1, \ldots, x'_n/\omega, x_1, \ldots, x_n](B)^c$ for all ω'-variants v of w^{∞} and $x'_1, \ldots, x'_n$-variants s' of s such that $v_\omega R v_{\omega'}$ and for some $C \in K_{v_\omega, v_{\omega'}}$, $s'(x_1)Cs'(x'_1), \ldots, s'(x_n)Cs'(x'_n)$,

which we know is equivalent to (1). ■

For any formula A of $\mathcal{L}$, the corresponding $\mathcal{L}_c$-formula $(A)^c$ contains at most one free world variable, ω, as well as an arbitrary number of free individual variables. Universally binding these variables gives us an $\mathcal{L}_c$-sentence expressing that A holds at all points in a given model. What we're more interested in is validity, where we abstract away not only from the interpretation of the variables, but also of the predicates. To express the validity of, say, $\Box Fx \supset Fx$ in the correspondence language, we have to add second-order quantifiers, so that we can say $\forall F\forall x\forall\omega(\forall\omega'\forall x'(\omega x R^1 \omega' x' \supset F\omega' x') \supset F\omega x)$. A little reflection shows that this is equivalent to $\forall\omega\forall x(\omega x R^1 \omega x)$. (Let ξ range over world-individual pairs. Then $\forall F\forall x\forall\omega(\forall\omega'(\omega x R^1 \omega' x' \supset$

$Fx') \supset Fx)$ can be shortened to $\forall F \forall \xi (\forall \xi' (\xi R^1 \xi' \supset F\xi') \supset F\xi)$, which is the standard translation for the validity of $\Box p \supset p$ in PML.) The validity of $\Box Fx \supset Fx$ in a structure therefore corresponds to reflexivity of R^1. Similarly, validity of $\Box Gxy \supset Gxy$ corresponds to reflexivity of R^2, and so on. (The validity of the *schema* $\Box\Phi \supset \Phi$ cannot be expressed as a single sentence of the correspondence language.)

[To be continued…]

7 Canonical models

We will use the canonical model technique for proving strong completeness. Let me briefly review the basic idea.

Let's hold fixed a particular language $\mathcal{L}$, with or without substitution. A formula A of $\mathcal{L}$ is a *syntactic consequence in a logic* L of a set Γ of $\mathcal{L}$-formulas L, for short: $\Gamma \vdash_L A$, iff there are 0 or more sentences $B_1, \ldots, B_n \in \Gamma$ such that $\vdash_L B_1 \wedge \ldots \wedge B_n \supset A$. (For $n = 0$, $B_1 \wedge \ldots \wedge B_n \supset A$ is A.) Γ is *L-consistent* iff there are no members $A_1, \ldots, A_n$ of Γ such that $\vdash_L \neg(A_1 \wedge \ldots \wedge A_n)$.

A logic L is *weakly complete* with respect to a class of models $\mathbb{M}$ iff L contains every formula valid in $\mathbb{M}$: whenever A is valid in $\mathbb{M}$, then $\vdash_L A$. Equivalently, every formula $A \notin L$ is false at some world in some model in $\mathbb{M}$. L is *strongly complete* with respect to $\mathbb{M}$ iff whenever A is a semantic consequence of a set of formulas Γ in $\mathbb{M}$, then $\Gamma \vdash_L A$. Since $\Gamma \nvdash_L A$ iff $\Gamma \cup \{\neg A\}$ is L-consistent, and A is a semantic consequence of Γ in $\mathbb{M}$ iff no world in any model in $\mathbb{M}$ verifies all members of $\Gamma \cup \{\neg A\}$, this means that L is strongly complete with respect to $\mathbb{M}$ iff for every L-consistent set of formulas Γ there is a world in some model in $\mathbb{M}$ at which all members of Γ are true.

To prove strong completeness, we associate with each logic L a *canonical model* $\mathcal{M}_L$. The worlds of $\mathcal{M}_L$ are construed as maximal L-consistent sets of formulas, and it is shown that a formula A is true at a world in $\mathcal{M}_L$ iff A is a member of that world. Since every L-consistent set of formulas can be extended to a maximal L-consistent set, it follows that every L-consistent set of formulas is verified at some world in $\mathcal{M}_L$, and therefore that L is strongly complete with respect to every model class that contains $\mathcal{M}_L$.

Since A should be true at w iff $A \in w$, the worlds must be *maximal* sets: for every sentence A and world w, either $A \in w$ or $\neg A \in w$. Moreover, if the logic L is sound with respect to $\mathcal{M}_L$, the worlds must be *L-consistent*. Otherwise there would be sentences $A_1, \ldots, A_n \in w$ and hence $w, V \Vdash A_1 \wedge \ldots \wedge A_n$ while $\vdash_L \neg(A_1 \wedge \ldots \wedge A_n)$. However, we will see that the worlds in our canonical models only have an L-consistent fragment.

To ensure that what's true at a world are precisely the formulas it contains, the interpretation V in a canonical model assigns to each variable x at each world w some individual $[x]_w$ and to each predicate P at w the set of n-tuples $\langle [x_1]_w, \ldots, [x_n] \rangle$ such that $Px_1 \ldots x_n \in w$. The customary way to make this work is to identify $[x]_w$ with the class of variables z such that $x = z \in w$. The domains therefore consist of equivalence classes of variables.

A well-known problem now arises from the fact that first-order logic does not require every individual to have a name. This means that there are consistent sets Γ that contain $\exists xFx$ as well as $\neg Fx_i$ for every variable x_i. If we extend Γ to a maximal consistent set w and apply the construction just outlined, then $V_w(F)$ would be the empty set. So we would have $w, V \Vdash \neg\exists xFx$, although $\exists xFx \in w$. To avoid this, one requires that the worlds in a canonical model are all *witnessed* so that whenever an existential formula $\exists xFx$ is in w, then some witnessing instance Fy is in w as well. But we still want the set Γ to be verified at some world. So the worlds are construed in a larger language $\mathcal{L}^*$ that adds infinitely many new variables to the original language $\mathcal{L}$. The new variables may then serve as witnesses. (In the new language, there are again consistent sets of sentences that are not included in any world, but not so in the old language.)

In modal logic, this problem reappears in another form. Assume Γ contains $\Diamond\exists xFx$ but also $\Box\neg Fx_i$ for every $\mathcal{L}^*$-variable x_i. Using Kripke semantic, we then need a world w' accessible from the Γ-world that verifies all instances of $\neg Fx_i$, as well as $\exists xFx$. But then w' isn't witnessed!

One way out is to stipulate that worlds in $\mathcal{M}_L$ must be *modally witnessed* in the sense that whenever $\Diamond\exists xA \in w$, then $\Diamond[y/x]A \in w$ for some (possibly new) variable y. Metaphysically speaking, this means that whenever it is possible that something is so-and-so, then we can point at some object at the actual world which is possibly so-and-so. In single-domain models, this has the unfortunate consequence of rendering the Barcan Formula valid. In dual-domain semantics, the "modal witness" can come from the outer domain, so that $\forall x\Box Fx$ does not entail $\Box\forall xFx$. However, if the relevant logic is classical rather than free, the Barcan Formula still comes out valid, despite the fact that it is not entailed by the principles of classical first-order logic and K. (For the Barcan Formula to be invalid, there must be a world where $\forall x\Box Fx$ is true and $\Box\forall xFx$ false. So $\Diamond\exists x\neg Fx$ is true, and modal witnessing requires $\Diamond\neg Fy$ for some y. But in classical logic, $\forall x\Box Fx$ entails $\Box Fy$, i.e. $\neg\Diamond\neg Fy$.)

Counterpart semantics provides a less drastic response that does not introduce undesired validities and thereby reduce the applicability of the canonical model technique. In counterpart semantics, the truth of $\Box\neg Fx_i$ at some world w only requires that $\neg Fx_i$ is true at w' under all *w'-images* V' of V at w – i.e. under interpretations V' such that $V'_{w'}(x_i)$ is some counterpart of $V_w(x_i) = [x_1]_w$. Suppose, for example, that $[x_1]_w = \{x_1\}$ and each individual $[x_i]_w$ at w has $[x_{i+1}]_{w'}$ as unique counterpart at w'. Then the truth of $\Box\neg Fx_1, \Box\neg Fx_2$, etc. at w only requires that $\neg Fx_1, \neg Fx_2$, etc. are true at w' under an assignment of $[x_2]_{w'}$ to x_1, $[x_3]_{w'}$ to x_2, etc. So $Fx_2, Fx_3, \ldots \in w'$, but the variable x_1 becomes available to serve as a witness for $\Diamond\exists xFx$.

Here we exploit the fact that in counterpart semantics truth at a world "considered as actual" can come apart from truth at a world "considered as counterfactual". In the canonical model, membership in a world only coincides with truth at the world "as actual": $A \in w$ iff $w, V \Vdash A$. If w' contains Fx_1, it follows that $w', V \Vdash Fx_1$. On the other hand, when we look at w' ("as counterfactual") from the perspective of w, we evaluate formulas not by the original interpretation function V, but by an image V' of V. Given that $V'_{w'}(x_1) = [x_2]_{w'}$ and $Fx_2 \notin w'$, $w', V' \nVdash Fx_1$.

Unfortunately, this creates a complication. Suppose we want to show that for every formula A and world w in $\mathcal{M}_L$,

(8) $\Box A \in w$ iff $w, V \Vdash \Box A$,

where V is the interpretation function of $\mathcal{M}_L$. Proceeding by induction on complexity of A, we can assume that for all w,

(9) $A \in w$ iff $w, V \Vdash A$.

In standard Kripke semantics, we now only need to stipulate that w' is accessible from w iff w' contains all A for which w contains $\Box A$. This means that $\Box A \in w$ iff $A \in w'$ for all w-accessible w'; by (9), the latter holds iff $w', V \Vdash A$ for all such w', i.e. iff $w, V \Vdash \Box A$ by the semantics of the box. In counterpart semantics, this line of thought no longer goes through, since $w, V \Vdash \Box A$ only means that A is true at all accessible worlds w' *considered as counterfactual*: $w', V' \Vdash A$. By contrast, (9) only considers worlds *as actual*; it does not tell us that $A \in w$ iff $w, V' \Vdash A$.

Take a concrete example. Suppose w and w' look as follows.

$$w : \{x \neq y, \Box x \neq y, \Box Fx, \Box Fy, \ldots\}$$
$$w' : \{\neg Fx, Fu, Fv, u \neq y, \ldots\}$$

We can tell that $[x]_{w'}$ does not qualify as counterpart of $[x]_w$, since it doesn't satisfy the "modal profile" w attributes to $[x]_w$: w contains $\Box Fx$, so all counterparts of $[x]_w$ should satisfy F. Both $[u]_{w'}$ and $[v]_{w'}$ meet this condition. So we might say that both of them are counterparts of $[x]_w$. But then they should also be counterparts of $[y]_w$, and we get a violation of the "joint modal profile" expressed by $\Box x \neq y$, which requires that no counterpart of $[x]_w$ is identical to any counterpart of $[y]_w$. Structurally, this is the "problem of internal relations" noted in [Hazen 1979]. In response, we assume that there can be multiple counterpart relations linking the individuals at w to those at w'. One relation links $[x]_w$ with $[u]_{w'}$ and $[y]_w$ with $[v]_w$, another $[x]_w$ with $[v]_{w'}$ and $[y]_w$ with $[u]_{w'}$. So $[x]_w$ does have both $[u]_{w'}$ and $[v]_{w'}$ as counterpart, but the pair $\langle [x]_w, [y]_w \rangle$ has only two rather than four counterparts.

It proves convenient to impose a further restriction on canonical counterpart relations: every counterpart relation C in a canonical model corresponds to a transformation τ

so that $[x]_w$ at w has $[x^\tau]_{w'}$ at w' as counterpart (unless $[x^\tau]_{w'}$ is empty – see below). This means that if an individual $[x]_w$ at w has two counterparts at w' under the same counterpart relation, then there must be at least two variables x, y in $[x]_w$, so that $[x^\tau]_{w'}$ and $[y^\tau]_{w'}$ can serve as the two counterparts.

So here is how we might construct the accessibility and counterpart relations. Let's say that w' is *accessibility from w' via* a transformation τ, for short: $w \xrightarrow{\tau} w'$, iff w' contains A^τ whenever w contains $\Box A$. Define wRw' to be true iff $w \xrightarrow{\tau} w'$ for some τ, and let C be a counterpart relation between w and w' iff $C = \{\langle [x]_w, [y]_{w'} \rangle : x^\tau = y\}$ for some $w \xrightarrow{\tau} w'$.

If $w \xrightarrow{\tau} w'$, then $[x]_w$ has $[x^\tau]_{w'}$ as counterpart. Since $[x^\tau]_{w'} = V_{w'}(x^\tau)$ and $V_{w'}(x^\tau) = V^\tau_{w'}(x)$, V^τ is a w'-image of V at w. If V^τ were the only w'-image of V at w, it would be easy to prove that $\Box A \in w$ iff $w, V \Vdash \Box A$. Assume $\Box A \in w$. Then $A^\tau \in w'$ whenever $w \xrightarrow{\tau} w'$. The induction hypothesis (9) tells us that $A^\tau \in w'$ iff $w', V \Vdash A^\tau$. By the transformation lemma (lemma 3.13), $w', V^\tau \Vdash A$ iff $w', V \Vdash A^\tau$: A is true at w' as counterfactual iff A^τ is true at w' as actual. So $\Box A \in w$ iff for all accessible w', there is a transformation τ such that $w', V^\tau \Vdash A$. If V^τ is the only w'-image of V_w at w', it follows that $\Box A \in w$ iff $w, V \Vdash \Box A$.

The remaining problem is that V^τ may not be the only w'-image of V at w – and not only because there can be several τ with $w \xrightarrow{\tau} w'$. For example, assume w contains $x = y$ but not $\Box x = y$. Then there is some world w' and transformation τ such that w' contains $x^\tau \neq y^\tau$. I.e., the individual $[x]_w = [y]_w = \{x, y, \ldots\}$ at w has two τ-induced counterparts at w', $[x^\tau]_{w'}$ and $[y^\tau]_{w'}$, which V^τ assigns (at w') to x and y, respectively. But then there will also be another w'-image of V at w which assigns, for example, $[y^\tau]_{w'}$ to both x and y.

Sometimes this is harmless. Assume w also contains $\Box Fx$. We want to show that $w, V \Vdash \Box Fx$ and thus that $w', V' \Vdash Fx$ for all accessible w' and w'-images V' of V. In particular, at the above world w', both $[x^\tau]_{w'}$ and $[y^\tau]_{w'}$ must fall in $V_{w'}(F)$. Since w contains $\Box Fx$, we know that w' contains Fx^τ, so $[x^\tau]_{w'} \in V_{w'}(F)$ by construction of the canonical interpretation V. What about $[y^\tau]_{w'}$? Well, since w contains $\Box Fx$ and $x = y$, then it also contains $\Box Fy$, by (LL*). So w' contains Fy^τ and $[y^\tau]_{w'} \in V_{w'}(F)$. The upshot is that the truth-value of Fx at w' considered as counterfactual does not vary between V^τ and other w'-images of V. Unfortunately, this is not always the case.

Here is a case where this leads to trouble. Assume we are working with a positive logic without explicit substitution. Again let w contain $x = y$ but not $\Box x = y$, so that for some $w \xrightarrow{\tau} w'$, w' contains $x^\tau \neq y^\tau$. Assume further that w contains $\Box\Diamond x \neq y$. So w' contains $\Diamond x^\tau \neq y^\tau$. To verify $\Box\Diamond x \neq y$ at w, we need to ensure that $w', V' \Vdash \Diamond x \neq y$ for all w'-images V' of V at w, not just for V^τ. Consider the image V' that assigns $[y^\tau]_{w'}$ to both x and y. To ensure that $w', V' \Vdash \Diamond x \neq y$, there must be some $w' \xrightarrow{\sigma} w''$ and $V'_{w'} \triangleright V''_{w''}$ with $w'', V'' \Vdash x \neq y$. $V'_{w'} \triangleright V''_{w''}$ means that there is a σ such that $w' \xrightarrow{\sigma} w''$

and $V'_{w'}(x) C^\sigma V''_{w''}(x)$ for all x, where C^σ is the counterpart relation induced by σ. In other words, we need a transformation σ, world w'' and interpretation V'' such that $w'', V'' \Vdash x \neq y$, where $w' \xrightarrow{\sigma} w''$ and V'' is such that for all x there is a $z \in V'_{w'}(x)$ with $z^\sigma \in V''_{w''}(x)$. Since $V'_{w'}(x) = V'_{w'}(y) = [y^\tau]_{w'}$, this means that $[y^\tau]_{w'}$ must have two counterparts at some w'' relative to the same transformation σ. But so far, we have no guarantee that this is the case. There has to be a variable z other than y^τ such that w' contains $z = y^\tau$ as well as $\Diamond z \neq y^\tau$. The latter ensures that $z^\sigma \neq (y^\tau)^\sigma \in w''$ for some $w \xrightarrow{\sigma} w''$; $[z^\sigma]_{w''}$ and $[(y^\tau)^\sigma]_{w''}$ are then both counterparts at w'' of $[y^\tau]_{w'}$. Hence we complicate the definition of $w \xrightarrow{\tau} w'$. We stipulate that if w' does not contain $z = y^\tau$ and $\Diamond z \neq y^\tau$ for some suitable z, then w' is not τ-accessible from w. In general, if w contains $\Box A$ as well as $x = y$, and x is free in A, then for w' to be accessible from w via τ, we require that it must contain not only A^τ, but also $z = y^\tau$ and $[z/x^\tau]A^\tau$, for some z not free in A^τ.

This requirement might be easier to understand if we consider the same situation in a language with substitution. Here $\Box\Diamond x \neq y$ and $x = y$ entail $\Box\langle y : x\rangle \Diamond x \neq y$ (by (LL_s) and (S$\Box$)). By the original, simple definition of $w \xrightarrow{\tau} w'$, each world w' accessible from w via τ must contain $\langle y^\tau : x^\tau\rangle \Diamond x^\tau \neq y^\tau$. This formula says that $[y^\tau]_{w'}$ has multiple counterparts at some accessible world w''. Before we worry about images other than V^τ, we ought to make sure that $\langle y^\tau : x^\tau\rangle \Diamond x^\tau \neq y^\tau$ is true at w' under V^τ. This requires that there is a variable z other than y^τ such that w' contains $z = y^\tau$ and $\Diamond z \neq y^\tau$. In effect, z is a kind of witness for the substitution formula $\langle y^\tau : x^\tau\rangle \Diamond x^\tau \neq y^\tau$. Just as an existential formula $\exists x A$ must be witnessed by an instance $[z/x]A$, a substitution formula $\langle y : x\rangle A$ must be witnessed by $[z/x]A$ together with $z = y$. Loosely speaking, $\langle y : x\rangle A(x)$ says that y is identical to some x such that $A(x)$. In a canonical model, we want a concrete witness z so that y is identical to z and $A(z)$. y itself may not serve that purpose, because $\langle y : x\rangle A(x)$ does not guarantee $A(y)$.

This requirement of substitutional witnessing entails that if w contains $\Box A$, then any τ-accessible w' contains not only A^τ, but also $z = y^\tau$ and $[z/x^\tau]A^\tau$ (for some suitable z). So we don't need to complicate the accessibility relation. In our example, since w' contains A^τ whenever w contains $\Box A$, w' contains $\langle y^\tau : x^\tau\rangle \Diamond x^\tau \neq y^\tau$, which settles that $[y^\tau]_{w'}$ has two counterparts at some accessible world. Without substitution, $\langle y^\tau : x^\tau\rangle \Diamond x^\tau \neq y^\tau$ is inexpressible (see lemma 3.10). So we have to limit the accessible worlds by requiring membership of the relevant witnessing formulas in addition to A^τ.

Wouldn't it suffice to require that if w contains $x = y$ and $\Box A$, then w' contains $z = y^\tau$ as well as A^τ? Since w also contains $y = x$, it follows that w' also contains $z' = x^\tau$. By (LL*), w' then also contains $[z/y^\tau]A^\tau$ and $[z'/x^\tau]A^\tau$. – Yes, but that's not enough. We get $\{z = y^\tau, z' = x^\tau, \Diamond z' \neq y^\tau, \Diamond x^\tau \neq z\} \subseteq w'$. But this is compatible with $[x^\tau, z']$ not having two counterparts at some accessible world.

We know how truth at a world as actual relates to membership: $w, V \Vdash A$ iff $A \in w$. What about truth at a world as counterfactual? I.e. what about truth at a world w' under some image V' (of some image ...) of V. This becomes important when we want to prove that $\Diamond A \in w$ iff $w, V \Vdash \Diamond A$. The latter tells us that A is true at some w' under some image V'. We would like to infer by induction hypothesis that w' contains some sentence(s) A', and conclude by definition of counterparthood and accessibility that $\Diamond A$ must be in w.

There are several ways to achieve this, corresponding to different choices for the canonical counterparthood and accessibility relations.

The two have to work together. Roughly, the main constraint is that if w contains a formula $\Box A$ with free variables $\underline{x}$, then if wRw' and $(\underline{[x]_w}, w)C(\underline{[y]_{w'}}, w')$, then w' contains $\langle \underline{y} : \underline{x} \rangle A$; and if w contains $\Diamond A$, then there are w' with wRw' and y with $(\underline{[x]_w}, w)C(\underline{[y]_{w'}}, w')$ such that w' contains $\langle \underline{y} : \underline{x} \rangle A$. (That's not quite right in negative semantics, because w can satisfy $\Diamond \neg Ex$ without $[x]_w$ having any counterpart at any accessible world.)

One idea: take two worlds w, w'. If there is any relation $C_{w,w'}$ between the two domains such that whenever $\Box A \in w$ and $\underline{[x]_w} C_{w,w'} \underline{[y]_{w'}}$, then $\langle \underline{y} : \underline{x} \rangle A \in w'$, then say that wRw' and let $C_{w,w'}$ be the counterpart relation restricted to w, w'. But what if there are several such relations? E.g., if the only non-trivial box sentence in w is $\Box(Fx \leftrightarrow \neg Fy)$, and w' contains $Fz, \neg Fu$, we could map x to z and z to u, or x to u and y to z, but we can't let x and y have both z and u as counterparts.

We don't have to read off the counterpart relation from the content of the relevant worlds. We can add it as a primitive extra ingredient into the model. For instance, we could stipulate that $[x_i]_w$ at w always has $[x_{2i}]_{w'}$ at w' as counterpart. This is more or less what I will do.

I used to pair each world with a partial, injective substitution function τ. The idea was that the counterparts of $[x]_w$ at w' are the $[\tau_{w'}(y)]_{w'}$ for all $y \in [x]_w$. To allow $[x]_w = \{x\}$ to have no counterpart at w', $\tau_{w'}$ could be undefined. One challenge is that in classical models, $Ex \in w$ for all x and w. But we want $\Diamond \neg Ex$. So evaluating formulas at worlds as counterfactual seems not just to evaluate them (in the obvious way) under a variable substitution. I defined $(Fx)^\tau = \bot$ if $\tau(x) = undef$, so that $\neg Ex$ could be true under τ. This made it hard to prove the existence lemma. Also, it doesn't generalise to positive systems, where x can have non-trivial properties even at worlds where it doesn't exist. Hence considered as counterfactual, w' must say that non-existent object x is F, non-existent object y isn't F, and so on.

Take a concrete example. w contains $\Box Fx, \Box Fy, \Diamond Gxy, \ldots$, thereby specifying a modal profile for x, y, etc. (more precisely, for $[x]_w, [y]_w$). Another Henkin set w' may contain $Fx, Fz, Gxz, \ldots$. From the perspective of w (i.e. considered as counterfactual), this isn't a state at which e.g. x is F. To say what w' represent about x, we first have to locate x at w', by finding its counterparts.

In Kripke semantics, that's easy: x is always its unique own counterpart at any world; more precisely, if $[x]_w = \{x, y, \ldots\}$, then $[x']_{w'}$ is a counterpart of $[x]_w$ iff $[x']_{w'} = \{x, y, \ldots\}$. Here there is an externally fixed counterpart relation. To allow for contingent identity, we could relax this clause and say that $[x]_w$ has $[x']_{w'}$ as counterpart iff there is some z that occurs both in $[x]_w$ and $[x']_{w'}$. (Now we can have $x = y \in w$ but $x \neq y \in w'$, in which case $[x]_w = [y]_w$ has both $[x]_{w'}$ and $[y]_{w'}$ at w' as counterparts.) To help with the problem of modal witnessing, we could fix a different counterpart relation on which, for example, v_n always has v_{n+1} as counterpart. That's not enough because it only frees a single variable. So we better pick some (arbitrary) substitution σ whose range excludes infinitely many variables and say that $[x]$ always has $[x^\sigma]$ as counterpart, i.e. $[x]_w$ has $[x']_{w'}$ as counterpart iff there is a $z \in [x]_w$ with $z^\sigma \in [x']_{w'}$. It proves convenient to let σ be a transformation τ. This approach works to some extent: one can prove completeness for all four basic logics. But it runs into problems when we look at stronger logics. For example, it is easy to see that $\mathsf{P}+\mathsf{T}$ is valid in a structure iff every world can see itself and all things are their own counterparts. So to prove structure completeness for $\mathsf{P}+\mathsf{T}$, we want the canonical model of $\mathsf{P}+\mathsf{T}$ to be reflexive in this sense. But it won't be. (Let Γ contain $x_1 \neq x_1^\tau$ as well as all $\mathcal{L}$-instances of $\Box A \supset A$. Γ is $\mathsf{P}+\mathsf{T}$-consistent. So it is part of a world w in the canonical model. If the model is reflexive, then for all w, wRw and for all d, $\langle d, w\rangle C \langle d, w\rangle$. On a plausible definition of canonical accessibility, the first condition requires that $A^\tau \in w$ whenever $\Box A \in w$. That already may fail, if e.g. we add to Γ the formulas $\Box Fx_1$ and $\neg Fx_1^\tau$. The second condition requires that $[x_1]_w C [x_1]_w$, i.e. there is some $z \in [x_1]_w$ with $z^\tau \in [x_1]_w$. This is a bit harder to render false explicitly, since we can't add $x_1 \neq z^\tau$ to Γ for all variables z as otherwise Γ contains every variable. However, obviously there are max cons extensions of Γ that contain no identity $x_1 = z^\tau$.)

This shows that we shouldn't define canonical counterparthood in a fixed, external manner. Compare accessibility: whether w' is accessible from w depends on whether it verifies all formulas A (or A^τ) which w claims to be true at all accessible worlds. By analogy, we should say that whether $[x']_{w'}$ is a counterpart of $[x]_w$ is determined by whether $[x']_{w'}$ satisfies the modal profile attributed to $[x]_w$ in w. The above proposal ensured that if $\Box A(x) \in w$, then $A(x^\tau) \in w'$ for accessible w', so that w' verifies that the counterpart $[x^\tau]_{w'}$ of $[x]_w$ satisfies condition A. But this doesn't tell us that *everything* that satisfies $A(x)$ for all $\Box A(x) \in w$ qualifies as counterpart of $[x]_w$. If we had this, it would be easy to show that the CM of $\mathsf{P}+\mathsf{T}$ is reflexive: since every w contains $\Box A(x) \supset A(x)$, $[x]_w$ at w must be a counterpart of itself at w.

So we need to define counterparthood in such a way that we can read off whether $\langle [x]_w, w\rangle C \langle [y]_{w'}, w'\rangle$ by comparing what w says about the boxed properties of x and what w' says about y. It's as if w', considered as counterfactual, were a merely qualitative description of a world (saying that there is some x, some u, some v, etc. satisfying such-and-such conditions), and now we need to figure out which of these x, u, v, etc. qualify as representatives of $[x]_w$, $[y]_w$, $[z]_w$, etc. (Although, of course, we don't stipulate that this is a matter of qualitative similarity: w comes with built-in claims about modal profiles.)

But now there's another problem. w doesn't just constrain the modal profile of individuals one by one, but also in relation to one another. It might say that $\Box Gxy$ or $\Box x \neq y$, etc. And this matters. Suppose w contains $\Box Gxy, \Diamond\neg Gyx, \Box Fx, \Box Fy$, and no other interesting modal statement about x and y. We need an accessible world w' that verifies $Gxy, \neg Gyx, Fx, Fy$ considered as counterfactual. We can easily find a w' containing $Gx'y', \neg Gy'x', Fx', Fy'$ for some variables x', y'. But now which of x', y' is counterpart of x, y, respectively? We must not say that x has both x' and y' as counterpart, and so does y. For then Gyx comes out true at w' as counterfactual.

In philosophy, this problem was noticed by Hazen and Lewis, who argued that we have to allow for counterparts between pairs, triples, etc., rather than just between individuals. This way, we can say that x', y' is a counterpart pair for x, y, and so is y', x', but not x', x' or y', y'. We could go this route...

An alternative to counterpart relations between sequences is to go haecceitistic and introduce "qualitatively indistinguishable", haecceitistic worlds. A haecceitistic world is one in which every individual carries a marker (a "haecceity") which specifies which actual individual it represents. So the world w' containing $Gx'y', \neg Gy'x', Fx', Fy'$ is really two worlds, one marking x' as counterpart of x and y' for y, the other marking them the other way round. This is easily achieved by defining worlds in the canonical model to be pairs of an arbitrary Henkin set and an arbitrary variable transformation. Where previously the pair x, y had two counterparts x', y' and y', x' at the relevant world, we now have two worlds $w, [x', y'/x, y]$ and $w', [y', x'/x, y]$ in each of which there is only one counterpart pair.

Won't that rule out contingent identity and distinctness? Won't worlds with multiple counterparts always be turned into multiple worlds with single counterparts? This happens in standard (extreme) haecceitism in philosophy. But not here. We've already seen the reason: even if we fix a particular transformation τ to find the counterparts of $[x]_w$ at w', we can get multiple counterparts.

A third option is Kutz's...

The fact that we need only consider the image V^τ doesn't mean that there's a "canonical counterpart" for each object. Rather, there's a single such counterpart for each *name* of each object. For instance, $w, V \Vdash x=y \wedge \Diamond x \neq y$ iff there is an accessible world w' with $[x^\tau]_{w'} \neq [y^\tau]_{w'}$.

I never actually prove below that $w, V \Vdash \Box A$ iff $w', V^\tau \Vdash A$ for all wRw'. The interesting direction is the one from right to left, and it follows directly from the truth lemma: assume $w', V^\tau \Vdash A$ for all wRw'; by the truth lemma, $A^\tau \in w'$ for all w'; by def R, $\Box A \in w$; by the truth lemma, $w, V \Vdash \Box A$.

The LTR direction requires showing that V^τ is a w'-image of V at w'. For by def. 2.7, $w, V \Vdash \Box A$ iff $w', V' \Vdash A$ for all wRw' and w'-images V' of V at w, i.e. for all V' such that for all variables z, $V'_{w'}(z)$ is a counterpart at w' of $V_w(z)$ at w, or undefined if there is no such counterpart. By construction of canonical counterparthood, $V_w(z)$ at w has a counterpart at w' iff there is a $z^* \in V_w(z)$ such that $[z^{*\tau}]_{w'} \neq \emptyset$, i.e. such that $z^{*\tau} = z^{*\tau} \in w'$. So V' is such that for all variables z, if there is a $z^* \in V_w(z)$ with $z^{*\tau} = z^{*\tau} \in w'$, then there is a $z^* \in V_w(z)$ with $z^{*\tau} \in V'_{w'}(z)$, else $V'_{w'}(z)$ is undefined. (Note that $V'_{w'}(z)$ is always undefined if $V_w(z)$ is undefined.)

For every variable z, if there is a $z^* \in V_w(z)$ with $z^{*\tau} = z^{*\tau} \in w'$, then $V_w(z)$ is defined and hence $z \in [z]_w = V_w(z)$. Moreover, since by construction of R things cannot go partly out of existence, if there is a $z^* \in V_w(z)$ with $z^{*\tau} = z^{*\tau} \in w'$, then this is true for all $z^* \in V_w(z)$, in particular for z. So then $(z)^\tau \in [z^\tau]_{w'} = V_{w'}(z^\tau) = V^\tau_{w'}(z)$. So V^τ is an interpretation of type V'. I.e. if $w, V \Vdash \Box A$ then $w', V^\tau \Vdash A$. ∎

To a certain extent, substitution operators are redundant in canonical models. Recall that the point of writing $\langle y : x \rangle \Diamond Gxy$ instead of $\Diamond Gyy$ is to introduce a new term x as coreferring with y that won't get captured by the other occurrence of y in the scope of the diamond. If we already have another term x that corefers with y, we could instead have said $\Diamond Gxy$. However, we can't easily get rid of the operator in $\Box \langle y : x \rangle \Diamond x \neq y$.

What we do have is a kind of total commutativity of substitution with the box. It's not that whenever $\langle \underline{x}^\sigma, \underline{x} \rangle \Diamond A \in w$, then $\Diamond \langle \underline{x}^\sigma, \underline{x} \rangle A \in w$, with $\underline{x} = Var(A)$. $\langle y, y : x, y \rangle \Diamond x \neq y \in w$ still doesn't imply $\Diamond \langle y, y : x, y \rangle x \neq y \in w$. Nor can we say $\langle \underline{x}^\sigma, \underline{x} \rangle \Diamond A \leftrightarrow \langle \underline{x}^\rho, \underline{x} \rangle \Diamond A \in w$, for some injective ρ. E.g. consider $\langle y, y : x, y \rangle \Diamond x = y$ in a case where $V_w(y)$ has no dual counterparts. Or consider $\langle y, y : x, y \rangle \Diamond\Diamond x \neq y$ in a case where $V_w(y)$ has no dual counterparts, but some of its counterparts do. What we can say is this:

Let $\underline{x}$ be the free variables in A, and w a world in a canonical model. If $\langle \underline{x^\sigma}, \underline{x}\rangle \Diamond A \in w$, then $\Diamond\langle \underline{x^\rho}, \underline{x}\rangle A \in w$ for some ρ such that for each $x \in \underline{x}$, $x^\sigma = x^\rho \in w$.

Proof By substitutional witnessing, if $\langle \underline{x^\sigma} : \underline{x}\rangle \Diamond A \in w$, then $[\underline{x^\rho}, \underline{x}]\Diamond A \in w$ for some new variables $\underline{x^\rho}$. So $\Diamond[\underline{x^\rho}, \underline{x}]A \in w$, and so $\Diamond\langle \underline{x^\rho} : \underline{x}\rangle A \in w$. ■

(This cannot be proved without substitutional witnessing.) Let's try it with $\langle y : x\rangle$.

Case 1: A is compatible with x being equal to y. I.e. $\langle y : x\rangle \Diamond(A \wedge x = y) \in w$. Then it follows by (S8) that $\Diamond\langle y : x\rangle A \in w$.

Case 2: A is compatible with x being equal to some y' with $y' = y \in w$. I.e. $\langle y : x\rangle \Diamond(A \wedge x = y') \in w$. Since $y = y' \in w$, it follows by LL that $\langle y : y'\rangle\langle y : x\rangle \Diamond(A \wedge x = y') \in w$. And this somehow entails that $\langle y' : x\rangle \Diamond(A \wedge x = y') \in w$. Then by (S8), somehow, $\Diamond\langle y' : x\rangle(A \wedge x = y') \in w$, and so $\Diamond\langle y' : x\rangle A \in w$.

Case 3: A is not compatible with x being equal to any y' with $y' = y \in w$. I.e. $\langle y : x\rangle \Diamond(A \wedge x = y') \notin w$ for all those y'. This can happen for several reasons.

Case 3a: A is compatible with $y' = y''$ for some distinct $y', y'' \in [y]_w$. I.e. $\langle y : x\rangle \Diamond(A \wedge y' = y'') \in w$. (E.g. A entails $y' = y''$ and so $\langle y : x\rangle \Box(A \supset y' = y'') \in w$. We have to liberate one of those terms for $[y]_w$ to make it free for x.) Since $y' = y'' \in w$, it follows by LL that $\langle y' : y''\rangle\langle y : x\rangle \Diamond(A \wedge y' = y'') \in w$. Then by (S8), somehow, $\langle y : x\rangle \Diamond\langle y'' : y'\rangle A \in w$. Now take $\langle y'' : y'\rangle A$ as the new A. Since y'' and y' are distinct, y' does not occur free in $\langle y'' : y'\rangle A$, so $\langle y'' : y'\rangle A$ must be compatible with x being equal to some y'. We continue as in case 2 and reach $\Diamond\langle y' : x\rangle\langle y'' : y'\rangle A \in w$. And so $\Diamond\langle y', y'' : x, y'\rangle A \in w$.

Case 3b: A is not compatible with $y' = y''$ for any distinct $y', y'' \in [y]_w$. I.e. $\langle y : x\rangle \Diamond(A \wedge y' = y'') \notin w$ for all those y', y''. I.e., A entails that all the y, y', y'' are distinct from one another, and that x is distinct from each of them. If y has any counterparts at the A world, it would have to have more counterparts than there are members of $[y]_w$. This is impossible. But how do we show it?

From $\langle y : x\rangle \Diamond(A \wedge y' = y'') \notin w$, we have $\langle y : x\rangle \Box(A \supset y' \neq y'') \in w$. By Cont, $\Box\langle y : x\rangle(A \supset y' \neq y'') \in w$. So for all w', $\langle y^\tau : x^\tau\rangle(A^\tau \supset y'^\tau \neq y''^\tau) \in w$...

This won't work. There's nothing inconsistent about $\{\langle y : x\rangle \Diamond(x \neq y \wedge y = y), y \neq z : z \in Var \backslash \{y\}\}$. I guess I need an extended version of the existence lemma: if $\langle y : x\rangle \Diamond A \in w$, then there are σ, w' s.t. $y = y^\sigma \in w$ and $\langle y^{\sigma\tau} : x^\tau\rangle A^\tau \in w'$. – Well, that won't work either. If $\{\langle y : x\rangle \Diamond(x \neq y \wedge y = y), y \neq z : z \in Var \backslash \{y\}\}$ is L-consistent, it will be contained in some Henkin sets. Looks like I have to restrict the worlds by another clause like the witnessing requirement. Like the witnessing requirement, this clause must not compromise the fact that every L-consistent set can be embedded in a world (so it won't do to say that all worlds must satisfy some formula A that is not a theorem of L). What we need actually looks much like witnessing: whenever $\langle y : x\rangle A(x) \in w$, there is a y' such that $A(y') \in w$.)

On to the details. Let $\mathcal{L}$ be some language with or without substitution and L a positive or strongly negative quantified modal logic in $\mathcal{L}$. Define the extended language $\mathcal{L}^*$ by adding infinitely many new variables Var^+ to $\mathcal{L}$.

Definition 7.1 (Henkin set)

A *Henkin set* for L is a set H of $\mathcal{L}^*$-formulas that is

1. *L-consistent*: there are no $A_1, \ldots, A_n \in H$ with $\vdash_{L(\mathcal{L}^*)} \neg(A_1 \land \ldots \land A_n)$,
2. *maximal*: for every $\mathcal{L}^*$-formula A, H contains either A or $\neg A$,
3. *witnessed*: whenever H contains an existential formula $\exists x A$, then there is a variable $y \notin \mathit{Var}(A)$ such that H contains $[y/x]A$ as well as Ey, and
4. *substitutionally witnessed*: whenever H contains a substitution formula $\langle y : x\rangle A$ as well as $y = y$, then there is a variable $z \notin \mathit{Var}(\langle y : x\rangle A)$ such that H contains $y = z$.

I write $\mathcal{H}_L$ for the class of Henkin sets for L in $\mathcal{L}^*$.

If L is without substitution, the fourth clause is trivial.

Above I said that witnessing a substitution formula $\langle y : x\rangle A$ requires $y = z$ as well as $[z/x]A$, but in fact $y = z$ is enough, since $[z/x]A$ follows from $\langle y : x\rangle A$ and $y = z$ by (LV2) (lemma 5.13). I have also added the condition that H contains $y = y$. In negative logics, a Henkin set may contain $y \neq y$ as well as $\langle y : x\rangle A$; adding $y = z$ would render the set inconsistent.

The requirement of substitutional witnessing generalises to substitution sequences: if H contains a substitution formula $\langle y_1, \ldots, y_n : x_1, \ldots, x_n\rangle A$ as well as $y_i = y_i$ for all y_i in $y_1, \ldots, y_n$, then there are (distinct) new variables $z_1, \ldots, z_n$ such that H contains $y_1 = z_1, \ldots, y_n = z_n$ as well as $[z_1, \ldots, z_n/x_1, \ldots, x_n]A$. This is easily proved by induction on n. Suppose H contains $\langle y_1, \ldots, y_n : x_1, \ldots, x_n\rangle A$. By definition 3.14, this is $\langle y_n : v\rangle\langle y_1, \ldots, y_{n-1} : x_1, \ldots, x_{n-1}\rangle\langle v : x_n\rangle A$, where v is new. Witnessing requires $y_n = z_n \in H$ and (hence) $[z_n/v]\langle y_1, \ldots, y_{n-1} : x_1, \ldots, x_{n-1}\rangle\langle v : x_n\rangle A = \langle y_1, \ldots, y_{n-1} : x_1, \ldots, x_{n-1}\rangle\langle z_n : x_n\rangle A \in H$ for some new z_n. By induction hypothesis, the latter means that there are (distinct) $z_1, \ldots, z_{n-1} \notin \mathit{Var}(\langle z_n : x_n\rangle A)$ such that H contains $y_1 = z_1, \ldots, y_{n-1} = z_{n-1}$ as well as $[z_1, \ldots, z_{n-1}/x_1, \ldots, x_{n-1}]\langle z_n : x_n\rangle A$. Since all the x_i and z_i are pairwise distinct, $[z_1, \ldots, z_{n-1}/x_1, \ldots, x_{n-1}]\langle z_n : x_n\rangle A$ is $\langle z_n : x_n\rangle[z_1, \ldots, z_{n-1}/x_1, \ldots, x_{n-1}]A$. By (SC1), it follows that $[z_n/x_n][z_1, \ldots, z_{n-1}/x_1, \ldots, x_{n-1}]A = [z_1, \ldots, z_n/x_1, \ldots, x_n]A \in H$.

Perhaps a more natural definition for substitution logics would replace clause 3 with the condition that H is

3.$^\diamond$ *witnessed*$^\diamond$: whenever H contains an existential formula $\exists x A$, then there is a variable $y \notin \mathit{Var}(A)$ such that H contains $\langle y : x\rangle A$ as well as Ey.

If $y \notin \mathit{Var}(A)$, then $\vdash_L [y/x]A \leftrightarrow \langle y : x\rangle A$ by (SCI), so the two clauses are equivalent.

For substitutional witnessing, z must be new so that the witnessing formulas $[z/x]A, y=z$ entail the original $\langle y : x\rangle A$. E.g., $[y/x]\Box x=y$, which is $\Box y=y$, does not entail $\langle y : x\rangle\Box x=y$. Novelty of witnesses is used in the CML.

Clause 4 could have been restricted to *non-trivial substitutions*, where y and x are different variables. As it stands, it says that if H contains $\langle x : x\rangle A$ and $x=x$, then H must contain $[z/x]A$ and $x=z$ for some new z. Since $\langle x : x\rangle$ is an empty operator, this seems unnecessary. On the other hand, I don't think there's any harm.

For logics without substitution, we could have added the requirement that H is

4.* *substitutionally witnessed*: whenever H contains a formula A and an identity formula $x=y$, and y is not modally free in A, then there is a variable $z \notin Var(A)$ such that H contains $y=z$. [So y is m.f. in A, and H contains $[z/x]A$, by (LL).]

The extensibility lemma can be adjusted accordingly. But it isn't really necessary, and it doesn't help with the existence lemma.

Definition 7.2 (Variable classes)
For any Henkin set H, define $\sim_H$ to be the binary relation on the variables of $\mathcal{L}^*$ such that $x \sim_H y$ iff $x=y \in H$. For any variable x, let $[x]_H$ be $\{y : x \sim_H y\}$.

Lemma 7.3 ($\sim$-Lemma)
$\sim_H$ is transitive and symmetrical.

Proof Immediate from lemmas 4.10 and 5.12. ■

Definition 7.4 (Accessibility via transformations)
Let w, w' be Henkin sets and τ a transformation.

If $\mathcal{L}$ is with substitution, then w' *is accessible from* w *via* τ, for short: $w \xrightarrow{\tau} w'$, iff for every $\mathcal{L}$-formula A, if $\Box A \in w$, then $A^\tau \in w'$.

If $\mathcal{L}$ is without substitution, then $w \xrightarrow{\tau} w'$ iff for every $\mathcal{L}$-formula A and variables $x_1 \ldots x_n$, $y_1, \ldots, y_n$ $(n \geq 0)$ such that the $x_1 \ldots x_n$ are pairwise distinct members of $Varf(A)$, if $x_1=y_1 \wedge \ldots \wedge x_n=y_n \wedge \Box A \in w$ and $y_1^\tau=y_1^\tau \wedge \ldots \wedge y_n^\tau=y_n^\tau \in w'$, then there are variables $z_1 \ldots z_n \notin Var(A^\tau)$ such that $z_1=y_1^\tau \wedge \ldots \wedge z_n=y_n^\tau \wedge [z_1 \ldots z_n/x_1^\tau \ldots x_n^\tau]A^\tau \in w'$.

This generalises the witnessing requirements on accessible world as explained above to multiple variables and negative logics. (In this case, the generalised version for n

variable pairs is not entailed by the requirement for a single pair, unlike in the case of substitutional witnessing.) Note that the $x_1, \ldots, x_n$ need not be *all* the free variables in A. Also recall from p.9 that a conjunction of zero sentences is the tautology $\top$; so for $n = 0$, the accessibility requirement says that if $\top \land \Box A \in w$ and $\top \in w'$, then $\top \land A^\tau \in w'$ – equivalently: if $\Box A \in w$, then $A^\tau \in w'$.

We could have required that each x_i is distinct from y_i, mirroring the possible restriction of substitutional witnesses to non-trivial substitutions (see p. 106 above). If x_i is y_i, clause 5 says that if $x = x \land \Box A \in w$, and $x^\tau = x^\tau \in w'$, then $z = x^\tau \land [z/x^\tau]A^\tau \in w'$ for some new z. Again, this is pointless but I think it does no harm.

Why the generalisation to multiple variables? We didn't add that to the substitution witnesses. Previously, I had: wRw' iff for all x, y, A,

(i) if $\Box A \in w$, then $A^\tau \in w'$, and

(ii) for any two variables x, y, if $x = y \land \Box A \in w$, then there is a $z \notin \mathit{Var}(A^\tau)$ such that $x^\tau = x^\tau \supset z = x^\tau \land [z/y^\tau]A^\tau \in w'$.

(Note that I had y and x the other way round.) But this won't do. Suppose $x_1 = y_1 \land x_2 = y_2 \land \Box F y_1 y_2 \in w$. By (ii), then $z_1 = x_1^\tau \land F z_1 y_2 \in w'$ and $z_2 = x_2^\tau \land F y_1 z_2 \in w'$, but we also want $F z_1 z_2 \in w'$.

Again, assume $\Box\Diamond(x_1 \neq y_1 \land x_2 \neq y_2) \in w$, and $[x_i]_w = \{x_i, y_i\}$. This means that at every w', all counterparts a, b, c, d of $[x_1], [y_1], [x_2], [y_2]$ are such that some w'' harbours distinct counterparts for a and b and for c and d. The counterparts a of $[x_1]$ at w' are $[x_1^\tau]_{w'}$ and $[y_1^\tau]_{w'}$ (which may or may not be distinct, depending on what else w' says). Similarly for b, c, d. So w' must tell us that

1. some w'' harbours two counterparts for $[x_1^\tau]$ and $[y_1^\tau]$, and for $[x_2^\tau]$ and $[y_2^\tau]$ – i.e. $\Diamond(x_1^\tau \neq y_1^\tau \land x_2^\tau \neq y_2^\tau)$;
2. some w'' harbours two counterparts for $[x_1^\tau]$, and for $[x_2^\tau]$ and $[y_2^\tau]$ – i.e. $z = x_1 \land \Diamond(x_1^\tau \neq z \land x_2^\tau \neq y_2^\tau)$;
3. some w'' harbours two counterparts for $[y_1^\tau]$, and for $[x_2^\tau]$ and $[y_2^\tau]$ – i.e. $z = y_1 \land \Diamond(y_1^\tau \neq z \land x_2^\tau \neq y_2^\tau)$;
4. some w'' harbours two counterparts for $[x_1^\tau]$ and $[y_1^\tau]$, and for $[x_2^\tau]$ – similar;
5. some w'' harbours two counterparts for $[x_1^\tau]$ and $[y_1^\tau]$, and for $[y_2^\tau]$ – i.e. $z_1 \neq x_1^\tau \land z_2 \neq x_2^\tau \land \Diamond(x_1^\tau \neq z \land x_2^\tau \neq z)$;
6. some w'' harbours two counterparts for $[x_1^\tau]$, and for $[x_2^\tau]$ – similar;
7. some w'' harbours two counterparts for $[y_1^\tau]$, and for $[x_2^\tau]$ – similar;
8. some w'' harbours two counterparts for $[x_1^\tau]$, and for $[y_2^\tau]$ – similar;
9. some w'' harbours two counterparts for $[y_1^\tau]$, and for $[y_2^\tau]$ – similar.

Clause (i) ensures the first of these, clause (ii) gives us 2-4, but we also need the others.

Definition 7.5 (Canonical model)
The *canonical model* $\langle W, R, U, D, K, V \rangle$ for L is defined as follows.

1. The *worlds* W are the Henkin sets $\mathcal{H}_L$.
2. For each $w \in W$, the *outer domain* U_w comprises the non-empty sets $[x]_w$, where x is a $\mathcal{L}^*$-variable.

3. For each $w \in W$, the *inner domain* D_w comprises the sets $[x]_w$ for which $Ex \in w$.

4. The *accessibility relation* R holds between world w and world w' iff there is some transformation τ such that $w \xrightarrow{\tau} w'$.

5. C is a *counterpart relation* $\in K_{w,w'}$ iff there is a transformation τ such that (i) $w \xrightarrow{\tau} w'$ and (ii) for all $d \in U_w, d' \in U_{w'}$, dCd' iff there is an $x \in d$ such that $x^\tau \in d'$.

6. The *interpretation* V is such that for all $\mathcal{L}^*$-variables x, $V_w(x)$ is either $[x]_w$ or undefined if $[x]_w = \emptyset$, and for all non-logical predicates P, $V_w(P) = \{\langle [x_1]_w, \ldots, [x_n]_w \rangle : Px_1 \ldots x_n \in w\}$.

Clause 6 takes into account the fact that in negative logics, $\neg Ex$ entails $x \neq y$ for every variable y. So if $\neg Ex \in w$, then $[x]_w$ is the empty set. However, we don't want to say that empty terms denote the empty set (so that $\emptyset \in D_w$, and $x = x$ would have to be true). Instead, the canonical interpretation assigns to each variable x at w the set $[x]_w$, *unless that set is empty*, in which case $V_w(x)$ remains undefined. Similarly, clause 5 ensures that $[x]_w$ at w has $[x^\tau]_{w'}$ as counterpart at w' only if $[x^\tau]_{w'} \neq \emptyset$.

(Note that even if $[x]_w$ at w has no counterpart at w', $w', V' \Vdash Fx$ iff $Fx^\tau \in w'$. So in our canonical model, x^τ still represents x, despite the fact that $V'(x) = undef \neq [x^\tau]_{w'}$. We could alternatively change the definition of τ so that x^τ can be undefined, but this complicates things a lot.)

Here we will need the strongly negative axioms (NA) and (TE). (NA) says that $\neg Ex \supset \Box\neg Ex$. This means that if $\neg Ex \in w$, then $\neg Ex^\tau \in w'$ for all wRw', which guarantees that non-existent objects don't have counterparts. (TE) guarantee that individuals do not go partly out of existence: $x = y \supset \Box(Ex \supset Ey)$. So if $x = y \in w$ and $x^\tau \neq x^\tau \in w'$, then $y^\tau \neq y^\tau \in w'$.

The term '$\{\langle [x_1]_w, \ldots, [x_n]_w \rangle : Px_1 \ldots x_n \in w\}$' in clause 4 is meant to denote the set of n-tuples $\langle d_1, \ldots, d_n \rangle$ for which there are variables $x_1, \ldots, x_n$ such that $d_1 = [x_1]_w$ and $\ldots$ and $d_n = [x_n]_w$ and $Px_1 \ldots x_n \in w$. These d_i are guaranteed to be non-empty because $x_i = x_i \in w$ whenever $Px_1 \ldots x_n \in w$: if L is positive, then $\vdash_L z_i = z_i$ by (=R); if L is negative, then $\vdash_L Pz_1 \ldots z_n \supset Ez_i$ by (Neg) and hence $\vdash_L Pz_1 \ldots z_n \supset z_i = z_i$ by ($\forall$=R) and (FUI*).

Lemma 7.6 (Charge of canonical models)
If L is positive, then the canonical model for L is positive. If L is strongly negative, then the canonical model for L is negative.

PROOF If L is positive, then for all $\mathcal{L}^*$-variables x, every Henkin set for L contains $x{=}x$ (by (=R)). So $V_w(x) = [x]_w$ is never empty. Nor is $[x^\tau]_{w'}$, for any world w' with $w \xrightarrow{\tau} w'$. So everything at any world has a counterpart at every accessible world under every counterpart relation. So the canonical model for a positive logic is positive.

If L is strongly negative, then every Henkin set for L contains $x = x \supset Ex$, for all $\mathcal{L}^*$-variables x (by (Neg)). So $V_w(x) = [x]_w \neq \emptyset$ iff $Ex \in w$, which means that $D_w = U_w$ for all worlds w in the model. So the canonical model for a strongly negative logic is negative. ■

LEMMA 7.7 (EXTENSIBILITY LEMMA)
If Γ is an L-consistent set of $\mathcal{L}^*$-sentences in which infinitely many $\mathcal{L}^*$-variables do not occur, then there is a Henkin set $H \in \mathcal{H}_L$ such that $\Gamma \subseteq H$.

PROOF Let $S_1, S_2, \ldots$ be an enumeration of all $\mathcal{L}^*$-sentences, and $z_1, z_2, \ldots$ an enumeration of the unused $\mathcal{L}^*$-variables in such a way that $z_i \notin Var(S_1 \wedge \ldots \wedge S_i)$. Let $\Gamma_0 = \Gamma$, and define Γ_n for $n \geq 1$ as follows.

(i) If $\Gamma_{n-1} \cup \{S_n\}$ is not L-consistent, then $\Gamma_n = \Gamma_{n-1}$;

(ii) else if S_n is an existential formula $\exists x A$, then $\Gamma_n = \Gamma_{n-1} \cup \{\exists x A, [z_n/x]A, Ez_n\}$;

(iii) else if S_n is a substitution formula $\langle y : x\rangle A$, then $\Gamma_n = \Gamma_{n-1} \cup \{\langle y : x\rangle A, y{=}y \supset y{=}z_n\}$;

(iv) else $\Gamma_n = \Gamma_{n-1} \cup \{S_n\}$.

Define w as the union of all Γ_n. We show that w is a Henkin set for L.

1. w is L-consistent. This is shown by proving that Γ_0 is L-consistent and that whenever Γ_{n-1} is L-consistent, then so is Γ_n. It follows that no finite subset of w is L-inconsistent, and hence that w itself is L-consistent. The base step, that Γ_0 is L-consistent is given by assumption. Now assume (for $n > 0$) that Γ_{n-1} is L-consistent. Then Γ_n is constructed by applying one of (i)–(iv).

   a) If case (i) in the construction applies, then $\Gamma_n = \Gamma_{n-1}$, and so Γ_n is also L-consistent.

   b) Assume case (ii) in the construction applies, and suppose that $\Gamma_n = \Gamma_{n-1} \cup \{\exists x A, [z_n/x]A, Ez\}$ is L-inconsistent. Then there is a finite subset $\{C_1, \ldots, C_m\} \subseteq \Gamma_{n-1}$ such that

      1. $\vdash_L \neg(C_1 \wedge \ldots \wedge C_m \wedge \exists x A \wedge [z_n/x]A \wedge Ez_n)$.

Let $\underline{C}$ abbreviate $C_1 \wedge \ldots \wedge C_m$. Then

2.	$\vdash_L \underline{C} \wedge \exists x A \supset (Ez_n \supset \neg[z_n/x]A)$	(1)
3.	$\vdash_L \forall z_n(\underline{C} \wedge \exists x A) \supset \forall z_n E z_n \supset \forall z_n \neg[z_n/x]A$	(2, (UG), (UD))
4.	$\vdash_L \underline{C} \wedge \exists x A \supset \forall z_n(\underline{C} \wedge \exists x A)$	((VQ), z_n not in Γ_{n-1})
5.	$\vdash_L \underline{C} \wedge \exists x A \supset \forall z_n E z_n \supset \forall z_n \neg[z_n/x]A.$	(3, 4)
6.	$\vdash_L \underline{C} \wedge \exists x A \supset \forall z_n \neg[z_n/x]A.$	(5, ($\forall$Ex))
7.	$\vdash_L \forall z_n \neg[z_n/x]A \leftrightarrow \forall x \neg A$	((AC), $z_n \notin \mathit{Var}(A)$)
8.	$\vdash_L \underline{C} \wedge \exists x A \supset \neg \exists x A.$	(6, 7)

So $\{C_1, \ldots C_m, \exists x A\}$ is not L-consistent, contradicting the assumption that clause (ii) applies.

c) Assume case (iii) in the construction applies (hence L is with substitution), and suppose that $\Gamma_n = \Gamma_{n-1} \cup \{\langle y : x \rangle A, y{=}y \supset y{=}z_n\}$ is L-inconsistent. Then there is a finite subset $\{C_1, \ldots, C_m\} \subseteq \Gamma_{n-1}$ such that

1. $\vdash_L \neg(\underline{C} \wedge \langle y : x \rangle A \wedge (y{=}y \supset y{\neq}z)).$

(As before, $\underline{C}$ is $C_1 \wedge \ldots \wedge C_m$.) But then

2.	$\vdash_L \underline{C} \wedge \langle y : x \rangle A \supset y{=}y \wedge y{\neq}z_n$	(1)
3.	$\vdash_L \langle y : z_n \rangle(\underline{C} \wedge \langle y : x \rangle A \supset y{=}y \wedge y{\neq}z_n)$	(2, (Sub$_\mathrm{s}$))
4.	$\vdash_L \langle y : z_n \rangle(\underline{C} \wedge \langle y : x \rangle A) \supset \langle y : z_n \rangle y{=}y \wedge \langle y : z_n \rangle y{\neq}z_n$	(3, (S$\supset$), (S$\neg$))
5.	$\vdash_L \underline{C} \wedge \langle y : x \rangle A \supset \langle y : z_n \rangle(\underline{C} \wedge \langle y : x \rangle A)$	((VS), z_n not in Γ_{n-1}, S_n)
6.	$\vdash_L \underline{C} \wedge \langle y : x \rangle A \supset \langle y : z_n \rangle y{=}y \wedge \langle y : z_n \rangle y{\neq}z_n$	(4, 5)
7.	$\vdash_L \langle y : z_n \rangle y{\neq}z_n \leftrightarrow y{\neq}y$	(SAt)
8.	$\vdash_L \langle y : z_n \rangle y{=}y \leftrightarrow y{=}y$	(SAt)
9.	$\vdash_L \underline{C} \wedge \langle y : x \rangle A \supset (y{=}y \wedge y{\neq}y).$	(6, 7, 8)

So $\{C_1, \ldots, C_m, \langle y : x \rangle A\}$ is L-inconsistent, contradicting the assumption that clause (iii) applies.

d) Assume case (iv) in the construction applies. Then $\Gamma_n = \Gamma_{n-1} \cup \{S_n\}$ is L-consistent, since otherwise (i) would have applied.

2. w is maximal. Assume some formula S_n is not in w. Then case (i) applied to S_n, so $\Gamma_{n-1} \cup \{S_n\}$ is not L-consistent. So there are $C_1, \ldots, C_m \in \Gamma_{n-1}$ such that $\vdash_L C_1 \wedge \ldots C_m \supset \neg S_n$. Similarly, if $S_k = \neg S_n$ is not in w, then there are $D_1, \ldots, D_l \in \Gamma_{k-1}$ such that $\vdash_L D_1 \wedge \ldots D_l \supset \neg S_k$. By (PC), it follows that there are $C_1, \ldots, C_m, D_1, \ldots D_l \in w$ such that

$$\vdash_L C_1 \wedge \ldots \wedge C_m \wedge D_1 \wedge \ldots \wedge D_l \supset (\neg S_n \wedge \neg\neg S_n).$$

But then w is inconsistent, contradicting what was just shown under 1.

3. w is witnessed. This is guaranteed by clause (ii) of the construction and the fact that the $z_n \notin \mathit{Var}(S_n)$. ■

4. w is substitutionally witnessed. This is guaranteed by clause (iii) and the fact that the $z_n \notin \mathit{Var}(S_n)$. ■

Some comments.

1. I used to have two versions of clause (ii), one for substitution logics and one for non-substitution logics:

   (ii) else if S_n is an existential formula $\exists xA$ and L is with substitution, then $\Gamma_n = \Gamma_{n-1} \cup \{\exists xA, \langle z_n : x\rangle A, Ez_n\}$;

   (ii*) else if S_n is an existential formula $\exists xA$ and L is without substitution, then $\Gamma_n = \Gamma_{n-1} \cup \{\exists xA, [z_n/x]A, Ez_n\}$;

   This version of (ii) is more natural for substitution logics (and it immediately shows that w is witnessed$^\Diamond$), but it isn't really necessary to distinguish the two cases; in particular, the proof above, for (ii*), also works in substitution logics. The proof for (ii) used to go as follows.

   - Assume case (ii) in the construction applied, and suppose $\Gamma_n = \Gamma_{n-1} \cup \{\exists xA, \langle z_n : x\rangle A, Ez\}$ is L-inconsistent. Then there is a finite subset $\{C_1, \ldots, C_m\} \subseteq \Gamma_{n-1}$ such that

     1. $\vdash_L \neg(C_1 \wedge \ldots \wedge C_m \wedge \exists xA \wedge \langle z_n : x\rangle A \wedge Ez_n)$.

     Let $\underline{C}$ abbreviate $C_1 \wedge \ldots \wedge C_m$. Then

2.	$\vdash_L \underline{C} \wedge \exists xA \supset (Ez_n \supset \neg\langle z_n : x\rangle A)$	(1)
3.	$\vdash_L \forall z_n(\underline{C} \wedge \exists xA) \supset \forall z_n Ez_n \supset \forall z_n\neg\langle z_n : x\rangle A$	(2, (UG), (UD))
4.	$\vdash_L \underline{C} \wedge \exists xA \supset \forall z_n(\underline{C} \wedge \exists xA)$	((VQ), z_n not in Γ_{n-1})
5.	$\vdash_L \underline{C} \wedge \exists xA \supset \forall z_n Ez_n \supset \forall z_n\neg\langle z_n : x\rangle A.$	(3, 4)
6.	$\vdash_L \underline{C} \wedge \exists xA \supset \forall z_n\neg\langle z_n : x\rangle A.$	(5, ($\forall$Ex))
7.	$\vdash_L \neg\langle z_n : x\rangle A \supset \langle z_n : x\rangle\neg A$	(S$\neg$)
8.	$\vdash_L \forall z_n\neg\langle z_n : x\rangle A \supset \forall z_n\langle z_n : x\rangle\neg A$	(7, (UG), (UD))
9.	$\vdash_L \forall z_n\langle z_n : x\rangle\neg A \supset \forall x\neg A$	((SBV), $z_n \notin \mathit{Var}(A)$)
10.	$\vdash_L \underline{C} \wedge \exists xA \supset \neg\exists xA.$	(6, 8, 9)

     So $\{C_1, \ldots C_m, \exists xA\}$ is not L-consistent, contradicting the assumption that clause (ii) applies.

2. With the additional constraint (4*) on Henkin sets mentioned above, we would need a further construction clause

   (ii*) else if S_n is a formula $x = y \wedge A$ and L does not contain substitution, then $\Gamma_n = \Gamma_{n-1} \cup \{x = y \wedge A, y = z_n\}$ (and perhaps, redundantly, $\cup\{[z_n/x]A\}$);

   and a further clause in the proof:

- Assume case (ii*) in the construction applied, and suppose that $\Gamma_n = \Gamma_{n-1} \cup \{x{=}y \wedge A, y{=}z_n\}$ is L-inconsistent, although $\Gamma_{n-1} \cup \{x{=}y \wedge A\}$ is L-consistent. Then there are $\underline{C} \in \Gamma_{n-1}$ such that

  $$\vdash_L \underline{C} \wedge x{=}y \wedge A \supset y{\neq}z_n.$$

  But since z_n does not occur in $\Gamma_{n-1}, x{=}y \wedge A$, and thus y is modally free for z_n in the whole formula, we have by (Subs*),

  $$\vdash_L \underline{C} \wedge x{=}y \wedge A \supset y{\neq}y.$$

  But since $\vdash_L x{=}y \supset y{=}y$, then

  $$\vdash_L \underline{C} \wedge A \supset x{\neq}y,$$

  contradicting the assumption that $\Gamma_{n-1} \cup \{x{=}y \wedge A\}$ is L-consistent.

3. Can we ensure substitutional witnessing for languages without substitution? I.e. can we ensure that whenever w contains a formula A and an identity formula $x{=}y$, and y is not modally free in A, then there is a variable $z \notin Var(A)$ such that H contains $y{=}z$ [so that H contains $[z/x]A$, by (LL)]? Not obvious. Whenever we add a new formula S_n, we have to check whether Γ_{n-1} contains some identities $x{=}y$ such that y is not m.f. in S_n; if so, we add a corresponding formula $y{=}z_{n+i}$, increasing the z_n counter for unused new variables. But that is not enough: if we later add $x{=}y$, we have to go through all previous formulas again and add the required substitutional witnesses.

   It might be easier to start off by associating with each $\mathcal{L}^*$-variable y infinitely many $\mathcal{L}^*$-variables z, z' and set $\Gamma_0 = \Gamma \cup \{y{=}z, y{=}z', \ldots\}$. Then whenever $x{=}y \in \Gamma_n$, we must eventually have $x{=}z$ etc. in w as well. And because A can at most contain finitely many variables, if $A \in \Gamma_n$, we therefore have $y{=}z$ in w for some z that doesn't occur in A.

   Let's see how that goes.

   Divide the $\mathcal{L}^*$ variables so that each $\mathcal{L}^*$-variable y gets paired with infinitely many (disjoint) $\mathcal{L}^*$-variables $Eq(y) = \{y', y'', \ldots\}$,, leaving infinitely many further $\mathcal{L}^*$-variables $z_1, z_2, \ldots$ unused. (Think of an infinite 2D matrix into which the $\mathcal{L}^*$-variables are filled diagonally, with column 1 labeled 'unused' and the columns from 2 onwards labeled by the $\mathcal{L}^*$ variables.) Let $Id = \{y{=}z : y \in Var(\mathcal{L}^*), z \in Eq(y)\}$, and define $\Gamma_0 = \Gamma \cup Id \cup \{Ex : x \in Var(\mathcal{L})\}$ if $L = [[\Lambda]_{Nec}, Ex]$, else $\Gamma_0 = \Gamma \cup Id$. Let $S_1, S_2, \ldots$ be an enumeration of all $\mathcal{L}^*$-sentences, and $z_1, z_2, \ldots$ an enumeration of the unused variables from $\mathcal{L}^*$ such that $z_i \notin Var(S_1 \wedge \ldots \wedge S_n)$.

   Now one problem is that we no longer have $z_n \notin Var(\Gamma_{n-1})$, as Γ_0 already contains all variables as part of Id. We could have left $Eq(y)$ empty for pure $\mathcal{L}^*$-variables y. But then how would we ensure substitutional witnessing for formulas with pure $\mathcal{L}^*$-formulas? Could we somehow fill in Id incrementally, or would that lead us back to the starting point?

Lemma 7.8 (Existence Lemma)

If w is a world in the canonical model for L, A a formula with $\Diamond A \in w$, and τ any transformation whose range excludes infinitely many variables of $\mathcal{L}$, then there is a world w' in the model such that $w \xrightarrow{\tau} w'$ and $A^\tau \in w'$.

PROOF I first prove the lemma for logics L with substitution. Let $\Gamma = \{A^\tau\} \cup \{B^\tau : \Box B \in w\}$. E.g., Γ might be $\{\exists x Fx\} \cup \{Fx : x \in \mathit{Var}(\mathcal{L}^*)$. Suppose Γ is not L-consistent. Then there are $B_1^\tau, \ldots, B_n^\tau$ with $\Box B_i \in w$ such that $\vdash_L B_1^\tau \wedge \ldots \wedge B_n^\tau \supset \neg A^\tau$. By definition 3.3, this means that $\vdash_L (B_1 \wedge \ldots B_n \supset \neg A)^\tau$, and so $\vdash_L B_1 \wedge \ldots \wedge B_n \supset \neg A$ by (Sub$^\tau$). By (Nec) and (K), $\vdash_L \Box B_1 \wedge \ldots \wedge \Box B_n \supset \Box\neg A$. But then w contains both $\Diamond A$ and $\neg\Diamond A$, which is impossible because w is L-consistent. So Γ is L-consistent.

Since the range of τ excludes infinitely many variables, by the extensibility lemma, $\Gamma \subseteq H$ for some Henkin set H. Moreover, $w \xrightarrow{\tau} w'$ because $B^\tau \in H$ whenever for $\Box B \in w$.

Now for logics without substitution.

We can't use the same simple construction as before, since the world w' must satisfy the more sophisticated definition of accessibility. Recall that for substitution-free logics, $w \xrightarrow{\tau} w'$ iff for every formula A and variables $x_1 \ldots x_n$, $y_1, \ldots, y_n$ ($n \geq 0$) such that the $x_1 \ldots x_n$ are pairwise distinct members of $\mathit{Varf}(A)$, if $x_1 = y_1 \wedge \ldots \wedge x_n = y_n \wedge \Box A \in w$ and $y_1^\tau = y_1^\tau \wedge \ldots \wedge y_n^\tau = y_n^\tau \in w'$, then there are variables $z_1 \ldots z_n \notin \mathit{Var}(A^\tau)$ such that $z_1 = y_1^\tau \wedge \ldots \wedge z_n = y_n^\tau \wedge [z_1 \ldots z_n / x_1^\tau \ldots x_n^\tau]A^\tau \in w'$.

(Consider again the case where w contains $x = y \wedge \Box\Diamond x \neq y$. Assuming positive models, this says that every counterpart of the individual denoted by x and y has multiple counterparts at some further world. If w' is τ-accessible from w, we therefore want $[y^\tau]_{w'}$ to have multiple counterparts at some world w''. Hence we require that if w contains $x = y \wedge \Box\Diamond x \neq y$, then there is a z such that w' contains $y^\tau = z \wedge \Diamond z \neq y^\tau$.)

The first task in building w' is to find suitable variables z. Since w might contain $\Box x \neq z$ for all z (except x), and thus w' might have to contain all $x^\tau \neq z^\tau$, z must not be in the range of τ. So we choose the "unused" variables for the z role.

We start constructing w' by defining a set that meets the condition imposed by $w \xrightarrow{\tau} w'$. Let's ignore negative models for now, so that we can ignore the requirement that $y^\tau = y^\tau \in w'$. Let $S_1, S_2 \ldots$ enumerate all sentences in w of the form

$$x_1 = y_1 \wedge \ldots \wedge x_n = y_n \wedge \Box B,$$

where $x_1, \ldots, x_n$ are zero or more distinct variables free in B, and each y_i is distinct from x_i. For each $S_i = (x_1 = y_1 \wedge \ldots \wedge x_n = y_n \wedge \Box B)$, let $Z(S_i)$ be a set of unused variables $z_1, \ldots, z_n$ such that $Z(S_i) \cap \bigcup_{j<i} Z(S_j) = \emptyset$, and let S_i^* be the sentence

$$y_1^\tau = z_1 \wedge \ldots \wedge y_n^\tau = z_n \wedge [z_1, \ldots, z_n / x_1^\tau, \ldots, x_n^\tau]B.$$

Then let $\Gamma = \bigcup_i S_i^* \cup \{A^\tau\}$. Γ satisfies the accessibility condition.

We have to show that Γ is consistent (and that infinitely many variables do not occur in it). To bring out the main proof ideas, pretend that Γ contains just one simple S_i^*, so that

$$\Gamma = \{A^\tau, y^\tau = z \wedge [z/x^\tau]B\}.$$

Suppose for reductio that Γ is inconsistent. I.e.

(1) $\qquad \vdash_L \neg(A^\tau \wedge y^\tau = z \wedge [z/x^\tau]B^\tau)$.

Following the proof for substitution logics, we would rephrase this as

$$\vdash_L y^\tau = z \wedge [z/x^\tau]B^\tau \supset \neg A^\tau,$$

then apply (Sub^τ) to remove the τ superscripts, use necessitation and (K), and show that the result contradicts the consistency of w. Unfortunately, z is not in the range of τ, so we can't use (Sub^τ) to remove the superscripts. Worse, we have no guarantee that $\Box(y = z^{-\tau} \wedge [z^{-\tau}/x]B)$ is in w. Indeed, $\Box y = z^{-\tau}$ entails that y has a unique counterpart at every accessible world, which may well be false.

Let's apply necessitation directly to (1):

(2) $\quad \vdash_L \Box\neg(A^\tau \wedge y^\tau = z \wedge [z/x^\tau]B^\tau)$.

Pretend for a moment that z is in the range of τ, so that (2) entails

(2') $\quad \vdash_L \Box\neg(A \wedge y = z^{-\tau} \wedge [z^{-\tau}/x]B)$.

Now consider the following application of (CS):

(3) $\quad \vdash_L x = y \wedge \Box B \supset \Box(y = z^{-\tau} \supset [z^{-\tau}/x]B)$.

Since w contains $x = y \wedge \Box B$, this means that it contains $\Box(y = z^{-\tau} \supset [z^{-\tau}/x]B)$. If we could show that w also contains $\Diamond(A \wedge y = z^{-\tau})$, it would follow that w contains $\Diamond(A \wedge y = z^{-\tau} \wedge [z^{-\tau}/x]B)$. But w is consistent, so $\nvdash_L \Box\neg(A \wedge y = z^{-\tau} \wedge [z^{-\tau}/x]B)$ – in contradiction to (2').

Now we have no guarantee that w contains $\Diamond(A \wedge y = z^{-\tau})$. However, this time no untoward consequences would ensue if such a formula were in w. E.g. this would not entail that y has a unique counterpart at some world. We could therefore add a new variable z^* to the language, stipulate that $\tau(z^*) = z$ (so that z^* is the so-far undefined $z^{-\tau}$), and extend w by the formula $\Diamond(A \wedge y = z^*)$ without losing consistency.

What I do instead is slightly simpler: I make room for the new variable z^* by first applying a transformation σ to w that leaves infinitely many variables unused. By (Sub^τ), this preserves consistency. In fact, I use τ for this purpose. The new variables $\underline{z^*}$ are then the unused variables of τ. The formula we add to w^τ becomes $\Diamond(A^\tau \wedge y^\tau = z)$, where z is unused. Let Δ be this extension of w^τ. We will see that Δ is indeed consistent. Then we can reason as above: since w^τ contains $x^\tau = y^\tau \wedge \Box B^\tau$, by (CS), it contains $\Box(y^\tau = z \supset [z/x^\tau]B)$. So Δ contains $\Diamond(A^\tau \wedge y^\tau = z \wedge [z/x^\tau]B)$ – in contradiction to (2).

OK. Let's remove the simplification that Γ contains just one simple S_i^*. If Γ is inconsistent, there are $S_1^*, \ldots, S_m^*$ such that

(1) $\quad \vdash_L \neg(A^\tau \wedge S_1^* \wedge \ldots \wedge S_m^*)$

By necessitation,

(2) $\quad \vdash_L \Box\neg(A^\tau \wedge S_1^* \wedge \ldots \wedge S_m^*)$

Each S_i^*, remember, has the form

$$y_1^\tau = z_1 \wedge \ldots \wedge y_n^\tau = z_n \wedge [z_1, \ldots, z_n/x_1^\tau, \ldots, x_n^\tau]B.$$

Instead of (CS), we use

(CS$_\mathrm{n}$) $\vdash_L x_1\!=\!y_1\wedge\ldots\wedge x_n\!=\!y_n \supset \Box A \supset \Box(y_1\!=\!z_1\wedge\ldots\wedge y_n\!=\!z_n \supset [z_1,\ldots,z_n/x_1,\ldots,x_n]A)$, provided $z_1,\ldots,z_n$ are not free in A.

The relevant applications have the form

(3) $\vdash_L x_1^\tau\!=\!y_1^\tau \wedge\ldots\wedge x_n^\tau\!=\!y_n^\tau \wedge \Box B^\tau \supset$
$$\Box(y_1^\tau\!=\!z_1\wedge\ldots\wedge y_n^\tau\!=\!z_n \supset [z_1,\ldots,z_n/x_1^\tau,\ldots,x_n^\tau]B).$$

Since w^τ contains all the antecedents, it contains all the consequents. The consequents are just the formulas S_i^*. Let $\Delta = w^\tau\cup\{\Diamond(A^\tau\wedge y_{11}^\tau\!=\!z_{11}\wedge\ldots)\}$. Then Δ implies $\Diamond(A^\tau\wedge S_1^*\wedge\ldots\wedge S_m^*)$. We show that Δ is consistent – in contradiction to (2).

Let $S_1, S_2\ldots$ enumerate all sentences in w of the form

$$x_1\!=\!y_1\wedge\ldots\wedge x_n\!=\!y_n\wedge\Box B,$$

where $x_1,\ldots,x_n$ are zero or more distinct variables free in B. Let U be the "unused" $\mathcal{L}$-variables that are not in the range of τ. Let Z be an infinite subset of U such that $Z\backslash U$ is also infinite. For each $S_i = (x_1\!=\!y_1\wedge\ldots\wedge x_n\!=\!y_n\wedge\Box B)$, let Z_{S_i} be a set of distinct variables $z_1,\ldots,z_n\in Z$ such that $Z_{S_i}\cap\bigcup_{j<i}Z_{S_j}=\emptyset$ (i.e. none of the z_i has been used for any earlier S_j). Abbreviate

$$\begin{aligned}
B_i &=_{df} [z_1,\ldots,z_n/x_1^\tau,\ldots,x_n^\tau]B^\tau;\\
X_i &=_{df} x_1\!=\!y_1\wedge\ldots\wedge x_n\!=\!y_n;\\
Y_i &=_{df} y_1^\tau\!=\!y_1^\tau\wedge\ldots\wedge y_n^\tau\!=\!y_n^\tau;\\
Z_i &=_{df} y_1^\tau\!=\!z_1\wedge\ldots\wedge y_n^\tau\!=\!z_n.
\end{aligned}$$

(S_i^* is $Y_i\supset Z_i\wedge B_i$). (For $n=0$, X_i, Y_i and Z_i are the tautology $\top$, and B_i is B^τ.)

Let $\Gamma^- = \{(Y_i\supset Z_i\wedge B_i) : S_i\in S_1,S_2,\ldots\}$, and let $\Gamma=\Gamma^-\cup\{A^\tau\}$.

Suppose for reductio that Γ is inconsistent. Then there are sentences $(Y_1\supset Z_1\wedge B_1),\ldots,(Y_m\supset Z_m\wedge B_m)\in\Gamma^-$ such that

(1) $\vdash_L \neg(A^\tau\wedge(Y_1\supset Z_1\wedge B_1)\wedge\ldots\wedge(Y_m\supset Z_m\wedge B_m))$.

By (Nec),

(2) $\vdash_L \Box\neg(A^\tau\wedge(Y_1\supset Z_1\wedge B_1)\wedge\ldots\wedge(Y_m\supset Z_m\wedge B_m))$.

Any member $(Y_i\supset Z_i\wedge B_i)$ of Γ^- has the form

$$y_1^\tau\!=\!y_1^\tau\wedge\ldots\wedge y_n^\tau\!=\!y_n^\tau\supset y_1^\tau\!=\!z_1\wedge\ldots\wedge y_n^\tau\!=\!z_n\wedge[z_1,\ldots,z_n/x_1^\tau,\ldots,x_n^\tau]B^\tau.$$

By (CS$_\mathrm{n}$),

(3) $\vdash_L x_1^\tau\!=\!y_1^\tau \wedge\ldots\wedge x_n^\tau\!=\!y_n^\tau \wedge \Box B^\tau \supset$
$$\Box(y_1^\tau\!=\!z_1\wedge\ldots\wedge y_n^\tau\!=\!z_n \supset [z_1,\ldots,z_n/x_1^\tau,\ldots,x_n^\tau]B^\tau).$$

Now w contains $x_1 = y_1 \wedge \ldots \wedge x_n = y_n \wedge \Box B$. So w^τ contains $x_1^\tau = y_1^\tau \wedge \ldots \wedge x_n^\tau = y_n^\tau \wedge \Box B^\tau$, which is the antecedent of (3). The consequent of (3) is $\Box(Z_i \supset B_i)$. Thus

(4) $\quad w^\tau \vdash_L \Box(Z_1 \supset B_1) \wedge \ldots \wedge \Box(Z_m \supset B_m)$.

Let $\Delta = w^\tau \cup \{\Diamond(A^\tau \wedge (Y_1 \supset Z_1) \wedge \ldots \wedge (Y_m \supset Z_m))\}$. So

(5) $\quad \Delta \vdash_L \Box(Z_1 \supset B_1) \wedge \ldots \wedge \Box(Z_m \supset B_m)$;

(6) $\quad \Delta \vdash_L \Diamond(A^\tau \wedge (Y_1 \supset Z_1) \wedge \ldots \wedge (Y_m \supset Z_m))$.

By (K) and (Nec), (5) and (6) yield

(7) $\quad \Delta \vdash_L \Diamond(A^\tau \wedge (Y_1 \supset Z_1 \wedge B_1) \wedge \ldots \wedge (Y_m \supset Z_m \wedge B_m))$.

By (2), it follows that Δ is inconsistent. This means that

(8) $\quad w^\tau \vdash_L \neg\Diamond(A^\tau \wedge (Y_1 \supset Z_1) \wedge \ldots \wedge (Y_m \supset Z_m))$.

Now consider $Z_1 = (y_1^\tau = z_1 \wedge \ldots \wedge y_n^\tau = z_n)$. By ($\mathrm{LL}_\mathrm{n}^*$) (or repeated application of (LL^*)),

(9) $$\vdash_L y_1^\tau = z_1 \wedge \ldots \wedge y_n^\tau = z_n \supset \Box\neg(A^\tau \wedge (y_1^\tau = y_1^\tau \wedge \ldots \wedge y_n^\tau = y_n^\tau \supset y_1^\tau = z_1 \wedge \ldots \wedge y_n^\tau = z_n)) \\ \supset \Box\neg(A^\tau \wedge (y_1^\tau = y_1^\tau \wedge \ldots \wedge y_n^\tau = y_n^\tau \supset y_1^\tau = y_1^\tau \wedge \ldots \wedge y_n^\tau = y_n^\tau)),$$

because the z_i are not free in A^τ. In other words (and dropping the tautologous conjunct at the end),

(10) $\quad \vdash_L Z_1 \supset \Box\neg(A^\tau \wedge (Y_1 \supset Z_1)) \supset \Box\neg A^\tau$.

By the same reasoning,

(11) $\quad \vdash_L Z_1 \wedge \ldots \wedge Z_m \supset \Box\neg(A^\tau \wedge (Y_1 \supset Z_1) \wedge \ldots \wedge (Y_m \supset Z_m)) \supset \Box\neg A^\tau$.

By (PC), (Nec) and (K), this means

(12) $\quad \vdash_L Z_1 \wedge \ldots \wedge Z_m \supset \Diamond A^\tau \supset \Diamond(A^\tau \wedge (Y_1 \supset Z_1) \wedge \ldots \wedge (Y_m \supset Z_m))$.

Since $w^\tau \vdash_L \Diamond A^\tau$, (8) and (12) together entail

(13) $\quad w^\tau \vdash_L \neg(Z_1 \wedge \ldots \wedge Z_m)$.

So there are $C_1, \ldots, C_k \in w$ such that

(14) $\quad \vdash_L C_1^\tau \wedge \ldots \wedge C_k^\tau \supset \neg(Z_1 \wedge \ldots \wedge Z_m)$.

Each Z_i has the form $y_1^\tau = z_1 \wedge \ldots \wedge y_n^\tau = z_n$. All the z_i are pairwise distinct, and none of them occur in $C_1^\tau \wedge \ldots \wedge C_k^\tau$ (because the z_i are not in the range of τ) nor in any other Z_i. By (Sub^*), we can therefore replace each z_i in (14) by the corresponding y_i^τ, turning Z_i into Y_i:

(15) $\quad \vdash_L C_1^\tau \wedge \ldots \wedge C_k^\tau \supset \neg(Y_1 \wedge \ldots \wedge Y_m)$.

For any $Y_i = (y_1^\tau = y_1^\tau \land \ldots \land y_n^\tau = y_n^\tau)$, X_i is a sentence of the form $x_1 = y_1 \land \ldots \land x_n = y_n$. So X_i^τ is $x_1^\tau = y_1^\tau \land \ldots \land x_n^\tau = y_n^\tau$, and $\vdash_L X_i^\tau \supset Y_i$ by either (=R) or (Neg) and (∀=R). So (15) entails

(16) $\quad \vdash_L C_1^\tau \land \ldots \land C_k^\tau \supset \neg(X_1^\tau \land \ldots \land X_m^\tau)$.

Thus by (Sub$^\tau$),

(17) $\quad \vdash_L C_1 \land \ldots \land C_k \supset \neg(X_1 \land \ldots \land X_m)$.

Since $\{C_1, \ldots, C_k, X_1, \ldots, X_m\} \subseteq w$, it follows that w is inconsistent. Which it isn't. This completes the reductio.

So Γ is consistent. Since the infinitely many variables in $U \backslash Z$ do not occur in Γ, lemma 7.7 guarantees that $\Gamma \subseteq w'$ for some world w' in the canonical model for L. And of course, Γ was constructed so that w' satisfies the condition in definition 7.4 for $w \xrightarrow{\tau} w'$. This requires that for every formula B and variables $x_1 \ldots x_n$, $y_1, \ldots, y_n$ such that the $x_1 \ldots x_n$ are zero or more pairwise distinct members of $\mathit{Varf}(B)$, if $x_1 = y_1 \land \ldots \land x_n = y_n \land \Box B \in w$ and $y_1^\tau = y_1^\tau \land \ldots \land y_n^\tau = y_n^\tau \in w'$, then there are variables $z_1 \ldots z_n \notin \mathit{Var}(B^\tau)$ such that $z_1 = y_1^\tau \land \ldots \land z_n = y_n^\tau \land [z_1 \ldots z_n / x_1^\tau \ldots x_n^\tau]B^\tau \in w'$. By construction of Γ, whenever $x_1 = y_1 \land \ldots \land x_n = y_n \land \Box B \in w$, then there are suitable $z_1, \ldots, z_n$ such that $y_1^\tau = y_1^\tau \land \ldots \land y_n^\tau = y_n^\tau \supset y_1^\tau = z_1 \land \ldots \land y_n^\tau = z_n \land [z_1, \ldots, z_n / x_1^\tau, \ldots, x_n^\tau]B^\tau \in w'$. So if $y_1^\tau = y_1^\tau \land \ldots \land y_n^\tau = y_n^\tau \in w'$, then $y_1^\tau = z_1 \land \ldots \land y_n^\tau = z_n \land [z_1, \ldots, z_n / x_1^\tau, \ldots, x_n^\tau]B^\tau \in w'$. ■

Lemma 7.9 (Truth Lemma)
For any sentence A and world w in the canonical model $\mathcal{M}_L = \langle W, R, U, D, K, V \rangle$ for L,

$$w, V \Vdash A \text{ iff } A \in w.$$

Proof by induction on A.

1. A is $Px_1 \ldots x_n$. $w, V \Vdash Px_1 \ldots x_n$ iff $\langle V_w(x_1), \ldots, V_w(x_n) \rangle \in V_w(P)$ by definition 2.7. By construction of V_w (definition 7.5), $V_w(x_i)$ is $[x_i]_w$ or undefined if $[x_i]_w = \emptyset$, and $V_w(P) = \{\langle [z_1]_w, \ldots, [z_n]_w \rangle : Pz_1 \ldots z_n \in w\}$. (For non-logical P, this is directly given by definition 7.5; for the identity predicate, $V_w(=)$ is $\{\langle d, d \rangle : d \in U_w\}$ by definition 2.7, which equals $\{\langle [z]_w, [z]_w \rangle : z = z \in w\} = \{\langle [z_1]_w, [z_2]_w \rangle : z_1 = z_2 \in w\}$ because the members of U_w are precisely the non-empty sets $[z]_w$.) (For the last step: every $\langle [z_1]_w, [z_2]_w \rangle$ with $z_1 = z_2 \in w$ is a $\langle [z]_w, [z]_w \rangle$ with $z = z \in w$, and every $\langle [z]_w, [z]_w \rangle$ with $z = z \in w$ is a $\langle [z_1]_w, [z_2]_w \rangle$ with $z_1 = z_2 \in w$.)

   Now if $\langle V_w(x_1), \ldots, V_w(x_n) \rangle \in V_w(P)$, then $\langle [x_1]_w, \ldots, [x_n]_w \rangle \in \{\langle [z_1]_w, \ldots, [z_n]_w \rangle : Pz_1 \ldots z_n \in w\}$, where all the $[x_i]_w$ are non-empty (for $V_w(x_i)$ is defined). This means that there are variables $z_1, \ldots, z_n$ such that $\{x_1 = z_1, \ldots, x_n = z_n, Pz_1 \ldots z_n\} \subseteq w$. Then $Px_1 \ldots x_n \in w$ by (LL*).

In the other direction, if $Px_1 \dots x_n \in w$, then $x_i = x_i \in w$ for all x_i in $x_1 \dots x_n$ (see p. 108). Hence $\langle [x_1]_w, \dots, [x_n]_w \rangle \in \{\langle [z_1]_w, \dots, [z_n]_w \rangle : Pz_1 \dots z_n \in w\}$, i.e. $\langle V_w(x_1), \dots, V_w(x_n) \rangle \in V_w(P)$.

2. A is $\neg B$. $w, V \Vdash \neg B$ iff $w, V \nVdash B$ by definition 2.7, iff $B \notin w$ by induction hypothesis, iff $\neg B \in w$ by maximality of w.

3. A is $B \supset C$. $w, V \Vdash B \supset C$ iff $w, V \nVdash B$ or $w, V \Vdash C$ by definition 2.7, iff $B \notin w$ or $C \in w$ by induction hypothesis, iff $B \supset C \in w$ by maximality and consistency of w and the fact that $\vdash_L \neg B \supset (B \supset C)$ and $\vdash_L C \supset (B \supset C)$.

4. A is $\langle y : x \rangle B$. Assume first that $w, V \Vdash y \neq y$. So $V_w(y)$ is undefined, and it is not the case that $V_w(y)$ has multiple counterparts at any world. And then $w, V \Vdash \langle y : x \rangle B$ iff $w, V^{[y/x]} \Vdash B$ by definition 3.2, iff $w, V \Vdash [y/x]B$ by lemma 3.9, iff $[y/x]B \in w$ by induction hypothesis. Also by induction hypothesis, $y \neq y \in w$. By (SCN), $\vdash_L y \neq y \supset ([y/x]B \leftrightarrow \langle y : x \rangle B)$. So $[y/x]B \in w$ iff $\langle y : x \rangle B \in w$.

   Next, assume that $w, V \Vdash y = y$; so by induction hypothesis $y = y \in w$. Assume further that $\langle y : x \rangle B \notin w$. Then $\neg \langle y : x \rangle B \in w$ by maximality of w, and $\langle y : x \rangle \neg B \in w$ by (S$\neg$). Since w is substitutionally witnessed and $y = y \in w$, there is a variable $z \notin \mathit{Var}(\langle y : x \rangle \neg B)$ such that $y = z \in w$ and $[z/x]\neg B \in w$. By induction hypothesis, $w, V \Vdash y = z$. Moreover, by definition 3.3, $\neg [z/x]B \in w$, and so $[z/x]B \notin w$ by consistency of w. By induction hypothesis, $w, V \nVdash [z/x]B$. By definition 2.7, then $w, V \Vdash \neg [z/x]B$, i.e. $w, V \Vdash [z/x]\neg B$. Since z and x are modally separated in B, then $w, V^{[z/x]} \Vdash \neg B$ by lemma 3.9. But $V^{[z/x]}$ and $V^{[y/x]}$ agree on all variables at w, because $w, V \Vdash y = z$. So $w, V^{[y/x]} \Vdash \neg B$ by the locality lemma 2.11. So $w, V^{[y/x]} \nVdash B$ by definition 2.7, and $w, V \nVdash \langle y : x \rangle B$ by definition 3.2.

   In the other direction, assume $\langle y : x \rangle B \in w$. Since w is substitutionally witnessed and $y = y \in w$, there is a new variable z such that $y = z \in w$ and $[z/x]B \in w$. By induction hypothesis, $w, V \Vdash y = z$ and $w, V \Vdash [z/x]B$. Since z and x are modally separated in B, $w, V^{[z/x]} \Vdash B$ by lemma 3.9. As before $V^{[z/x]}$ and $V^{[y/x]}$ agree on all variables at w, because $w, V \Vdash y = z$; so $w, V^{[y/x]} \Vdash B$ by lemma 2.11 and $w, V \Vdash \langle y : x \rangle B$ by definition 3.2.

5. A is $\forall x B$. We first show that for any variable x, $w, V \Vdash Ex$ iff $Ex \in w$: $w, V \Vdash Ex$ iff $V_w(x) \in D_w$ by definition 3.3, iff $[x]_w \in D_w$ by definition 7.5, iff $Ex \in w$ by definition 7.5.

   Now assume $\forall x B \in w$, and let y be any variable such that $Ey \in w$. As just shown, $w, V \Vdash Ey$. By (FUI**), $\exists x (x = y \wedge B) \in w$. By witnessing, there is a $z \notin \mathit{Var}(B)$ such that $z = y \wedge [z/x]B \in w$, and thus $z = y \in w$ and $[z/x]B \in w$. By induction hypothesis, $w, V \Vdash z = y$ and $w, V \Vdash [z/x]B$. By lemma 3.9, then $w, V^{[z/x]} \Vdash B$. And since $V_w(z) = V_w(y)$, it follows by lemma 2.11 that $w, V^{[y/x]} \Vdash B$. So if $\forall x B \in w$, then $w, V^{[y/x]} \Vdash B$ for all variables y with $Ey \in w$, i.e. with $V_w(y) \in D_w$. Since every member $[y]_w$ of D_w is denoted by some variable y under V_w, this means that $w, V' \Vdash B$ for all existential x-variants V' of V on w. So $w, V \Vdash \forall x B$. (The last steps are only valid if V is the original interpretation function, not if it is an image: with images, members of D_w can be unnamed.)

Conversely, assume $\forall xB \notin w$. Then $\exists x \neg B \in w$; so by witnessing, $[y/x]\neg B \in w$ for some $y \notin Var(B)$ with $Ey \in w$. Then $\neg[y/x]B \in w$ and so $[y/x]B \notin w$. As shown above, $w, V \Vdash Ey$. Moreover, by induction hypothesis, $w, V \nVdash [y/x]B$. By lemma 3.9, then $w, V^{[y/x]} \nVdash B$. Let V' be the (existential) x-variant of V on w with $V'_w(x) = V_w^{[y/x]}(x)$. By the locality lemma, $w, V' \nVdash B$. So $w, V \nVdash \forall xB$.

(FUI**) here gives us a kind of witnessing for universal formulas, which amounts to something like (FUI): whenever $\forall xB \in w$, then for each $[y]_w \in D_w$ there is a z s.t. $z\!=\!y \in w$ and $[z/x]B \in w$. Compare (FUI$_s$), which ensures that whenever $\forall xB \in w$ then for each $[y]_w \in D_w$, $\langle y : x\rangle B \in w$ and so by substitutional witnessing, for each $[y]_w \in D_w$ there is a z s.t. $z\!=\!y \in w$ and $[z/x]B \in w$.

6. A is $\Box B$. Assume $w, V \Vdash \Box B$. Then $w', V' \Vdash B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$. We first show that if $w \xrightarrow{\tau} w'$, then $V_w \triangleright V^\tau_{w'}$. Recall that by definition 2.6, $V_w \triangleright V^\tau_{w'}$ if there is a counterpart relation between w and w' such that for every variable x, if $V_w(x)$ has a counterpart at w', then $V^\tau_{w'}(x)$ is one of these counterparts, otherwise it is undefined. Moreover, by definition 7.5, every τ with $w \xrightarrow{\tau} w'$ determines a counterpart relation that holds between $V_w(x)$ at w and $d' \in U_{w'}$ at w' iff there is a variable z with $z \in V_w(x)$ and $z^\tau \in d'$. Taken together, $V_w \triangleright V^\tau_{w'}$ iff there is a transformation σ such that $w \xrightarrow{\sigma} w'$ and for every variable x, if there is a $z \in V_w(x)$ such that $[z^\sigma]_{w'} \in U_{w'}$, then there is a $z \in V_w(x)$ such that $z^\sigma_{w'} \in V^\tau_{w'}(x)$. By definitions 2.6 and 7.5, $V_w \triangleright V^\tau_{w'}$ means that there is a transformation σ such that $w \xrightarrow{\sigma} w'$ and for every variable y, if there is a $z \in V_w(y)$ such that $[z^\sigma]_{w'} \in U_{w'}$ (i.e., if $V_w(y)$ has any σ-counterpart at w'), then there is a $z \in V_w(y)$ with $z^\sigma \in V^\tau_{w'}(y)$ (i.e., then $V^\tau_{w'}(y)$ is such a counterpart), otherwise $V^\tau_{w'}(y)$ is undefined. The relevant transformation σ will be τ. So what we'll show is this: for every variable y, if there is a $z \in V_w(y)$ such that $[z^\tau]_{w'} \in U_{w'}$, then there is a $z \in V_w(y)$ with $z^\tau \in V^\tau_{w'}(y)$, otherwise $V^\tau_{w'}(y)$ is undefined.

   Let y be any variable. Assume first that there is a $z \in V_w(y)$ such that $[z^\tau]_{w'} \in U_{w'}$. Then $z\!=\!y \in w$ and $z^\tau\!=\!z^\tau \in w'$. By either (Neg) and (EI) or (=R), $\vdash_L z\!=\!y \supset y\!=\!y$; so $y = y \in w$. Moreover, by either (TE), (EI), (Nec) and (K) or (=R) and (Nec), $\vdash_L z = y \supset \Box(z = z \supset y = y)$; so $\Box(z = z \supset y = y) \in w$. By definition of $w \xrightarrow{\tau} w'$, then $z^\tau = z^\tau \supset y^\tau = y^\tau \in w'$. So $y^\tau = y^\tau \in w'$. Hence $y \in V_w(y)$ and $y^\tau \in [y^\tau]_{w'} = V_{w'}(y^\tau) = V^\tau_{w'}(y)$. Alternatively, assume there is no $z \in V_w(y)$ with $z^\tau = z^\tau \in w'$. Then either $V_w(y) = \emptyset$, in which case $y \neq y \in w$, and so $\Box(y \neq y) \in w$ by (NA), (EI), (Nec) and (K), and $y^\tau \neq y^\tau \in w'$ by definition of $w \xrightarrow{\tau} w'$, or else $V_w(y) \neq \emptyset$, but $z^\tau \neq z^\tau \in w'$ for all $z \in V_w(y)$, in which case, too, $y^\tau \neq y^\tau \in w'$ since $y \in V_w(y)$. Either way, $V_{w'}(y^\tau) = V^\tau_{w'}(y)$ is undefined.

   We've shown that if $w, V \Vdash \Box B$, then for every w' and τ with $w \xrightarrow{\tau} w'$, $w', V^\tau \Vdash B$. By the transformation lemma, then $w', V \Vdash B^\tau$. By induction hypothesis, $B^\tau \in w'$. Now suppose $\Box B \notin w$. Then $\Diamond\neg B \in w$ by maximality of w. By the existence lemma, there is then a world w' and transformation τ with $w \xrightarrow{\tau} w'$ and $\neg B^\tau \in w'$. (Any transformation whose range excludes infinitely many variables will do.) But we've just seen that if $w \xrightarrow{\tau} w'$, then $B^\tau \in w'$. So if $w, V \Vdash \Box B$, then $\Box B \in w$.

   For the other direction, assume $w, V \nVdash \Box B$. So $w', V' \nVdash B$ for some w', V' with wRw' and $V_w \triangleright V'_{w'}$.

It may help to run through a simplified version of the proof first, assuming that B contains a single free variable x and ignoring negative logics. So we assume that $w, V \Vdash \Diamond\neg B(x)$, and hence $w', V' \Vdash \neg B(x)$ for suitable w', V'. If V' were V^τ for $\tau : w \xrightarrow{\tau} w'$, things would be easy: suppose for reductio that $\Diamond\neg B(x) \notin w$ and so $\Box B(x) \in w$; then $B(x^\tau) \in w'$ for all $w' : w \xrightarrow{\tau} w'$, and by i.h., $w', V^\tau \Vdash B(x)$ – contradiction. But V' may not be V^τ, i.e. $V'_{w'}(x)$ may not be $V^\tau_{w'}(x)$. What we do know is that there is some "guise" of x at w, i.e. some variable y with $x = y \in w$ for which $V'_{w'}(x) = [y^\tau]_{w'}$. Moreover, $[y^\tau]_{w'} = V_{w'}(y^\tau) = V^\tau_{w'}(y) = V^{\tau\cdot[y/x]}_{w'}(x)$. So we can think of V' as $V^{\tau\cdot[y/x]}$.

Thus our assumption is that $w', V^{\tau\cdot[y/x]} \Vdash \neg B(x)$. We suppose for reductio that $\Box B(x) \in w$. Since $x = y \in w$, it follows that $\langle y : x\rangle\Box B(x) \in w$ and so $\Box\langle y : x\rangle B(x) \in w$. By construction of R, then $\langle y^\tau : x^\tau\rangle B(x^\tau) \in w'$. By witnessing, there is a new variable z such that $z = y^\tau \in w'$ and $[z/x^\tau]B(x^\tau) \in w'$. By induction hypothesis, $w', V \Vdash z = y^\tau$ and $w', V \Vdash [z/x^\tau]B(x^\tau)$. Since z is new, $w', V^{[z/x^\tau]} \Vdash B(x^\tau)$ by lemma 3.9. By the transformation lemma, $w', V^{[z/x^\tau]\cdot\tau} \Vdash B(x)$. But $V^{[z/x^\tau]\cdot\tau}_{w'}(x) = V^{[z/x^\tau]}_{w'}(x^\tau) = V_{w'}(z) = V_{w'}(y^\tau)$ (because $w', V \Vdash y^\tau = z$) $= V^\tau_{w'}(y) = V^{\tau\cdot[y/x]}_{w'}(x)$. Hence by the locality lemma, $w', V^{\tau\cdot[y/x]} \Vdash B(x)$ – contradiction.

Without substitution, we don't have $\langle y : x\rangle B(x) \in w$ and hence $\langle y^\tau : x^\tau\rangle B(x^\tau) \in w'$. Instead we get the witnessing consequence – that there is a new variable z such that $z = y^\tau \in w'$ and $[z/x^\tau]B(x^\tau) \in w'$ – by definition of accessibility.

As before, $V_w \triangleright V'_{w'}$ means that there is a transformation τ with $w \xrightarrow{\tau} w'$ such that for every variable x, either there is a $y \in V_w(x)$ with $y^\tau \in V'_{w'}(x)$, or there is no $y \in V_w(x)$ with $y^\tau = y^\tau \in w'$, in which case $V'_{w'}(x)$ is undefined. Let τ be any transformation with $w \xrightarrow{\tau} w'$, and let $*$ be a substitution that maps each variable x in B to some member y of $V_w(x)$ with $y^\tau \in V'_{w'}(x)$, or to itself if there is no such y. Thus if $x \in \mathit{Var}(B)$ and $V'_{w'}(x)$ is defined, then $(*x)^\tau \in V'_{w'}(x)$, and so $V'_{w'}(x) = [(*x)^\tau]_{w'} = V^{\tau;*}_{w'}(x)$. Alternatively, if $V'_{w'}(x)$ is undefined (so $*x = x$), then $V^{\tau;*}_{w'}(x) = V^\tau_{w'}(x)$ is also undefined. The reason is that otherwise $V^\tau_{w'}(x) = [x^\tau]_{w'} \neq \emptyset$ and $x^\tau = x^\tau \in w'$; by definition of accessibility, then $\Box x \neq x \notin w$ and hence $x = x \in w$, as $\vdash_L x \neq x \supset \Box x \neq x$; so there is a $y \in V_w(x)$, namely x, such that $y^\tau = y^\tau \in w'$, in which case $V'_{w'}(x)$ cannot be undefined (by definition 2.6). So V' and $V^{\tau\cdot *}$ agree at w' on all variables in B. By lemma 2.10, $w', V^{\tau\cdot *} \nVdash B$.

Now suppose for reductio that $\Box B \in w$. Let $x_1, \ldots, x_n$ be the variables x in $\mathit{Var}(B)$ with $(*x)^\tau \in V'_{w'}(x)$ (thus excluding empty variables as well as variables denoting individuals without τ-counterparts at w'). For each such x_i, $*x_i \in V_w(x_i)$, and so $x_i = *x_i \in w$. If L is with substitution, then by (LL$_\mathrm{n}$), $\langle *x_1, \ldots, *x_n : x_1, \ldots, x_n\rangle\Box B \in w$; so $\Box\langle *x_1, \ldots, *x_n : x_1, \ldots, x_n\rangle B \in w$ by (S$\Box$). By definition of $w \xrightarrow{\tau} w'$, then $\langle (*x_1)^\tau, \ldots, (*x_n)^\tau : x_1^\tau, \ldots, x_n^\tau\rangle B^\tau \in w'$. By substitutional witnessing, it follows that there are new variables $z_1, \ldots, z_n$ such that $z_i = (*x_i)^\tau \in w'$ and (hence) $[z_1, \ldots, z_n/x_1^\tau, \ldots, x_n^\tau]B^\tau \in w'$. If L is without substitution, this fact – that there are new variables $z_1, \ldots, z_n$ such that $z_i = (*x_i)^\tau \in w'$ and $[z_1, \ldots, z_n/x_1^\tau, \ldots, x_n^\tau]B^\tau \in w'$ – is guaranteed directly by definition of $w \xrightarrow{\tau} w'$ and the fact that $\Box B \in w$.

By induction hypothesis, $w', V \Vdash z_i = (*x_i)^\tau$ and $w', V \Vdash [z_1, \ldots, z_n/x_1^\tau, \ldots, x_n^\tau]B^\tau$. Since the z_i are new, $w', V^{[z_1,\ldots,z_n/x_1^\tau,\ldots,x_n^\tau]} \Vdash B^\tau$ by lemma 3.9. By the transformation lemma 3.13, then $w', V^{[z_1,\ldots,z_n/x_1^\tau,\ldots,x_n^\tau]\cdot\tau} \Vdash B$. However, for each x_i, $V_{w'}^{[z_1,\ldots,z_n/x_1^\tau,\ldots,x_n^\tau]\cdot\tau}(x_i) = V_{w'}^{[z_1,\ldots,z_n/x_1^\tau,\ldots,x_n^\tau]}(x_i^\tau) = V_{w'}(z_i) = V_{w'}((*x_i)^\tau)$ (because $w', V \Vdash (*x_i)^\tau = z_i$) $= V_{w'}^\tau(*x_i) = V_{w'}^{\tau\cdot[*x_1,\ldots,*x_n/x_1,\ldots,x_n]}(x_i) = V_{w'}^{\tau\cdot*}(x_i)$. Similarly, if $x \in \mathit{Var}(B)$ is none of the $x_1, \ldots, x_n$, so $(*x)^\tau \notin V'_{w'}(x)$, then $*x$ is x by definition of $*$, and so $V_{w'}^{[z_1,\ldots,z_n/x_1^\tau,\ldots,x_n^\tau]\cdot\tau}(x) = V_{w'}^\tau(x) = V_{w'}^\tau(*x) = V_{w'}^{\tau\cdot*}(x)$. So $V^{[z_1,\ldots,z_n/x_1^\tau,\ldots,x_n^\tau]\cdot\tau}$ and $V_{w'}^{\tau\cdot*}$ agree at w' on all variables in B. By lemma 2.11, then $w', V^{\tau\cdot*} \Vdash B$ – contradiction. ∎

Where do I need the strong form of (LL) that allows $x = y \supset \Box Fx \supset \Box Fy$? Suppose we remove it from the logic. Then $\{x{=}y, \Box Fx, \neg\Box Fy = \Diamond\neg Fy\}$ is consistent, and can be extended to a Henkin set w. The existence lemma requires that $\{\neg Fy^\tau, Fx^\tau\}$ can be extended to a Henkin set w'. That's fine. But it also requires that wRw', which by clause (x) of R means that there is a new variable z s.t. $\{z{=}y^\tau, Fz\} \subseteq w'$. But then w' is inconsistent. So the strong form of (LL) is needed in the existence lemma to ensure wRw'.

8 Completeness

Recall that a logic L in some language of quantified modal logic is *(strongly) complete with respect to a class of models* $\mathbb{M}$ if every L-consistent set of formulas Γ is verified at some world in some model in $\mathbb{M}$. L is *characterised by* $\mathbb{M}$ if L is sound and complete with respect to $\mathbb{M}$.

The minimal positive and negative logics from sections 4 and 5 were designed to be complete with respect to the class of all positive and negative models, respectively. Let's confirm that this is the case.

Theorem 8.1 (Completeness of P and P_s)
The logics P and P_s are (strongly) complete with respect to the class of positive counterpart models.

Proof Let L range over P and P_s. We have to show that whenever a set of $\mathcal{L}$-formulas Γ is L-consistent, then there is some world in some positive counterpart model that verifies all members of Γ. By lemma 7.6, the canonical model $\mathcal{M}_L = \langle \mathcal{S}_L, V_L \rangle$ for L is a positive model. By the Extensibility Lemma, $\Gamma \subseteq w$ for some world w in $\mathcal{M}_L$, since none of the infinitely many variables Var^+ occur in Γ. By the truth lemma, then $w, V_L \Vdash_{\mathcal{S}_L} A$ for each $A \in \Gamma$. ∎

Theorem 8.2 (Completeness of N and N_s)
The logics N and N_s are (strongly) complete with respect to the class of negative counterpart models.

PROOF Let L range over N and N_s, and let Γ be an L-consistent set of $\mathcal{L}$-formulas. By lemma 7.6, the canonical model $\mathcal{M}_L = \langle \mathcal{S}_L, V_L \rangle$ for L is a negative model. By the Extensibility Lemma, $\Gamma \subseteq w$ for some world w in $\mathcal{M}_L$, since none of the infinitely many variables Var^+ occur in Γ. By the truth lemma, then $w, V_L \Vdash_{\mathcal{S}_L} A$ for each $A \in \Gamma$. ■

Together with the soundness theorems 4.3, 4.6, 5.4 and 5.5, it follows that P and P_s are characterized by the class of positive models, and N and N_s by the class of negative models.

In footnote 1 (on page 7) I mentioned that the introduction of multiple counterpart relations makes little difference to the base logic. Let's call a counterpart structure in which any two worlds are linked by at most one counterpart relation *unirelational.* As it turns out, P and P_s are also characterized by the class of unirelational positive models, and N and N_s by the class of unirelational negative models. The easiest way to see this is perhaps to note that all the lemmas in the previous section still go through if we define accessibility and counterparthood in canonical models by a fixed transformation τ whose range excludes infinitely many variables. The extensibility lemma 7.7 and existence lemma 7.8 are unaffected by this change; the only part that needs adjusting is the clause for $\Box B$ in the proof of the truth lemma 7.9, but the adjustments are straightforward.

Assume $w, V \Vdash \Box B$. Then $w', V' \Vdash B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$. We first show that $V_w \triangleright V^\tau_{w'}$. By definitions 2.6 and 7.5, this means that for every variable y, either there is a $z \in V_w(y)$ with $z^\tau \in V^\tau_{w'}(y)$, or there is no $z \in V_w(y)$ with $z^\tau = z^\tau \in w'$, in which case $V^\tau_{w'}(y)$ is undefined. So let y be any variable. Assume there is a $z \in V_w(y)$ with $z^\tau = z^\tau \in w'$. Then all $z \in V_w(y)$ are such that $z^\tau = z^\tau \in w'$, since $\vdash_L x = y \supset \Box(x = x \supset y = y)$, by either (TE) and (Neg) or (=R). Moreover, then $y \in V_w(y)$. So $y^\tau \in [y^\tau]_{w'} = V_{w'}(y^\tau) = V^\tau_{w'}(y)$. Alternatively, assume there is no $z \in V_w(y)$ with $z^\tau = z^\tau \in w'$. Then either $V_w(y) = \emptyset$, in which case $y \neq y \in w$, and so $\Box(y \neq y) \in w$ by (NA) and $y^\tau \neq y^\tau \in w'$ by definition of R, or else $V_w(y) \neq \emptyset$, but $z^\tau \neq z^\tau \in w'$ for all $z \in V_w(y)$, in which case, too, $y^\tau \neq y^\tau \in w'$ since $y \in V_w(y)$. Either way, $V_{w'}(y^\tau) = V^\tau_{w'}(y)$ is undefined. So $V_w \triangleright V^\tau_{w'}$.

We've shown that if $w, V \Vdash \Box B$, then for every w' with wRw', $w', V^\tau \Vdash B$. By the transformation lemma, then $w', V \Vdash B^\tau$, and by induction hypothesis, $B^\tau \in w'$. Now suppose $\Box B \notin w$. Then $\Diamond \neg B \in w$ by maximality of w. By the existence lemma, there is then a world w' with wRw' and $\neg B^\tau \in w'$. But we've just seen that if wRw', then $B^\tau \in w'$. So if $w, V \Vdash \Box B$, then $\Box B \in w$.

For the other direction, assume $w, V \nVdash \Box B$. So $w', V' \nVdash B$ for some w', V' with wRw' and $V_w \triangleright V'_{w'}$. As before, $V_w \triangleright V'_{w'}$ means that for every variable y, either there is a $z \in V_w(y)$ with $z^\tau \in V'_{w'}(y)$, or there is no $z \in V_w(y)$ with $z^\tau = z^\tau \in w'$, in which case $V'_{w'}(y)$ is undefined. Let $*$ be a substitution that maps each variable y in B to some member z of $V_w(y)$ with $z^\tau \in V'_{w'}(y)$, or to itself if there is no such z. Thus if $y \in Var(B)$ and $V'_{w'}(y)$ is defined, then $(*y)^\tau \in V'_{w'}(y)$, and so $V'_{w'}(y) = [(*y)^\tau]_{w'} = V^{\tau \cdot *}_{w'}(y)$. If $V'_{w'}(y)$ is undefined, then $V^{\tau \cdot *}_{w'}(y) = V^\tau_{w'}(y)$ is undefined, as otherwise $V^\tau_{w'}(y) = [y^\tau]_{w'} \neq \emptyset$ and $y^\tau = y^\tau \in w'$. So V' and $V^{\tau \cdot *}$ agree at w' on all variables in B. By the coincidence lemma, $w', V^{\tau \cdot *} \nVdash B$.

Now we have to distinguish logics with and without substitution.

First, with substitution. Suppose for reductio that $\Box B \in w$. Let $y_1, \ldots, y_n$ be the variables y in B with $(*y)^\tau \in V'_{w'}(y)$. (Note that this excludes all empty variables, as well as variables denoting things without counterparts.) For each such y, $*y \in V_w(y)$, and so $y = *y \in w$. By ($\mathrm{LL_n}$), $\langle *y_1, \ldots, *y_n : y_1, \ldots, y_n \rangle \Box B \in w$. By (S$\Box$), $\Box \langle *y_1, \ldots, *y_n : y_1, \ldots, y_n \rangle B \in w$. By construction of R, then $(\langle *y_1, \ldots, *y_n : y_1, \ldots, y_n \rangle B)^\tau \in w'$. By induction hypothesis, it follows that $w', V \Vdash (\langle *y_1, \ldots, *y_n : y_1, \ldots, y_n \rangle B)^\tau$. By the transformation lemma 3.13, then $w', V^\tau \Vdash \langle *y_1, \ldots, *y_n : y_1, \ldots, y_n \rangle B$. So $w', V^{\tau \cdot [*y_1, \ldots, *y_n / y_1, \ldots, y_n]} \Vdash B$ by lemma 3.15. But $[*y_1, \ldots, *y_n : y_1, \ldots, y_n]$ is $*$. (If $(*y)^\tau \notin V'_{w'}(y)$, then $V'_{w'}(y)$ is undefined, and there is no $z \in V_w(y)$ with $z^\tau \in V'_{w'}(y)$, so then $*y = y$.) So $w', V^{\tau \cdot *} \Vdash B$. Contradiction.

Next, without substitution. (Remember: wRw' iff for every formula B and variables $x_1 \ldots x_n$, $y_1, \ldots, y_n$ such that the $x_1 \ldots x_n$ are zero or more pairwise distinct members of *Varf*(B) and each x_i is distinct from y_i (xxx check!), if $x_1 = y_1 \wedge \ldots \wedge x_n = y_n \wedge \Box B \in w$ and $y_1^\tau = y_1^\tau \wedge \ldots \wedge y_n^\tau = y_n^\tau \in w'$, then there are variables $z_1 \ldots z_n \notin \mathit{Var}(B^\tau)$ such that $z_1 = y_1^\tau \wedge \ldots \wedge z_n = y_n^\tau \wedge [z_1 \ldots z_n / x_1^\tau \ldots x_n^\tau] B^\tau \in w'$.) Let $x_1, \ldots, x_n$ be the free variables x in B such that $*x$ is not x and $(*x)^\tau \in V'_{w'}(x)$ xxx that's redundant?!, and suppose for reductio that $\Box B \in w$. By definition of $*$, for any x_i in $x_1, \ldots, x_n$, $*x_i$ is a member of $V_w(x_i)$, so $x_i = *x_i \in w$. By definition of R (definition 7.5), there are variables $z_1, \ldots, z_n \notin \mathit{Var}(B^\tau)$ such that $z_1 = (*x_1)^\tau \wedge \ldots \wedge z_n = (*x_n)^\tau \wedge [z_1 \ldots z_n / x_1^\tau \ldots x_n^\tau] B^\tau \in w'$. So for all i, $z_i = (*x_i)^\tau \in w'$ and $[z_1 \ldots z_n / x_1^\tau \ldots x_n^\tau] B^\tau \in w'$. By induction hypothesis, $w', V \Vdash z_i = (x_i^*)^\tau$ and $w', V \Vdash [z_1 \ldots z_n / x_1^\tau \ldots x_n^\tau] B^\tau$. Since $z_i \notin \mathit{Var}(B^\tau)$, $[z_1 \ldots z_n / x_1^\tau \ldots x_n^\tau]$ is $[z_1 / x_1^\tau] \cdot \ldots \cdot [z_n / x_n^\tau]$ XXXX?, and so $w', V^{[z_1 \ldots z_n / x_1^\tau \ldots x_n^\tau]} \Vdash B^\tau$ by part (i) of lemma 3.9. And as $V_{w'}(z_i) = V_{w'}((*x_i)^\tau)$, this means that $w', V^{[(*x_1)^\tau \ldots (*x_n)^\tau / x_1^\tau \ldots x_n^\tau]} \Vdash B^\tau$. So $w', V^{[(*x_1)^\tau \ldots (*x_n)^\tau / x_1^\tau \ldots x_n^\tau] \cdot \tau} \Vdash B$ by the transformation lemma 3.13. But $V^{[(*x_1)^\tau \ldots (*x_n)^\tau / x_1^\tau \ldots x_n^\tau] \cdot \tau}$ coincides at w' with $V^{\tau \cdot *}$: for any variable z, $V_{w'}^{[(*x_1)^\tau \ldots (*x_n)^\tau / x_1^\tau \ldots x_n^\tau] \cdot \tau}(z) = V_{w'}^{[(*x_1)^\tau \ldots (*x_n)^\tau / x_1^\tau \ldots x_n^\tau]}(z^\tau) = V_{w'}([(*x_1)^\tau \ldots (*x_n)^\tau / x_1^\tau \ldots x_n^\tau] z^\tau) = V_{w'}^\tau([*x_1 \ldots *x_n / x_1 \ldots x_n] z) = V_{w'}^{\tau \cdot *}(z)$. So $w', V^{\tau \cdot *} \Vdash B$. Contradiction.

On the other hand, fixing the counterpart relation in this way gets inconvenient when we want to characterise stronger logics. Recall from section 6 that the schema $\mathsf{T} = \Box A \supset A$ is valid in a structure $\mathcal{S}$ iff (i) every world in $\mathcal{S}$ can see itself and (ii) every finite sequence of individuals at every world is its own counterpart at that world. We can check that the canonical model of $\mathsf{P} + \mathsf{T}$ satisfies (i) and (ii), and therefore that $\mathsf{P} + \mathsf{T}$ is characterized by the class of positive counterpart models satisfying (i) and (ii). As for (i), suppose there is some world w in the model that can't see itself. By definition of canonical models in section 7, this means that there is no transformation τ such that $A^\tau \in w$ whenever $\Box A \in w$. But if w contains all instances of $\Box A \supset A$, then there is such a transformation, namely the identity transformation. Similarly for (ii). Suppose some sequence $\langle [x_1]_w, \ldots, [x_n]_w \rangle \in U_w$ is not its own counterpart at w, i.e. there is no $C \in K_{w,w'}$ such that $[x_i]_w C [x_i]_w$ for all $1 \leq i \leq n$. In the canonical model, this means that there is no transformation τ such that (a) $A^\tau \in w$ whenever $\Box A \in w$, and (b) for all $1 \leq i \leq n$, there is a $y \in [x_i]_w$ with $y^\tau \in [x_i]_w$. (See definition 7.5.) But if w contains all instances of $\Box A \supset A$, then again the identity transformation satisfies (a) and (b).

Every quantified modal logic is strongly complete with respect to every class of models that contains its canonical model. However, on the traditional idea that logical truths should be true on any interpretation of the non-logical terms, an arguably more important kind of completeness is completeness with respect to all models with a certain type of *structure*. Strictly speaking, we have two such notions, one for positive and one for negative logics. As in section **??**, let me focus on positive logics for the moment.

Definition 8.3 (Positive completeness and characterisation)
A logic L in some language of quantified modal logic is *(strongly) positively complete with respect to a class of structures* $\mathbb{S}$ if every L-consistent set of formulas Γ is verified at some world in some positive model $\langle \mathcal{S}, V \rangle$ with $\mathcal{S} \in \mathbb{S}$. L is *positively characterised by* $\mathbb{S}$ if it is sound and positively complete with respect to $\mathbb{S}$.

Now we might try to show, first, that if a PML is categorical, then so is its quantified counterpart (whatever that means). Then we could show that the canonical model of the PML is in a class of frames F iff the opaque propositional guise of the canonical model of the quantified counterpart is in F. Then we'd have completeness transfer for all categorical logics.

Theorem 8.4 ((Positive) completeness transfer)
If L is a (unimodal) propositional modal logic that is complete with respect to the Kripke frames in some class F, then the PL is positively complete with respect to the total counterpart structures whose opaque propositional guise is in F.

One way to

[To be continued...]

(Here we also see that we need multiple counterpart relations: if we only had the identity transformation from w to w, we would validate more principles than P+T?)

Definition 8.5 (Canonicity)
A logic L is *canonical* if all members of L are valid in the structure of the canonical model of L.

Every canonical logic is structure-complete.

If we have canonicity transfer, we get completeness transfer

Model-completeness is a rather weak property, since models contain interpretations. A *logical* truth should be true on any interpretation of the non-logical terms, not just on this or that particular interpretation. So genuine logics ought to be structure-complete, not only model-complete. In fact, they ought to be more. For arguably the choice of W or D is also part of the interpretation: a genuine logical truth should be true independently of how many things there are. So ideally, we'd want completeness with respect to a class of structures that is not constrained by particular choices of W or D. On the other hand, since a modal logic can be seen as the logic of a particular (accessibility) relation, expressed by the box, which is a logical term, logical truths need not be true independently of the features of this relation. Thus it is nice that many well-known logics are characterised by classes of structures that are constrained precisely by the accessibility relation. (The logic T is characterised by the class of structures where accessibility is reflexive, and so on.)

Quantified modal logic, on the present picture, is the logic of *two* relations: one linking worlds and one linking individuals. So we would like to characterise interesting logics by classes of structures constrained only by the accessibility and the counterpart relation. To simplify the following discussion, I will restrict myself to positive models and assume that no individual at any world can have a counterpart at an inaccessible world. As mentioned on p. **??**, this assumption makes no difference to the evaluation of any formula because the counterpart relation is only invoked when we move from one world to another by the accessibility relation. Thus modal formulas can only constrain the "visible part" of the counterpart relation, restricted to accessible worlds, and we can simplify things by assuming there is no invisible part. The restriction to positive models further simplifies things because the accessibility relation can then be defined by the counterpart relation: wRw' iff there are d, d' with $\langle d, w\rangle C\langle d', w'\rangle$ (as also mentioned on p. **??**). So our logical systems characterize a single relation after all. In addition, the relevant properties of the counterpart relation are easier to describe if we ignore negative models. For example, we'd like $A \supset \Box\Diamond A$ to be valid in a structure iff both R and C are symmetrical. But in non-total structures (for negative models), we have to be careful because it is not enough that whenever $\langle d, w\rangle C\langle d', w'\rangle$ then $\langle d', w'\rangle C\langle d, w\rangle$. Assume $\mathcal{S}$ is symmetrical and suppose for reductio that some instance of $A \supset \Box\Diamond A$ is not valid on $\mathcal{S}$, so that $w, V \Vdash A$ and $w, V \nVdash \Box\Diamond A$, i.e. $w, V \Vdash \Diamond\Box\neg A$. The latter means that there are w', V' with wRw' and $V_w \triangleright V_{w'}$ such that for all w'', V'' with $w'Rw''$ and $V_{w'} \triangleright V_{w''}$, $w'', V'' \Vdash \neg A$. But since wRw' and R is symmetrical, one such w'' is w. And since C is symmetrical, we'd like to say that V is a w-image of V' at w, so that we get $w, V \Vdash \neg A$, completing the reductio. But if some d at w has no counterpart at all at w', then $V'_{w'}(x)$ may be empty and thus not denote any counterpart of $V_w(x)$. So when we shift back from w' to w and from V' to V, the empty variable x turns non-empty. That is not OK by definition 2.6, which says that in order for $V'_{w'} \triangleright V_w$, it must be the case that whenever $\langle V'_{w'}(x), w'\rangle$ has a counterpart at w, then $V_w(x)$ is one of these counterparts, *otherwise* $V_w(x)$ *is undefined.* So we'd somehow have to strengthen the definition of symmetry, or find a different line of proof here.

In PML, modal principles express constraints on the accessibility relation. Ideally, this should carry over to de dicto modal principles in QML. De re principles would then express constraints on accessibility and counterparthood. But it isn't true that e.g. (T) corresponds to reflexive R, without any constraint on counterparthood. Suppose F is not a frame for (T), i.e. it is not the case that all instances of $\Box A \supset A$ are true at all w, V, U, D, C based on F. So some instance of $\Box A \supset A$ is false for some w, V, U, D, C on F. So $w, V \Vdash \Box A$ and $w, V \neg \Vdash A$. The former means that $w', V' \Vdash A$ for all wRw' and $V_w \triangleright V'_{w'}$. If wRw, then in particular $w, V' \Vdash A$ for all $V_w \triangleright V_w$. But now it doesn't follow that $w, V \Vdash A$, for V need not be an image of itself at w. So not all reflexive frames are frames for (T). In fact, hardly any non-trivial reflexive frame is a frame for (T), for we can always cook up an interpretation and a counterpart relation that renders (T) false at some world that can see itself. We just need to make Fx false at w, Fy true, and stipulate that $V_w(y)$ is the counterpart at w of $V_w(x)$, so that $\Box Fx$ is true. This also shows that $\mathsf{P} + T$ is not sound and complete wrt reflexive frames, as it isn't even sound.

If features of R by themselves don't determine interesting systems, we might consider merging R and C. What's the combined feature of R and C that corresponds to $\Box A \supset A$? $\Box A \supset A$ is valid iff $w, V \Vdash A$ whenever $w', V' \Vdash A$ for all wRw' and $V_w \triangleright V'_{w'}$. This does mean, I think, that (a) wRw for all worlds and (b) everything is at least a counterpart of itself.

Maybe we can say that whenever a PML L is characterised by a class of frames F, then PL is characterised by a corresponding, but different class F^*? E.g., perhaps P+T is characterised by the frames in which R is reflexive and everything is a counterpart of itself. What is the general *-operation? Does the counterpart relation always have to mirror the features of R: reflexive, transitive, final, etc.? Let's see.

–

Now for completeness theory.

> Theorem 8.6 (The logic of C-functional structures)
> QK.

Let's write X+(Y) for the logic that results by adding the principle (Y) to the axiomatic basis of X, as given above. (xxx is this relative to a choice of axiomatisation? Probably: e.g. if we axiomatise N by all its instances with zero rules, than adding e.g. (T) will not have the same effect as if we axiomatise it sensibly.)

Consider the class of C-functional structures. By lemma 3.9, in those models $\langle y/x \rangle A$ can be replaced by $[y/x]A$, and thus the unmodified versions of (LL) and (FUI) from non-modal logic are sound. So is, therefore, the necessity of identity, but not the necessity of non-identity. Thus functional models provide a general semantics for naive (extensions of) mixtures of F with K. And functional models in which things don't go out of existence provide the semantics for naive (extensions of) mixtures of Q with K.

> Theorem 8.7 (Completeness of QK)
> The logic Q+K is characterised by the class of C-functional C-total structures.

Proof TODO.

In PML, modal principles express constraints on the accessibility relation. Ideally, this should carry over to de dicto modal principles in QML. De re principles would then express constraints on accessibility and counterparthood. But is it true that e.g. (T) corresponds to reflexive R, without any constraint on counterparthood? And is it true that $\mathsf{P} + T$ is characterized by R-reflexive structures?

Define: A is valid in a *frame* F iff it is valid in all structures based on F.

To show: F is a frame for (T) iff F is R-reflexive.

Proof LTR. Suppose some w in F cannot see itself. Define V such that $V_w(p) = 0$ and $V_{w'}(p) = 1$ for all $w' \neq w$, where p is a zero-ary predicate. (Alternatively, we could use Fx with $C = Id$.) Then $w, V \nVdash \Box p \supset p$.

RTL. Suppose F is not a frame for (T), i.e. it is not the case that all instances of $\Box A \supset A$ are true at w, V, U, D, C based on F. So some instance of $\Box A \supset A$ is false for some w, V, U, D, C on F. So $w, V \Vdash \Box A$ and $w, V \neg \Vdash A$. The former means that $w', V' \Vdash A$ for all wRw' and $V_w \triangleright V'_{w'}$. If wRw, then in particular $w, V' \Vdash A$ for all $V_w \triangleright V_w$. But now it doesn't follow that $w, V \Vdash A$, for V need not be an image of itself at w!

So not all reflexive frames are frames for (T). In fact, is *any* non-trivial reflexive frame a frame for (T)? Can't we always cook up an interpretation and a counterpart relation that renders (T) false at some world that can see itself? We can. We just need to make Fx false at w, Fy true, and stipulate that $V_w(y)$ is the counterpart at w of $V_w(x)$, so that $\Box Fx$ is true.

Note that the present observation also shows that $\mathsf{P}+T$ is not sound and complete wrt reflexive frames, as it isn't sound.

Oh dear.

If features of R by themselves don't determine interesting systems, we might consider merging R and C. What's the combined feature of R and C that corresponds to $\Box A \supset A$? $\Box A \supset A$ is valid iff $w, V \Vdash A$ whenever $w', V' \Vdash A$ for all wRw' and $V_w \triangleright V'_{w'}$. This does mean, I think, that (a) wRw for all worlds and (b) everything is at least a counterpart of itself. Then

Maybe we can say that whenever a PML L is characterised by a class of frames F, then PL is characterised by a corresponding, but different class F^*? E.g., perhaps P+T is characterised by the frames in which R is reflexive and everything is a counterpart of itself. What is the general *-operation? Does the counterpart relation always have to mirror the features of R: reflexive, transitive, final, etc.?

Let's look at (free) quantified S4M, i.e. P combined with

(T) $\Box A \supset A$,

(4) $\Box A \supset \Box\Box A$,

(M) $\Box\Diamond A \supset \Diamond\Box A$.

$\mathsf{S4M}$ implies the following rule:

($\mathrm{R_{S4M}}$) if $\vdash_{S4M} \Diamond A$ and $\vdash_{S4M} \Diamond B$, then $\vdash_{S4M} \Diamond(A \wedge B)$.

Propositional S4M is characterised by the class of reflexive, transitive and final frames, where a frame is *final* if every world can see an "end" world that can only see itself. (S4M is also characterised by other sets of frames, e.g. by the frames in which every chain of accessible worlds is finite. This is not quite the same as saying that each world can see an "end" world: finality entails that R is anti-symmetrical, since otherwise there would be an infinite chain going back and forth between two worlds. But the "end" requirement does not entail anti-symmetry; e.g., we could have three worlds w_1, w_2, w_3 such that w_2 and w_3 can see everyone and w_1 can only see itself.)

To show this, we first show that $\mathsf{S4M}$ is sound wrt. reflexive, transitive and final frames. We know soundness of (T) and (4), so we only need to establish that $\langle W, R, V\rangle, w \Vdash \Box\Diamond A \supset \Diamond\Box A$ whenever $\langle W, R\rangle$ is final. So suppose for reductio that $\langle W, R\rangle$ is reflexive, transitive and final but (i) $\langle W, R, V\rangle, w \Vdash \Box\Diamond A$ and (ii) $\langle W, R, V\rangle, w \nVdash \Diamond\Box A$, i.e. $\langle W, R, V\rangle, w \Vdash \Box\Diamond\neg A$. (i) means that all worlds accessible from w can see a world where A is true. (ii) means that all worlds accessible from w can see a world where A is false. But any world that can be seen from any accessible world is itself accessible. So that world w_1 required by (i) where A is true is accessible from w and hence by (ii) it can see a world $w_2 \neq w_1$ where A is false; w_2 is also accessible from w so by (i) it can see a world $w_3 \neq w_2$ where A is true, and so on. So we never reach a world that can only see itself, contradicting the finality of $\langle W, R \rangle$.

To show that S4M is complete wrt. this class, we have to show that its canonical model is reflexive, transitive and final. We know that every world in the canonical model contains $\Box\Diamond A \supset \Diamond\Box A$, and that wRw' iff $A \in w'$ whenever $\Box A \in w$. What does the frame of this model look like? Let w be an arbitrary world, and let $\Box(w) = \{A : \Box A \in w\}$. A world is accessible from w iff it contains $\Box(w)$. We want to show that some such world can only see itself. Let $Fin = \{A \supset \Box A : A \in \mathcal{L}\}$. Evidently, $Fin \in w$ iff w can only see itself. (LTR because if $w' \neq w$, then there is some A false at w' and true at w, so if wRw' then $\Box A \notin w$; RTL because $A \wedge \neg\Box A$ requires that wRw' for some $w' \neq w$.) So all we have to show is that $\Box(w) \cup Fin$ is S4M-consistent. Suppose it is not. Then there are some $\underline{A}, \underline{B}$ such that $\underline{\Box A} \in w$ and

$$\vdash_{S4M} \neg(\underline{A} \wedge (\underline{B \supset \Box B})).$$

So

$$\vdash_{S4M} \underline{A} \supset \neg(\underline{B \supset \Box B}).$$

So by (Nec) and (K)

$$\vdash_{S4M} \underline{\Box A} \supset \neg\Diamond(\underline{B \supset \Box B}).$$

Since $\underline{\Box A} \in w$, it follows that $\neg\Diamond(\underline{B \supset \Box B}) \in w$. But by (T), $\vdash_{S4M} \Box\neg B \supset \neg B$, so $\vdash_{S4M} B \supset \Diamond B$. And so by ($\mathrm{R_{S4M}}$), $\vdash_{S4M} \Diamond(\underline{B \supset \Box B})$. So w contains $\Diamond(\underline{B \supset \Box B})$, contradicting its consistency.

Now what happens with quantified S4M in Kripke semantics? Consider constant domain semantics, where we must add the Barcan Formula to QS4M. The first part of the above proof goes through as before: QS4M+BF is valid in every RTF frame. (This follows from a general *frame transfer theorem*: if S is any normal propositional logic, then the frames for S are exactly the frames for the corresponding quantified system QS+BF.) For the second part, we have to show that the canonical model of QS4M+BF has an RTF frame. The problem arises when we want to prove finality. In canonical models of Kripke semantics, wRw' iff $A \in w'$ whenever $\Box A \in w$; so we want to show that for any given world w, some world w' with $\Box(w) \subseteq w'$ can only see itself. Now $Fin = \{A \supset \Box A : A \in \mathcal{L}^*\} \in w'$ iff w' is final. But even if $\Box(w) \cup Fin$ is QS4M+BF-consistent, we cannot directly conclude that there is a world w' with $\Box(w) \cup Fin \subseteq w'$, as $\Box(w)$ and Fin are not restricted to formulas in $\mathcal{L}$, so the extensibility lemma doesn't apply. (Why does it help to add $\Box\exists x Ax \supset \Diamond\exists x\Box Ax$?)

By contrast, in counterpart semantics (for substitution logics), wRw' iff $A^\tau \in w'$ whenever $\Box A \in w$. So we need to show that for any given world w, some world w' with $\Box^\tau(w) = \{A^\tau : \Box A \in w\} \in w'$ can only see itself. Note that the infinitely many variables of $\mathcal{L}$ do not occur in $\Box^\tau(w)$! We still have a problem with Fin, which can't just contain all instances of $A^\tau \supset \Box A^\tau$, otherwise some w' might contain Fin and yet see another world that only differs from w' wrt formulas in $\mathcal{L}$. (Can we extend $\Box^\tau(w) \cup Fin$ to a Henkin set, despite the fact that Fin contains all variables (although that doesn't look like a very general solution)? Fin itself is obviously not ω-inconsistent, i.e. it doesn't contain $\Phi(x)$ for all x but also $\exists x \neg \Phi(x)$, for no formula in Fin is equivalent to an existential formula. However, what if $\Box^\tau(w)$ contains $\exists x \neg(\Phi(x) \supset \Box\Phi(x))$, for some Φ? Then $\Box^\tau(w) \cup Fin$ is ω-inconsistent. $\exists x \neg(\Phi(x) \supset \Box\Phi(x))$ is equivalent to $\exists x(\Phi(x) \wedge \Diamond\neg\Phi(x))$...) (Note that if $A^\tau \in w'$, then $\langle x_1^\tau, \ldots, x_n^\tau : x_1, \ldots, x_n \rangle A \in w'$ for $x_1, \ldots, x_n = Var(A)$. By substitutional witnessing, there are new variables z_i such that w' contains $z_i = x_i^\tau$ as well as $[z_i/x_i]A$...)

In counterpart semantics, things need not be their own counterparts. So even if w can only see itself, $w, V \Vdash A$ does not entail $w, V \Vdash \Box A$ or even $w, V \Vdash \Diamond A$. For instance, we could have $w, V \Vdash Fx$ but $w, V \nVdash \Diamond Fx$. (What does this mean for the system Ver?!) In a canonical model, w can see itself iff w contains $\Box^\tau(w)$. How can we check if w can see another world? Will there be some A for which $A^\tau \in w$ and $\Box A \notin w$? If $\Box A \notin w$, then $\Diamond\neg A \in w$, so $\neg A^\tau \in w'$ for some accessible w'. So if w can only see itself, then $\Box A \in w$ whenever $A^\tau \in w$. The converse is given by wRw. So yes, w can only see itself iff for all A, $\Box A \in w$ iff $A^\tau \in w$, i.e. $\Box A \leftrightarrow A^\tau \in w$. To show that every world w sees an end world, we need to show that there is a w' with $\Box^\tau(w) \subseteq w'$ (so wRw') and $\{\Box A \leftrightarrow A^\tau : A \in \mathcal{L}^*\} \subseteq w'$. Can we weaken extensibility so that infinitely many variables do not occur outside modal operators?

How would that generalize? We'd like to explain how every completeness proof in PML carries over. It's no good if we have to ingeneously come up with a translation of e.g. Fin, case by case. For S4M in PML, we had to show that for every $w \in CM$, there is a world w' with $\Box(w) \subseteq w'$ and $\{A \supset \Box A : A\} \subseteq w'$. This turns into $\Box^\tau(w) \subseteq w'$ and $\{A^\tau \supset \Box A : A\} \subseteq w'$.

Take another example. To proof canonicity of T in PML, we have to show that every world in the CM of T can see itself. Every world w contains $\Box A \supset A$ for all A, and thus $\Box(w) = \{A : \Box A \in w\}$. So wRw by definition R. In QML, w again contains $\Box A \supset A$, but that doesn't give us $\Box^\tau(w)$. In fact, T is not characterized by reflexivity of R at all: it is characterized by reflexivity of C! This means that every point $\langle w, s \rangle$ of the CM can see itself in the sense that $\langle w, s_i \rangle C \langle w, s_i \rangle$ for all $s_i \in s$. This renders $\Box A \supset A$ sound, because $\Box A$ is true at $\langle w, s \rangle$ iff A is true at all points $\langle w', s' \rangle$ counterpart-related to $\langle w, s \rangle$, and then A is true at $\langle w, s \rangle$. It may help to think of points as $\langle w, V \rangle$ pairs, where all that matter in V is the assignment to variables. Now for completeness, note that in the CM of T, a point $\langle w, V \rangle$ can see $\langle w', V' \rangle$ iff each $s_i = V_w(x_i)$ has a member z_i such that $z_i^\tau \in s_i' = V_{w'}'(x_i)$. Since $\Box A \supset A \in w$, w contains A whenever $\Box A \in w$. How does this tell us that everything is a counterpart of itself? Can we say that $\Box A \in w$ iff $A^\tau \in w'$ for all accessible w'? Let's think of the worlds in CM as points $\langle w, V \rangle$, where w is a Henkin set and V is a counterfactual interpretation? $\langle w', \bar{d}' \rangle$ is accessible from $\langle w, \bar{d} \rangle$ iff each d_i' is $[x^\tau]_{w'}$ for some $x \in d_i$. E.g. an "initial point" might look like this:

$$\langle w, \langle \{x_1\}, \{x_2, x_3\}, \{x_2, x_3\}, \{x_4\}, \ldots \rangle \rangle$$

assuming $x_2 = x_3 \in w$. This point could see e.g.

$$\langle w', \langle \{x_1^\tau\}, \{x_2^\tau\}, \{x_2^\tau\}, \{x_4^\tau, x_5^\tau\}, \ldots \rangle \rangle$$

and

$$\langle w', \langle \{x_1^\tau\}, \{x_2^\tau\}, \{x_3^\tau\}, \{x_4^\tau, x_5^\tau\}, \ldots \rangle \rangle$$

assuming $x_2^\tau \neq x_3^\tau \in w'$ and $x_4^\tau = x_5^\tau \in w'$. Here all the x_i^τ are from $\mathcal{L}^+$, so the point is clearly not initial. In general, $\langle w, \bar{d} \rangle$ can see $\langle w', \bar{d}' \rangle$ iff there is a "thinning" $\bar{d}^t$ of $\bar{d}$, choosing a single member x from all non-singletons in $\bar{d}$ (not necessarily the same for each occurrence of d) such that $d'_i = \tau(d^t_i)$, i.e. $[x^\tau_i]_{w'}$ where x_i is the sole member left in d^t_i.

Now reflexivity means that $\langle w, \bar{d}\rangle$ can see $\langle w, \bar{d}\rangle$. How is that possible if x^τ is not even in $\mathcal{L}$? I guess we need $x_i \!=\! x^\tau_i \in w$ for all x_i. Right, that could also give us $\Box^\tau(w)$ from $\Box(w)$. So can we show that if w contains all instances of $\Box A \supset A$, then it also contains $x_i \!=\! x^\tau_i$? Surely not: the construction of Henkin sets assigns no special role to pairings x_i, x^τ_i.

Can we show that CM of T is *not* reflexive? Let Γ contain $x_1 \neq x^\tau_1$ as well as all $\mathcal{L}$-instances of $\Box A \supset A$. Γ is T-consistent. So it is part of a world w in CM(T). If CM(T) is reflexive, then for all w, wRw and for all d, $\langle d, w\rangle C \langle d, w\rangle$. The first condition requires that $A^\tau \in w$ whenever $\Box A \in w$. That may already fail, if e.g. we add to Γ the formulas $\Box F x_1$ and $\neg F x^\tau_1$. The second condition requires that $[x_1]_w C [x_1]_w$, i.e. there is some $z \in [x_1]_w$ with $z^\tau \in [x_1]_w$. This is a bit harder to render false explicitly, since we can't add $x_1 \!\neq\! z^\tau$ to Γ for all variables z as otherwise Γ contains every variable. However, obviously there are max cons extensions of Γ that contain no identity $x_1 \!=\! z^\tau$. Fuck.

(Is it even true that $\Box A \supset A$ is *true* at w? (It better be, by the truth lemma!) Suppose some instance is false. Then $w, V \Vdash \Box A$ and $w, V \nVdash A$. So $w', V' \Vdash A$ for all $wRw', V_w \triangleright V'_{w'}$. Then $w', V^\tau \Vdash A$ for all wRw'. Then $w', V \Vdash A^\tau$ by the transformation lemma. By induction, $A^\tau \in w'$. But then $\Box A \in w$ by definition of R. Also by induction, $A \notin w$. But then w is inconsistent given that it contains $\Box A \supset A$. The interesting point here is that whenever $\Box A$ is true at a world in a CM, then A^τ is true at all accessible worlds. It's not really surprising that $\Box A \supset A$ is valid in many non-reflexive models. It only requires that whenever all counterparts of $\bar{d}$ are Φ, then so is $\bar{d}$. Clearly it doesn't follow that $\bar{d}$ must be a counterpart of itself.)

…Return to the points in a CM. Since we have w as well as $\bar{d}$ to our disposal, we don't really need "think" individuals like $\{x_2, x_3\}$. We can recover

$$\langle w, \langle \{x_1\}, \{x_2, x_3\}, \{x_2, x_3\}, \{x_4\}, \ldots \rangle\rangle$$

from

$$\langle w, \langle \{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \ldots \rangle\rangle$$

from the fact that $x_2 = x_3 \in w$. So we can take worlds to be pairs of Henkin sets w and simple sequences of *variables*. But now it's a bit harder to explain accessibility. Why can $\langle w, \langle x_1, x_2, x_3, x_4, \ldots \rangle\rangle$ see $\langle w', \langle x^\tau_1, x^\tau_2, x^\tau_2, x^\tau_4, \ldots \rangle\rangle$? At any rate, we may need to redefine accessibility among points in CMs so that the CM for T comes out reflexive. ...... No, I have to make sure that $A \in w$ iff $A^\tau \in w$. Can't I simply add all $x_i \!=\! x^\tau_i$ to every Henkin set, perhaps saying that the substitutional witness of x^τ_i is always x_i? Probably not: wouldn't that destroy the difference between truth at a world as actual and as counterfactual? E.g. in the case where w contains all instances of $\Box\neg F x_i$ as well as $\Diamond\exists x F x$, I need an accessible world with all $\neg F x^\tau_i$ as well as a witness $F x_k$. That's not possible if $x_k = x^\tau_k$ is in the world. OTOH, couldn't I adjust the extensibility lemma so that we provide an arbitrary mapping from every variable x^τ in Γ to an unused variable x so that x will be used as substitutional witness for x^τ? No, I will clash with the existential witnesses: even if Γ contains all $\neg F x^\tau$, I need some witness Fz.

Now return to S4M: we need to assume finality of C, not R!

–

We can easily show that every counterpart structure for S4M must be reflexive, transitive and final. (Let $\mathcal{S}$ be a non-reflexive structure in which w cannot see itself, and let V be such that $w, V \nVdash \exists x F x$ and $w', V \Vdash \exists x F x$ for all $w' \neq w$. Since $\exists x F x$ has no free variables, its truth-value at a world w' does not vary between V and any image V' of V. So $w, V \Vdash \Box \exists x F x$, and $w, V \nVdash \exists x F x$. So (T) is not valid in $\mathcal{S}$. Similarly for (4). As to (M), let $\mathcal{S}$ be a reflexive, transitive, non-final structure in which w cannot see an end world. Then w can see some other world w' which (by transitivity) itself cannot see an end world, i.e. w' can see some other world w'' (possibly w), etc. Let V be such that $w, V \Vdash \exists x F x$, $w', V \nVdash \exists x F x$, $w'', V \Vdash \exists x F x$, etc., with alternating truth-values for $\exists x F x$ on the whole chain (though not necessarily consecutively alternating truth-values). Then $w, V \Vdash \Box \Diamond \exists x F x$, but $w, V \nVdash \Diamond \Box \exists x F x$. So (M) is not valid in $\mathcal{S}$.)

So every structure for P+S4M is reflexive, transitive and final. I.e., P+S4M is not sound with respect to any class of structures with non-reflexive, non-reflexive or non-final members: some theorems of P+S4M will be invalid in those structures. So if P+S4M is sound and complete with respect to some class of structures $\mathbb{S}$, then these structures are all reflexive, transitive and final.

All this holds also in Kripke semantics. But in Kripke semantics,

(10) $\quad \Box \exists x A \supset \Diamond \exists x \Box A$

is also valid in every reflexive, transitive and final structure. But it is not a theorem of P+S4M (see [Hughes and Cresswell 1996: 266–270] for a proof). Why is (10) valid in reflexive, transitive and final structures? We have to show that $w, V \Vdash \Box \exists x A$ entails $w, V \Vdash \Diamond \exists x \Box A$. So assume $w, V \Vdash \Box \exists x A$. Then $w', V \Vdash \exists x A$ at some final world w' accessible from w. Then $w', V' \Vdash A$ for some existential V'-variant V' of V at w'. Since w' can only see itself, then $w', V' \Vdash \Box A$. And then $w', V \Vdash \exists x \Box A$. But then $w, V \Vdash \Diamond \exists x \Box A$. (See [Hughes and Cresswell 1996: 266,283]).

This reasoning does not go through in counterpart semantics. Assuming $w, V \Vdash \Box \exists x A$, we know that $w', V' \Vdash \exists x A$ for some final w' accessible from w and some w'-image V' of V at w. Then $w', V'' \Vdash A$ for some existential x-variant V'' of V' at w'. Now $w', V'' \Vdash \Box A$ would require (given that w' can only see itself) that $w', V''' \Vdash A$ for all V''' with $V''_{w'} \triangleright V'''_{w'}$. This may fail if some individual at w' has some other individual at w' as a counterpart (perhaps in addition to itself). Suppose this is not so, and we have $w', V'' \Vdash \Box A$. Next, $w', V' \Vdash \exists x \Box A$ requires that $w', V'^{*} \Vdash \Box A$ for some existential x-variant V'^{*} of V' at w'. Since V'' is some such x-variant, this step goes through. Finally, $w, V \Vdash \Diamond \exists x \Box A$ requires that $w^{*}, V^{*} \Vdash \exists x \Box A$ for some wRw^{*} and $V_w \triangleright V^{*}_{w^{*}}$. Since w' and V' fit that condition, this step also goes through.

So to regain the incompleteness result, one would have to show that if P+S4M is valid in a structure $\mathcal{S}$, then no individual at any world in $\mathcal{S}$ has a different individual at the same world as its counterpart. But while (T) forces every individual to have itself as a counterpart, nothing in P+S4M rules out further counterparts at the same world. E.g. let $\mathcal{S}$ consist of $W = \{w\}$, $R = \{\langle w, w \rangle\}$, $D_w = \{x, y\}$, $C = \{\langle \langle w, x \rangle, \langle w, x \rangle \rangle, \langle \langle w, x \rangle, \langle w, y \rangle \rangle, \langle \langle w, y \rangle, \langle w, y \rangle \rangle\}$. Note that any interpretation V on this structure is an image of itself (at w). V may also have other images V' that assign y to some variables previously assigned x. Each such image V' still has itself as an image, but may have further images that re-assign more variables from x to y. Any such further image V'' of V' is also directly an image of V. Then (T) is valid: if $w, V \Vdash \Box A$, then $w, V' \Vdash A$ for all $V_w \triangleright V'_{w'}$, and since any V on $\mathcal{S}$ is one of its own images at w, then $w, V \Vdash A$. (4) is also valid: if $w, V' \Vdash A$ for all $V_w \triangleright V'_{w'}$, then $w', V'' \Vdash A$ for all V'' such that for some V'_w,

$V_w \triangleright V'_w \triangleright V''_w$, because each such V'' is also an image of V. Lastly, (M) is valid: if $w, V \Vdash \Box\Diamond A$, then for every image V' of V there is an image V'' of V' with $w, V'' \Vdash A$. Let V' be the image of V that maps all variables to y that were previously mapped to x. V' is its only image. So $w, V' \Vdash A$. Moreover, V' is an image of V such that $w, V'' \Vdash A$ for every image V'' of V'. So there is an image V' of V such that $w, V'' \Vdash A$ for every image V'' of V'. So $w, V \Vdash \Diamond\Box A$.

Can I prove completeness of $sP + (\text{S4M})$, using canonical models? Isn't it enough to show that the canonical model of that logic is final, transitive and reflexive, and that the logic is sound on that model?

In section 7 of [Kracht and Kutz 2005], Kracht and Kutz present a logic that is essentially P together with

(T) $\Box A \supset A$.

(4) $\Box A \supset \Box\Box A$.

(alt_2) $\Diamond A \wedge \Diamond B \wedge \Diamond C \supset \Diamond(A \wedge B) \vee \Diamond(A \wedge C) \vee \Diamond(B \wedge C)$.

(D2) $\exists x \exists y (x \neq y \wedge \forall z (z = x \vee z = y))$.

(W1) $\forall x \forall y (Fx \wedge \neg Fy \supset \Box(Fx \wedge \neg Fy \vee \neg Fx \wedge Fy))$.

(C2) $\forall x \forall y (Fx \wedge \neg Fy \supset \Diamond(\neg Fx \wedge Fy))$.

(The main difference between this and their logic is that they have outer quantifiers in place of our inner quantifiers. But this won't matter in what follows.) Suppose that this logic is complete with respect to some structure $\mathcal{S}$. If there is a world w in $\mathcal{S}$ that can see three worlds w_1, w_2, w_3, then we can find an interpretation such that three sentences A, B, C (e.g. $\exists x F_i x$, for three predicates F_i) are true at exactly one of w_1, w_2, w_3 each, and so (alt_2) is false at w. Hence no world in $\mathcal{S}$ can see more than two worlds. By (T), each world can see itself. By (4), each world can therefore reach at most one world besides itself.

(D2) ensures that each world contains exactly two individuals (in the inner domain). (W1) says that if two (inner) individuals differ in a certain respect F, then all their counterparts at accessible worlds also differ in that respect. Thus let x, y be distinct (inner) individuals at w, so that we can let them differ w.r.t F. Suppose there is a world w' accessible from w. Since $\mathcal{S}$ is positive, x and y must have counterparts x', y' at w'. This is OK as long as $w' = w$, and x', y' are x and y themselves (in some order). But if $w' \neq w$, or $w' = w$ and x', y' are the same individual, or $w' = w$ and at one or both of x', y' is not x or y, then there will be an interpretation under which x' and y' agree with respect to F, in contradiction to (W1). So all worlds in w can only see themselves, and their two inhabitants x, y must have either x, y or y, x or both as counterparts. (Note that we don't need the extra axioms (NI) and (NNI) that Kutz includes. Can we also get rid of T, 4, alt?)

For (C2), choose an interpretation on which x is F and y isn't. Then (C2) says that, first, there are counterparts x', y' of x and y at some w' such that x' is F and y' isn't, and second, that there are also counterparts x', y' at some w' such that y' is F and x' isn't. This means that x, y must have both x, y and y, x as counterparts.

So this is a logic that requires non-trivial intra-world counterparts. It therefore presents trouble for "conceptual" rivals to counterparts semantics (e.g. the "coherence models" of [Kracht and Kutz 2005]). But there is no problem for our own account here.

xxx Since CP and CN are just strengthened versions of P and N, I might drop them from the base logics at this point, and talk about them later, when I look at extensions of the base logics. At that point, I could also point out that substitution is fully definable in CC, and that lambda substitution might do in FUI and LL in CN (and perhaps in CP).

9 Starting over

There is a problem with my canonical models. Consider the model for $\mathsf{P}+\mathsf{T}$. We'd like the structure of this model to be reflexive in some sense, to prove completeness. (Arguably $\mathsf{P}+\mathsf{T}$ is valid in a structure iff that structure is reflexive, in the sense that every world can see itself and everything is its own counterpart at its own world. So if the canonical model for $\mathsf{P}+\mathsf{T}$ isn't reflexive, its structure is not a structure for $\mathsf{P}+\mathsf{T}$, and we can't use the technique to prove completeness.) But on the above construction, the canonical model of $\mathsf{P}+\mathsf{T}$ is not reflexive. Let Γ contain $x_1 \neq x_1^\tau$ as well as all $\mathcal{L}$-instances of $\Box A \supset A$. Γ is $\mathsf{P}+\mathsf{T}$-consistent. So it is part of a world w in CM($\mathsf{P}+\mathsf{T}$). If CM($\mathsf{P}+\mathsf{T}$) is reflexive, then for all w, wRw and for all d, $\langle d, w\rangle C\langle d, w\rangle$. The first condition requires that $A^\tau \in w$ whenever $\Box A \in w$. That may already fail, if e.g. we add to Γ the formulas $\Box Fx_1$ and $\neg Fx_1^\tau$. The second condition requires that $[x_1]_w C[x_1]_w$, i.e. there is some $z \in [x_1]_w$ with $z^\tau \in [x_1]_w$. This is a bit harder to render false explicitly, since we can't add $x_1 \neq z^\tau$ to Γ for all variables z as otherwise Γ contains every variable. However, obviously there are max cons extensions of Γ that contain no identity $x_1 = z^\tau$.

What went wrong? The problem is that we've defined canonical counterparthood in a fixed, external manner. Compare accessibility: whether w' is accessible from w depends on whether it verifies all formulas A (or A^τ) which w claims to be true at all accessible worlds. By contrast, whether $[y]_{w'}$ is a counterpart of $[x]_w$ has nothing to with whether $[y]_{w'}$ at w' verifies every condition that w claims to be true for all counterparts of x. Of course, we didn't completely ignore this – it was supposed to be taken care of by requiring $A^\tau \in w'$ in the clause for R: this ensured that if $\Box A(x) \in w$, then $A(x^\tau) \in w'$, so that w' verifies that the counterpart $[x^\tau]_{w'}$ of $[x]_w$ satisfies condition A. On the other hand, this doesn't tell us that *anything* that satisfies $A(x)$ for all $\Box A(x) \in w$ qualifies as counterpart of $[x]_w$. If we had this, it would be easy to show that the CM of $\mathsf{P}+\mathsf{T}$ is reflexive: since every w contains $\Box A(x) \supset A(x)$, $[x]_w$ at w must be a counterpart of itself at w.

So that's what we need: we need to define counterparthood in such a way that we can read off whether $\langle [x]_w, w\rangle C\langle [y]_{w'}, w'\rangle$ by comparing what w says about the boxed properties of x and what w' says about y. (Call this the *new idea*.)

This doesn't immediately show that our old approach can't be fixed. Perhaps we only need to add a clause to the effect that whenever $[y]_{w'}$ at w' satisfies the boxed properties of x at w, then w' contains $y = x^\tau$. Now that won't work, because $[x]_w$ can have several counterparts at w', not all identical to $[x^\tau]_{w'}$. So we'd need to say that whenever $[y]_{w'}$ at w' satisfies the boxed properties of x at w, then $[x]_w$ contains some z with $y = z^\tau$. Perhaps this could be somehow turned into a constraint on R? (Worlds w' that don't satisfy the condition are not accessible.)

But maybe we should make bigger changes. Recall that the true points of evaluation in a counterpart model are pairs of a world and an assignment function, $\langle w, V\rangle$. Both $\Box$ and $\forall x_i$ shift the point of evaluation. In the canonical model, the relevant assignment function V can be represented by a substitution s; the corresponding V (i.e., V_w restricted to variables) is then recoverable as $V(x) = \{z : s(x){=}z \in w\}$. The initial interpretation V at w is represented by the identity substitution, since V maps each x to $\{z : x{=}z \in w\}$. Now we could say that a point $\langle w', s' \rangle$ is accessible from $\langle w, s \rangle$ iff $s'(A) \in w'$ whenever $s(\Box A) \in w$. The truth lemma would say that for every point w, s in the canonical model, $w, s \Vdash A$ iff $s(A) \in w$. The proof for $\Box B$ would be simple:

$$
\begin{aligned}
& w, s \Vdash \Box B \\
\Leftrightarrow & w', s' \Vdash B \text{ for all } w', s' \text{ s.t. } s'(A) \in w' \text{ whenever } s(\Box A) \in w \\
\Leftrightarrow & s'(B) \in w' \text{ for all such } w', s', \text{ by i.h.} \\
\Leftrightarrow & s(\Box B) \in w
\end{aligned}
$$

One worry with this approach is that the equivalence between $w', s' \Vdash B$ and $s'(B) \in w'$ effectively relies on an unrestricted substitution lemma, for $s'(B) \in w'$ also means that $w', = \Vdash s'(B)$, so we'd assume equivalence between $w', =^{s'} \Vdash B$ and $w', = \Vdash s'(B)$. However, perhaps we can stipulate that in every point $\langle w, s \rangle$, the substitution s is a transformation. This would rule out that at some canonical world under some image of V, two distinct variables x, y denote the same *singleton* $\{z\}$, so that $s(x) = s(y) = z$. But it's not clear why we'd need that. Of course x and y can corefer under some image at some w', but they needn't corefer to a singleton. (OTOH, this is not enforced by our old construction, where we might have $[x]_w = [y]_w = \{x, y\}$ and $[x^\tau]_{w'} = \{x^\tau\} \neq \{y^\tau\}$, so that one image maps both x and y to $\{x^\tau\}$.)

Now can we express the present accessibility relation on canonical points in terms of a relation R between worlds and a relation C between world-indidvidual pairs? Well, we could say that wRw' iff for every variable-free A, $A \in w'$ whenever $\Box A \in w$, and that $\langle [x]_w, w \rangle C \langle [y]_{w'}, w' \rangle$ iff $A(y) \in w'$ whenever $\Box A(x) \in w$, or some such – in line with the *new idea* above. But I don't think this works because it doesn't keep track of relations between the counterparts of different variables. A counterpart relation between sequents may not be determined by a counterpart relation between their elements. And perhaps we need a non-trivial (non-elementwise) relation between sequents to ensure injectivity (transformativity).

In fact, there is an obvious problem with the new idea. Suppose w contains $x \neq y, \Box(x \neq y), \Box Fx, \Box Fy$ and no other non-trivial boxed formula involving x or y. The idea is that x and y are said to be distinct and necessarily distinct, but that they have the same modal profile. An elementwise counterpart relation, following the new idea, threatens to assign the same counterparts to x and y at every world, making their distinctness contingent. But in fact, how would we even apply the idea, given that one of the boxed formulas is $\Box(x{\neq}y)$? This doesn't just say something about the counterparts of x in isolation. It constrains the counterparts of x and y together.

All this is reminiscent of the old observations in Hazen and Lewis against elementwise counterpart relations: it might be that the pair $\langle x, y \rangle$ at w has two joint counterparts at w', $\langle x', y' \rangle$ and $\langle x^*, y^* \rangle$, while e.g. $\langle x', y^* \rangle$ is not an eligible choice of counterparts. My present semantics doesn't respect that, although we can mimick the desired results by assuming qualitatively indistinguishable worlds, as long as we don't have counterfactuals or nominals that would specifically direct us at a particular world, in which case $\textit{at } a : (Fx \land Fy) \land \textit{at } a : (Gx \land Gy)$ would entail $\textit{at } a : (Fx \land Gy)$.

So should we change the notion of a structure to allow for counterparts between finite sequences? Finite sequences are enough because every de re formula has the form $\langle \bar{y} : \bar{x} \rangle \Box A$ or $\langle \bar{y} : \bar{x} \rangle \Diamond A$

etc., and then we only need to look for counterparts of $\bar{y}$. I might still mention that the C relation between sequences might be determined elementwise by a simpler C relation between individuals and that this does not affect the logic.

Moving to sequences also fits the idea of counterpoints as world-sequence pairs, which isn't tied to canonical models. (We also talk that way in correspondence theory.) Strictly speaking, a counterpoint is a pair of a world w and an infinite sequence of members of U_w, representing the assignment function. When we shift the point of evaluation, wouldn't we thus need a counterpart relation between *infinite* sequences? In a sense yes, but only a finite segment will matter. Still, does that mean we have to look at the variables that occur in A in order to say how to evaluate $\Box A$ at a point? $w, s \Vdash \Box A$ iff $w', s' \Vdash A$ for all wRw' and s' such that for the sequence of variables $\bar{x}$ in A, $s'(\bar{x})$ is a counterpart of $s(\bar{x})$. That's inelegant, and looks like Lewis's account. For what it's worth, the philosophical motivations for moving to sequences surely apply to infinite sequences as well.

There are different ways of formally defining counterparthood for sequences. In the end, what we need is a notion that allows us to define the imaging relation $\triangleright$ between interpretations, i.e. between infinite sequences of individuals (with repetition). This was easy with counterparts for single individuals: we only needed to say that each position in the first sequence is counterpart-related to the corresponding position in the second. (Note that this does not require repetitions to be preserved: if the first sequence begins $a, a, \ldots$, the second may begin $a', b', \ldots$, allowing for the fact that $x = y$ is true at w under V but false at w' under V'.) If we also allow pairs and triples etc. to have counterparts, it is less obvious how this translates into an imaging relation. If a and b both have a' and b' as counterparts, but a, b only has a', b' as counterpart, should we allow e.g. $b', a', \ldots$ as image of $a, b, \ldots$? More strangely, what if a, b has a', b' as counterpart, but a alone does not have a' as counterpart? Should this be forbidden? Intuitively, we might want to say that as long as we evaluate a formula that only contains x free, then we ought to look at images that assign to a counterparts of a alone, while if the relevant formula contains x and y free, we need to look at counterpart pairs. But this is hard to enforce in the framework of standard Tarski semantics, where all formulas are always evaluated relative to infinite sequences. – It motivates moving to a Lawvere or Jacobson style alternative.

The point where I really need counterpart sequences in the canonical model construction, where I want to read off the counterparts of individuals (variable classes) at other worlds by what these worlds say: if w' says Fx' for every $\Box Fx \in w$, then $[x']_{w'}$ is a counterpart of $[x]_w$ (loosely speaking). Analogously, if w' says $Fx'y'$ for every $\Box Fxy \in w$, then $[x']_{w'}, [y']_{w'}$ is a counterpart pair of $[x]_w, [y]_w$. Call a structure *reflexive* if every individual at every world is a counterpart of itself at that world, as is every pair, triple, etc. Since every world in the canonical model for T contains $\Box A(x) \supset A(x)$ and $\Box A(x, y) \supset A(x, y)$ etc., it comes out as reflexive – which is what we want. Now how do we show that $\Box A(x) \supset A(x)$ is *true* at every world? Well, $\Box A(x)$ is true at w, V iff $A(x)$ is true at all accessible w', V'. Can we say that whether w', V' is accessible depends on the free variables in A? Then we get the strange results of Lewis's logic! We might have $\Box(Fx \wedge Gy)$ without $\Box Fx$. (As in the example above: a', b' are the only counterparts of a, b, but a alone also has b' as counterpart.) If we don't want that, we must say that whether or not $\Box A$ is true at w, V does not depend on which variables occur in A. Put differently, we must assume that counterparts between single individuals always show up in counterparts of sequences, so that we get the very same a counterparts whether we look at a alone or at any sequence beginning with a.

So when is $\Box A$ true at w, V? Whenever A is true at all accessible w', V', i.e. for all wRw', $V_w \triangleright V_{w'}$. In effect, V_w is a sequence of individuals from U_w, perhaps with repetitions (or blanks). $V_{w'}$ is such a sequence from $U_{w'}$. We can't just compare the sequences individual-wise. What we could do is say that if V_w is $a_1, a_2, a_3, \ldots$ then $V_{w'} = a'_1, a'_2, a'_3, \ldots$ must be such that each a'_i is a counterpart of a_i, each a'_i, a'_j is a counterpart of a_i, a_j, etc., for all (finite) sequences taken from the elements of V_w and $V_{w'}$. That requires counterparthood to be defined for all finite sequences of individuals, which is a bit odd. Alternatively, we could simply look at the one infinite sequence V_w and require $V_{w'}$ to a counterpart sequence of it. This requires counterparthood to be defined for all countable sequences of individuals, which is in a sense even worse, but it makes it clear that this kind of counterparthood is a technical notion to be defined in terms of intuitive constraints ("a can only be mapped to a', b and c must be mapped to either b', c' or c', b', etc.").

A further issue concerns repetitions. Suppose V_w is $a, a, \ldots$. We certainly want to allow for an image of V_w where the first two positions come apart: $a', b', \ldots$. But do we want a', b' or $a', b', \ldots$ as candidate *counterpart* of a, a or $a, a, \ldots$? What would that mean? At w, we have one individual a, denoted by both x and y. At w', there are two individuals a' and b'. It makes sense to say that a has both of these as counterparts. But does it make sense to say that the pair a, a has a', b' as counterpart, perhaps without having b', a' as counterpart? Then a would be a' at w' *qua* x but not *qua* y. We'd get a violation of $\vdash x = y \supset \Diamond Fx \supset \Diamond Fy$. Perhaps it's worth investigating this possibility – it would need even more restrictive substitution principles and would in many ways look like concept semantics. But it's not my main object of study.

One thing we could do is restrict counterpart relations so that if $a, a, \ldots$ has $a', b', \ldots$ as counterpart, it also has $b', a', \ldots$ as counterpart. But that looks a bit ad hoc. Perhaps it's better to say that counterparthood is only defined between sequences without repetitions. So multiple counterparthood is not reflected by something like $\langle w, a, a, \ldots \rangle C \langle w', b, c, \ldots \rangle$, but rather by a pair like $\langle w, a, \ldots \rangle C \langle w', b, \ldots \rangle$ and $\langle w, a, \ldots \rangle C \langle w', c, \ldots \rangle$. But how do we represent the fact that a and b at w both have a' as unique counterpart at w'? Doesn't that mean that the pair a', a' is a counterpart of a, b? Which other pairs (or longer sequences) would be counterparts of a, b (or $a, b, \ldots$)? It looks like we have to allow for repetitions on the right-hand side of counterparthood, if not on the left.

In any case, this means we can't define counterparthood in general as a relation between ω-sequences, since U_w may not contain infinitely many individuals. We could still try the alternative approach, where we take counterparthood between finite sequences (on the left, without repetitions) as basic and say that $a_1, a_2, a_3, \ldots \triangleright a'_1, a'_2, a'_3, \ldots$ iff each a'_i is a counterpart of a_i, each a'_i, a'_j with a_i, a_j non-identical is a counterpart of a_i, a_j, etc. Or something like that.

This is getting ugly. But let's march on.

We can plausibly impose some constraints on counterparthood between sequences. In particular, counterparthood for individuals is always reflected in counterparthood for sequences: if a has both a' and a'' as counterpart and b has b' and b'' (say), then ab must have a counterpart beginning with a' and one with a'', and one ending with b' and one with b'' (i.e. $\{a'b', a''b''\}$ or $\{a'b'', a'', b'\}$ or $\{a'b', a''b', a''b''\}$ etc.). So the strange cases mentioned above can be stipulated away. In this case, we could even let the box only impose a re-interpretation of the free variables in its scope (which could be made more explicit by moving to a syntax like Ghilardi's etc.), without getting a deviant logic: $\Diamond(Fx \wedge Gy)$ would always entail $\Diamond Fx$. But with the present syntax, that's not the most natural rule. So we change the whole sequence V_w when moving from w to w'.

Similarly with repetitions. It makes little sense to think that a, a could have a', b' as counterpart but not b', a'. That's because the ordering in the pairs is really irrelevant: we want to know how to locate a and b, or a and a at w', no matter how they are ordered. In a sense, it is therefore misleading to construe counterparthood as a relation between sequences. Rather, we want multiple relations between individuals. For example, relation 1 links a with a' and b with b', relation 2 a with b' and b with a', while no relation links a with a' and b'; or: relation 1 links a with both a' and b' and therefore, trivially, with both b' and a'. Aha! That's what Kutz was after!

Is there another solution to my canonicity problems? Suppose we have only individual counterparts. How do we read off the counterparts of $[x], [y]$ in w at another Henkin set w'? Suppose w contains $\Box Fxy$. Can't we take this as a joint constraint on the counterpart(s) assigned individually to $[x]$ and $[y]$? Any assignment of individual counterparts determines an assignment of counterparts to sequences. Suppose w' contains $Fx'y'$. At this point, $[x']$ is a candidate counterpart for $[x]$ and $[y']$ for $[y]$. What's the problem? Well, remember the case where w contains $x \neq y, \Box(x \neq y), \Box Fx, \Box Fy$ and no other non-trivial box formula. An elementwise counterpart relation seems to assign the same counterparts to x and y at every world, making their distinctness contingent. But couldn't I here say that canonical counterparts are not to be *defined* elementwise? The individual counterparts assigned to $[x], [y], \ldots$ have to fit together. Like so perhaps: $[x'], [y'], \ldots$ at w' are (individually) counterparts of $[x], [y], \ldots$ at w iff there is some transformation τ such that (i) $A^\tau \in w'$ whenever $\Box A \in w$ and (ii) there are $\underline{z}$ such that $z_i \in [x]$ and $z_i^\tau \in [x']$. In the example, any relevant w' must contain $x' \neq y'$ as well as Fx' and Fy' for some x', y', in order to meet (i). Now what are the counterparts of $[x]$ and $[y]$ at w'? The problem is still there. For both $[x'], [y']$ *and* $[y'], [x']$ seem to qualify. But then $[x]$ seems to have both $[x']$ and $[y']$ as counterpart, and so does $[y]$.

Perhaps the problem can be overcome. After all, we can see which assignments of counterparts are OK and which aren't. Can't we turn that into a definition?

Let's think again. Take a concrete example. w contains $\Box Fx, \Box Fy, \Diamond Gxy, \ldots$, thereby specifying a modal profile for x, y, etc. (more precisely, for $[x]_w, [y]_w$). Another Henkin set w' may contain $Fx, Fz, Gxz, \ldots$. From the perspective of w (i.e. considered as counterfactual), this isn't a state at which e.g. x is F. To say what w' represent about x, we first have to locate x at w', by finding its counterparts.

In Kripke semantics, that's easy: x is always its unique own counterpart at any world; more precisely, if $[x]_w = \{x, y, \ldots\}$, then $[x']_{w'}$ is a counterpart of $[x]_w$ iff $[x']_{w'} = \{x, y, \ldots\}$. Here there is an externally fixed counterpart relation. To allow for contingent identity, we could relax this clause and say that $[x]_w$ has $[x']_{w'}$ as counterpart iff there is some z that occurs both in $[x]_w$ and $[x']_{w'}$. (Now we can have $x = y \in w$ but $x \neq y \in w'$, in which case $[x]_w = [y]_w$ has both $[x]_{w'}$ and $[y]_{w'}$ at w' as counterparts.) To help with the problem of modal witnessing, we could fix a different counterpart relation on which, for example, v_n always has v_{n+1} as counterpart. That's not enough because it only frees a single variable. So we better pick some (arbitrary) substitution σ whose range excludes infinitely many variables and say that $[x]$ always has $[x^\sigma]$ as counterpart, i.e. $[x]_w$ has $[x']_{w'}$ as counterpart iff there is a $z \in [x]_w$ with $z^\sigma \in [x']_{w'}$. It proves convenient to let σ be a transformation τ. This approach works to some extent: one can prove completeness for all four basic logics. But it runs into problems when we look at stronger logics. For example, it is easy to see that $\mathsf{P}+\mathsf{T}$ is valid in a structure iff every world can see itself and all things are their own counterparts. So to prove structure completeness for $\mathsf{P}+\mathsf{T}$, we want the canonical model of $\mathsf{P}+\mathsf{T}$ to be reflexive in this sense. But it won't be. (Let Γ contain $x_1 \neq x_1^\tau$ as well as all $\mathcal{L}$-instances of $\Box A \supset A$. Γ is $\mathsf{P}+\mathsf{T}$-consistent. So it is part of a world w in the canonical model. If the model is reflexive, then for all w, wRw and for all d, $\langle d, w\rangle C\langle d, w\rangle$. On a plausible definition of canonical accessibility, the first condition requires that $A^\tau \in w$ whenever $\Box A \in w$. That already may fail, if e.g. we add to Γ the formulas $\Box Fx_1$ and $\neg Fx_1^\tau$. The second condition requires that $[x_1]_w C[x_1]_w$, i.e. there is some $z \in [x_1]_w$ with $z^\tau \in [x_1]_w$. This is a bit harder to render false explicitly, since we can't add $x_1 \neq z^\tau$ to Γ for all variables z as otherwise Γ contains every variable. However, obviously there are max cons extensions of Γ that contain no identity $x_1 = z^\tau$.)

This shows that we shouldn't define canonical counterparthood in a fixed, external manner. Compare accessibility: whether w' is accessible from w depends on whether it verifies all formulas A (or A^τ) which w claims to be true at all accessible worlds. By analogy, we should say that whether $[x']_{w'}$ is a counterpart of $[x]_w$ is determined by whether $[x']_{w'}$ satisfies the modal profile attributed to $[x]_w$ in w. The above proposal ensured that if $\Box A(x) \in w$, then $A(x^\tau) \in w'$ for accessible w', so that w' verifies that the counterpart $[x^\tau]_{w'}$ of $[x]_w$ satisfies condition A. But this doesn't tell us that *everything* that satisfies $A(x)$ for all $\Box A(x) \in w$ qualifies as counterpart of $[x]_w$. If we had this, it would be easy to show that the CM of $\mathsf{P}+\mathsf{T}$ is reflexive: since every w contains $\Box A(x) \supset A(x)$, $[x]_w$ at w must be a counterpart of itself at w.

So we need to define counterparthood in such a way that we can read off whether $\langle [x]_w, w\rangle C\langle [y]_{w'}, w'\rangle$ by comparing what w says about the boxed properties of x and what w' says about y. It's as if w', considered as counterfactual, were a merely qualitative description of a world (saying that there is some x, some u, some v, etc. satisfying such-and-such conditions), and now we need to figure out which of these x, u, v, etc. qualify as representatives of $[x]_w$, $[y]_w$, $[z]_w$, etc. (Although, of course, we don't stipulate that this is a matter of qualitative similarity: w comes with built-in claims about modal profiles.)

But now there's another problem. w doesn't just constrain the modal profile of individuals one by one, but also in relation to one another. It might say that $\Box Gxy$ or $\Box x \neq y$, etc. And this matters. Suppose w contains $\Box Gxy, \Diamond\neg Gyx, \Box Fx, \Box Fy$, and no other interesting modal statement about x and y. We need an accessible world w' that verifies $Gxy, \neg Gyx, Fx, Fy$ considered as counterfactual. We can easily find a w' containing $Gx'y', \neg Gy'x', Fx', Fy'$ for some variables x', y'. But now which of x', y' is counterpart of x, y, respectively? We must not say that x has both x' and y' as counterpart, and so does y. For then Gyx comes out true at w' as counterfactual.

In philosophy, this problem was noticed by Hazen and Lewis, who argued that we have to allow for counterparts between pairs, triples, etc., rather than just between individuals. This way, we can say that x', y' is a counterpart pair for x, y, and so is y', x', but not x', x' or y', y'. We could go this route...

An alternative to counterpart relations between sequences is to go haecceitistic and introduce "qualitatively indistinguishable", haecceitistic worlds. A haecceitistic world is one in which every individual carries a marker (a "haecceity") which specifies which actual individual it represents. So the world w' containing $Gx'y', \neg Gy'x', Fx', Fy'$ is really two worlds, one marking x' as counterpart of x and y' for y, the other marking them the other way round. This is easily achieved by defining worlds in the canonical model to be pairs of an arbitrary Henkin set and an arbitrary variable transformation. Where previously the pair x, y had two counterparts x', y' and y', x' at the relevant world, we now have two worlds $w, [x', y'/x, y]$ and $w', [y', x'/x, y]$ in each of which there is only one counterpart pair.

Won't that rule out contingent identity and distinctness? Won't worlds with multiple counterparts always be turned into multiple worlds with single counterparts? This happens in standard (extreme) haecceitism in philosophy. But not here. We've already seen the reason: even if we fix a particular transformation τ to find the counterparts of $[x]_w$ at w', we can get multiple counterparts.

A third option is Kutz's, who introduces multiple counterpart relations between worlds and interprets the box to require truth on all of them. I haven't thought carefully about whether and why this works, because I find it inelegant.

Now let's see what happens when we take canonical worlds to be pairs of a Henkin set and a substitution, staying with individual-wise counterpart relations. This is what I already did in ct2, and then gave up for unclear reasons. It has some nice features: (1) The very same account can be used for model theories in which there are different counterpart relations (indexing different modal operators); here the canonical worlds would be simply the Henkin sets and the associated substitutions would represent different paths of getting to the world. (2) Perhaps the same account could even be used to build canonical models for sequenced counterpart theories; here again the canonical worlds are simply the Henkin sets and the substitutions determine counterparts of sequences: if H' is accessible from H via τ, then $\underline{[x^\tau]_{H'}}$ is a counterpart of $\underline{[x]_H}$, roughly.

(Compare Hazen's model theory: "A stipulational world is either the actual world or an ordered pair of a world and a function from the domain of the actual world to its domain. Truth at a stipulational world on an assignment is defined in the usual way (for atomic formulas, truth-functional compounds, and quantified formulas) if the world is the actual world, and in the usual way, but relative to a reinterpretation of the individual constants assigning them the images, if any, under the function, of their usual denotata" [Hazen 1979: 336].)

Let's first check that we get a reflexive canonical model for $\mathsf{P} + \mathsf{T}$. A world $\langle H, \tau \rangle$ in the CM can see itself iff $A^\tau \in H$ for each $\Box A \in H$. Oops: $\Box A \supset A$ ensures that $\langle H, 1 \rangle$ can see itself for all H, but not $\langle H, \tau \rangle$ for arbitrary τ. Maybe we have to move to pointed models and say that the centre point always has the identity transformation? (Why? Would that help e.g. with $\Box T$ systems?) Maybe we can instead redefine accessibility: $\langle H, \tau \rangle$ can see $\langle H', \tau' \rangle$ iff $(A^\tau)^{\tau'} \in H'$ whenever $\Box A^\tau \in H$. After all, truth at $\langle H, \tau \rangle$ isn't membership in H, but membership in H under τ. If all instances of $\Box A \supset A$ are in H, then in particular $\Box A^\tau \supset A^\tau \in H$, for any transformation τ. Does it follow that $\langle H, \tau \rangle$ can see itself? That would require that $(A^\tau)^\tau \in H$ whenever $A^\tau \in H$, which isn't guaranteed.

Hm. I kinda want to invoke transformations only when looking at other worlds as counterfactual. When evaluating formulas at a Henkin set H, I want to look at *other* Henkin sets H' under transformations that respect the modal profiles in H (including joint profiles).

Now recall that the true points of evaluation in a counterpart model are pairs of a world and an assignment function, $\langle w, s\rangle$. The "initial" assignment s at w maps every variable x to $[x]_w$. Other assignments s' might map x to $[y]_w$ etc. Given the identity statements in w, these assignments can be represented by a substitution σ such that $\sigma(x)$ is some element of $s'(x)$; we can recover $s'(x)$ as $\{z : z{=}\sigma(x) \in w\}$. If $\{y\}$ and $\{z\}$ both have $\{x\}$ at w as counterpart, then $s'(y) = s'(z) = \{x\}$, so $\sigma(y) = \sigma(z) = x$ and σ is not a transformation. That complicates things. We proceed by hoping that nothing breaks if we restrict counterparthood so that it is always reflected by a transformation. (This rules out that at some canonical world under some image of V, two distinct variables x, y denote the same *singleton* $\{z\}$, so that $s(x) = s(y) = z$. But it's not clear why we'd need that. Of course x and y can corefer under some image at some w', but they needn't corefer to a singleton. OTOH, this is not enforced by our old construction, where we might have $[x]_w = [y]_w = \{x, y\}$ and $[x^\tau]_{w'} = \{x^\tau\} \neq \{y^\tau\}$, so that one image maps both x and y to $\{x^\tau\}$.)

Now the idea is that since every assignment function in canonical models is induced by a transformation, we can imagine that formulas are evaluated at pairs of a world and a transformation. The 'initial' point for w is $\langle w, 1\rangle$. Another point $\langle w', s\rangle$ is accessible from $\langle w, 1\rangle$ iff $1^s(A) \in w'$ whenever $1(\Box A) \in w$. For the truth lemma, we then show that $w, s \Vdash A$ iff $s(A) \in w$. For $\Box A$: $w, s \Vdash \Box A$ iff $w', s' \Vdash A$ for all w', s' with $s'(A) \in w'$ whenever $s(\Box A) \in w$, iff $s'(A) \in w'$ for all w', s' with $s'(A) \in w'$ whenever $s(\Box A) \in w$ (i.h.), iff $s(\Box A) \in w$.

We must be clear on the relation between assignments, sequences, and substitutions. The objects of the counterpart relation are sequences of individuals, not substitutions, so we can't directly apply them to formulas. Moreover, a sequence of individuals generally corresponds to multiple different substitutions. If w contains $x_2 = x_3$, the initial assignment s at w might look like this (as a sequence):

$$\{x_1\}, \{x_2, x_3\}, \{x_2, x_3\}, \{x_4\}, \ldots$$

Substitutions can also be pictured as sequences, this time of variables. For example, the identity substitution 1 is

$$x_1, x_2, x_3, x_4, \ldots$$

The above assignment s is generated by the identity substitution, but also by three others:

$$x_1, x_2, x_2, x_4, \ldots$$
$$x_1, x_3, x_2, x_4, \ldots$$
$$x_1, x_3, x_3, x_4, \ldots$$

Perhaps it does not matter which substitution is used, because $s_1(A) \in w$ iff $s_2(A) \in w$ for all candidates? Or perhaps we can stipulate here that the relevant substitutions must be transformations?

Can we define the counterpart relation C in a canonical model to hold between $\langle w, \omega\rangle$ and $\langle w', \omega'\rangle$ iff there is some transformation τ such that (i) $\omega'(x) = \omega(x^\tau)$ and (ii) w' contains A^τ whenever w contains $\Box A$? Condition (i) means that if ω is

$$\{x_1\}, \{x_2, x_3\}, \{x_2, x_3\}, \{x_4\}, \ldots$$

then ω' must look, for example, like this (for τ as $+1$):

$$\{x_2, x_3\}, \{x_2, x_3\}, \{x_4\}, \ldots$$

But that's odd: we preserve the identity facts from w when looking at w'. If w' doesn't contain $x_2 = x_3$, then $\{x_2, x_3\}$ isn't even an individual at w'. It's not obvious that this is a problem ($x_2 \neq x_3$ can still be false at w' as counterfactual, i.e. we can have $x_2^\tau \neq x_3^\tau \in w'$, as we do in the example), but it's still strange. So it might be better to replace condition (i) by saying that ω' is *generated* from ω by τ in the sense that for all positions i there are variables $v \in \omega_i, v' \in \omega'_i$ with $v' = v^\tau$. Conditions (i) and (ii) together then mean that $\langle w, \omega\rangle$ has $\langle w', \omega'\rangle$ as counterpart iff there is some transformation τ such that $A^\tau \in w'$ whenever $\Box A \in w$ and ω' is generated from ω by this transformation. For example, let

$$w = \{\Box Fx_1, \Box Fx_2, \ldots \Diamond\exists x \neg Fx\}$$
$$w' = \{Fx_2, Fx_3, \ldots \exists x \neg Fx, \neg Fx_1$$

Here $\langle w, \langle [x_1]_w, [x_2]_2, \ldots\rangle\rangle$ has $\langle w', \langle [x_2]_{w'}, [x_3]_{w'}, \ldots\rangle\rangle$ as counterpart.

(Here is why we need sequences for this approach. Take two worlds w, w'. Suppose the only non-trivial box sentence in w is $\Box(Fx \leftrightarrow \neg Fy)$, and w' contains $Fz, \neg Fu$. Then we could map x to z and z to u, or x to u and y to z, but we can't let x and y have both z and u as counterparts.)

There is a difference between Tarskian sequences and sequences as relata of counterparthood: Tarskian sequences may contain repetitions (as when x and y corefer), but the relata of counterparthood plausibly can't. Otherwise we could have $\langle w, a, a, \ldots\rangle$ counterpart-related to $\langle w', b, c, \ldots\rangle$ (and to nothing else at w'), but what would that mean? It better not mean that qua referent of 'x', a has b as counterpart, and qua referent of 'y', c. Then we'd get a violation of $\vdash x = y \supset \Diamond Fx \supset \Diamond Fy$. Perhaps it's worth investigating this possibility – it would need even more restrictive substitution principles and in many ways look like concept semantics. But it's not my main object of study here.

So multiple counterparthood is not reflected by something like $\langle w, a, a, \ldots\rangle C \langle w', b, c, \ldots\rangle$, but rather by a pair like $\langle w, a, \ldots\rangle C \langle w', b, \ldots\rangle$ and $\langle w, a, \ldots\rangle C \langle w', c, \ldots\rangle$. On the other hand, when we look at the "imaging" relation between Tarskian sequences introduced by such a counterpart relation, we will allow $\langle w, a, a, \ldots\rangle$ to be linked with $\langle w, b, c, \ldots\rangle$, for we want to validate $x = y \wedge \Diamond x \neq y$. So how exactly is $\langle w, s\rangle \triangleright \langle w', s'\rangle$ defined? (Not in the CM, but in general?)

Now let's see what exactly we need for the truth lemma. In the end, we want to show that $w, V \Vdash A$ iff $A \in w$. The crucial case is for $A = \Box B$.

$w, V \Vdash \Box B$ iff $w', V' \Vdash B$ for all w', V' with wRw' and $V_w \triangleright V'_{w'}$. What does that mean? wRw' means that for some τ, $A^\tau \in w'$ whenever $\Box A \in w$. In this case, let's say that τ *links* w and w' (for short, $w \xrightarrow{\tau} w'$). $V_w \triangleright V'_{w'}$ means that $\langle w, \underline{V_w(x)}\rangle C \langle w', \underline{V'_{w'}(x)}\rangle$ – no! This, in turn, means that for some (possibly different) τ, (i) $w \xrightarrow{\tau} w'$ and (ii) there is a sequence of variables $\underline{z}$ such that $z_i \in V_w(x_i)$ and $z_i^\tau \in V'_{w'}(x_i)$. Putting all this together, wRw' and $V_w \triangleright V'_{w'}$ means that there is a τ such that (i) $w \xrightarrow{\tau} w'$ and (ii) for some $\underline{z}$, $z_i \in V_w(x)$ and $z_i^\tau \in V'_{w'}(x_i)$.

Now, $V_w \triangleright V'_{w'}$ does not mean $\langle w, \underline{V_w(x)}\rangle C \langle w', \underline{V'_{w'}(x)}\rangle$. But perhaps we can defined C in such a way that in the CM, $V_w \triangleright V'_{w'}$ still means that for some τ, $w \xrightarrow{\tau} w'$ and for some $\underline{z}$, $z_i \in s_i$ and $z_i^\tau \in s'_i$. The old definition of C said that $\langle w, d\rangle C \langle w', d'\rangle$ iff there is a $z \in d$ with $z^\tau \in d'$. The obvious generalisation to sequences is: $\langle w, d_1, d_2, \ldots\rangle C \langle w', d'_1, d'_2, \ldots\rangle$ iff there is a $\underline{z}$ such that $z_i \in d_i$ and $z_i^\tau \in d_i'$. We also no longer have a fixed τ, so we could just say that $\langle w, d_1, d_2, \dots \rangle C \langle w', d_1', d_2', \dots \rangle$ iff what we just said is true for some τ. Or we could include the $w \xrightarrow{w}{}'$ condition, making R redundant. (We didn't do that in the old construction: there a counterpart of $V_w(x)$ could violate all sorts of things w says about the modal profile of x, as long as the world was inaccessible.) Finally, $V_w \triangleright V'_{w'}$ iff whenever a sequence of distinct individuals $d_1, d_2, \dots$ is the V_w-value of a sequence of variables $x_1, x_2, \dots$, then the $V'_{w'}$-value of these variables is a counterpart sequence of $d_1, d_2, \dots$. Does that work? (BTW, now that we've distinguished counterpart sequences from Tarski sequences it is clear that counterpart sequences need not be infinite. This is also important here, because it might be that all the variables denote just two different individuals.)

It would be nice if the only $V'_{w'}$ with $V_w \triangleright V'_{w'}$ were $V^\tau_{w'}$ with $w \xrightarrow{\tau} w'$.

What we can show is this: if $w \xrightarrow{\tau} w'$, then $V_w \triangleright V^\tau_{w'}$. Again, this means that $\langle w, \underline{V_w(x)} \rangle C \langle w', \underline{V^\tau_{w'}(x)} \rangle$, which in turn means that there is a σ such that (i) $w \xrightarrow{\sigma} w'$ and (ii) for some $\underline{z}$, $z_i \in V_w(x_i) = [x_i]_w$ and $z_i^\tau \in V^\tau_{w'}(x_i) = [x_i^\tau]_{w'}$. Now choose τ for σ and $\underline{x}$ for $\underline{z}$.

If we have a suitable existence lemma, that also gives us one direction of the truth lemma: assume $w, V \Vdash \Box B$. Then $w', V' \Vdash B$ for all wRw', $V_w \triangleright V'_{w'}$. Then in particular $w', V^\tau \Vdash B$ for all w' with $w \xrightarrow{\tau} w'$. So $w', V \Vdash B^\tau$ by the transformation lemma, and $B^\tau \in w'$ by induction. Suppose for reductio that $\Box B \notin w$. Then $\Diamond\neg B \in w$. By the existence lemma, there is a w' and τ such that $w \xrightarrow{\tau} w'$ and $\neg B^\tau \in w'$. But we've just seen that $B^\tau \in w'$.

Can we show that the only $V'_{w'}$ with $V_w \triangleright V'_{w'}$ are $V^\tau_{w'}$ with $w \xrightarrow{\tau} w'$? Assume $V_w \triangleright V'_{w'}$. I.e. there is a τ with (i) $w \xrightarrow{\tau} w'$ and (ii) for some $\underline{z}$, $z_i \in V_w(x_i)$ and $z_i^\tau \in V'_{w'}(x_i)$. Can we show that $V'_{w'}(x_i) = V^\tau_{w'}(x_i) = [x_i^\tau]_{w'}$? No. Assume w contains $\{x = y, \Diamond x \neq y\}$. So for some $w \xrightarrow{\tau} w'$, w' contains $\{x^\tau \neq y^\tau\}$. Let $\underline{z} = \langle y, x, z, \dots \rangle$, and let $V'_{w'}(x_i) = [z_i^\tau]_{w'}$. Then $V'_{w'}(x) = [y^\tau]_{w'} \neq [x^\tau]_{w'}$. But $V_w \triangleright V'_{w'}$ via τ and $\underline{z}$. (In particular, $y = z_1 \in V_w(x_1) = \{x, y\}$ and $y^\tau = z_1^\tau \in V'_w(x_1) = \{y^\tau\}$.)

Still, it might be that there is another transformation σ such that $V_w \triangleright V'_{w'}$ via σ and $\underline{z}$ and $V'_{w'} = V^\sigma_{w'}$. Indeed, in the present example, σ can be defined by $\sigma(x_i) = \tau(z_i)$, mapping y to x^τ, x to y^τ, etc. Now we'd still need to verify that $w \xrightarrow{\sigma} w'$. But in any case, we can't always use $\underline{z}$ and τ to determine a new transformation $\sigma = \tau \circ \underline{z}$, because $\underline{z}$ may contain repetitions in which case σ won't be injective. So the question is, can we restrict the $\underline{z}$ in the definition of counterparthood to be pairwise distinct, so that $\underline{z}$ is in effect another transformation $[\underline{z}/\underline{x}]$? So: $\langle w, \underline{d} \rangle C \langle w', \underline{d'} \rangle$ iff (i) ... and (ii) there is some transformation ζ such that for all i, $x_i^\zeta \in d_i$ and $(x_i^\zeta)^\tau \in d_i'$. – Arguably this can't work, because quantifiers can easily generate a Tarskian sequence $\{x\}, \{x\}, \{x\}, \{x\}, \dots$ and then the z_i can't be distinct. The only way to ensure this would be if all variable classes were infinite.

So we have something like the situation in the old model, where $V^\tau_{w'}$ wasn't the only image of V_w.

10 Locally classical and two-dimensional logics

The semantics given above lends itself nicely to various 2D interpretations: the reference of an individual constant x varies in two ways from world to world; first, there is the variation between $V_w(x)$ and $V_{w'}(x)$, which tracks the differences in reference depending on the ("actual") world at which we begin the evaluation of sentences. (Here a world may of course be a centered world, i.e. a context.) Second, there is the variation between $V_w(x)$ and $V'_{w'}(x)$, where V' is a w'-image of V at w. This tracks differences (or lack thereof) in reference between worlds "considered as counterfactual", holding fixed w as the actual world.

There are many ways to flash this out. E.g. you might give an epistemic interpretation to V so that $V_w(x)$ captures what you would (in principle) identify as x on the supposition that w is actual (cf. Chalmers). Or you might give a straightforward contextual interpretation to V so that $V(x)$ only varies from world to world if x is a "demonstrative" like 'I' or 'that' (cf. Jackson). Or you might give a meta-semantic interpretation to V so that $V_w(x)$ is what x refers to in the language used in w (cf. Stalnaker).

If you dislike any of these options, or think they don't belong into semantics, don't worry. We can e.g. completely avoid the variability of V by including an actual world/context in the model.

> Definition 10.1 (Quantified modal logics)
> A *positive (negative) quantified modal logic* is a function that maps each standard language $\mathcal{L}$ of quantified modal logic to a subset $L \subseteq \mathcal{L}$ that contains P (N) and is closed under (MP), (UG), (Sub$*$) and (Nec).

> Definition 10.2 (Quantified modal logics with substitution)
> A *positive (negative) quantified modal logic with substitution* is a function that maps each standard language $\mathcal{L}_s$ of quantified modal logic to a subset $L \subseteq \mathcal{L}_s$ that contains P_s (N_s) and is closed under (MP), (UG), (Sub$_\mathrm{s}$) and (Nec).

I have been using free logic as a base logic in order to allow for worlds where things do not have counterparts. But we may distinguish models that are genuinely free from models that are "remotely free".

Consider a formula like $\Box Fx$. In modal logic, this can only be evaluated as true or false from a specific point w within a model. So assume we dive into some such point w. The box in $\Box Fx$ then instructs us to look at the points w' accessible from here and check if the embedded formula Fx is true at those points. In a simple semantics, the rules for evaluating formulas at other points are the same as the rules for evaluating formulas at the initial point w. But this need not be so. For example, in two-dimensional modal logic, the "actually" operator ACT behaves differently depending on whether the world of evaluation is the initial point w or a shifted point w'. Following 2D terminology, I will call the initial point at which a formula is evaluated a *world considered as actual*, and other points *worlds considered as counterfactual*. [Strictly, one can also consider the actual world as counterfactual.] An equivalent way of making the distinction is if we associate each model with a designated actual point w_{c}.

With this distinction, the reason for using free logic in QML are not entirely convincing. Why would we want to allow names for things that don't exist at the world considered as actual? – Not that there aren't any good answers to this question. Ordinary language, for instance, does have names for non-existent objects. So if we want to use QML to formalise statements from ordinary modal discourse, we may want a genuinely free logic. But there are many other uses of QML, in some of which we may well assume that there are no names for non-existent objects.

What is a remotely free logic? We somehow have to restrict necessitation. Kripke suggested restricting it to closed formulas. That would block the derivation of bad stuff, but it would also not give us the harmless $\Box(Fx \vee \neg Fx)$. Similarly for the idea that one should only move from $\vdash A(\underline{x})$ to $\vdash \Box(E\underline{x} \supset A(\underline{x}))$. Again, we couldn't prove $\Box(Fx \vee \neg Fx)$.

I used to play around with other restrictions on necessitation. E.g. a restriction to closed formulas, or the restricted form

if $\vdash_L E\underline{x} \supset A$, where $\underline{x} = \mathit{Var}(A)$, then $\vdash_L \Box(E\underline{x} \supset A)$.

This is indeed sound. (Assume $w, V \nVdash \Box(E\underline{x} \supset A)$ in some model $\mathcal{M}$. Then $w', V' \nVdash E\underline{x} \supset A$ for some $w' \in W$ with wRw' and V' some w'-image of V at w. Then $w', V' \Vdash E\underline{x}$ and $w', V' \nVdash A$. So V' is defined for all $x \in \underline{x}$, as otherwise $w', V' \nVdash E\underline{x}$. Let V^* be like V' except that V^* is total. Since V' and V^* agree on all free variables in A, $w', V^* \nVdash A$. Thus A is invalid in a model that is like $\mathcal{M}$ but with V^* in place of V.) But I think it's not enough. We want to prove $\Box(Fx \supset Fx)$ or $\Box(\neg\exists z(Fz \wedge \neg Fz) \vee Fx)$. We kinda want to put Ex in front of A only if A is really committed to x. But then A may be committed to x *or* y, without being committed to either.

Summing up, there are four main options.

1. locally and remotely negative (single-domain),
2. locally and remotely positive (dual-domain),
3. locally classical, remotely negative (single-domain),
4. locally classical, remotely positive (dual-domain).

The last two are extensions of the first two.

The minimal logics CP and CN aren't that hard to define: take the fully positive or negative modal logics P and N, considered as sets of sentences, add all instances of Ex, and close under (MP) and (UI). But what is an extension of e.g. CN? Do we add further axioms to CN and close under (MP) and (UI)? Or do we add them to NN, close under (MP), (UI) and (Nec), and then add Ex and close under (MP) and (UI)? Either way might be desired.

Standard predicate logic can be axiomatised by adding all instances of

(Ex) Ex

to either positive free logic or negative free logic. As mentioned in section xxx, there are reasons to have a logic that is *locally* classical in the sense that it contains all theorems of classical logic, without however containing all necessitations of all such theorems – in particular, without containing all instances of $\Box Ex$.

When we consider systems stronger than the minimal logics, we therefore do not require closure under necessitation. Rather, such systems may have a *global part* that is closed under Necessitation, and a *merely local part* that isn't. E.g. in two-dimensional modal logic, the merely local part might contain the axiom $p \leftrightarrow \mathrm{ACT}p$. In locally classical modal logics, the local part might contain the axiom Ex.

> Definition 10.3 (Quantified modal logics)
> A *positive quantified modal logic* is a function that maps each standard language $\mathcal{L}$ of quantified modal logic to a subset $L \subseteq \mathcal{L}$ that contains P and is closed under (MP), (UG) and (Sub∗).

Similarly, a *negative quantified modal logic* is a function that maps each standard language $\mathcal{L}$ to a subset $L \subseteq \mathcal{L}$ that contains N and is closed under (MP), (UG) and (Sub$*$).

The *global part* $\Box(L)$ of a quantified modal logic L is the set $\{A \in L : \Box A \in L$ and $\Box\Box A \in L$ and $\ldots\}$; the *merely local part* $@(L)$ of L is the set of L-members that are not in the global part.

Definition 10.4 (Minimal weakly negative quantified modal logic)
The *minimal weakly negative quantified modal logic* N^- maps each standard language $\mathcal{L}$ to the smallest set that contains all $\mathcal{L}$-instances of (Taut), (UD), (VQ), (FUI*), (LL*), (K), as well as

(Neg) $Px_1 \ldots x_n \supset Ex_1 \wedge \ldots \wedge Ex_n$,

($\forall$=R) $\forall x(x = x)$

and that is closed under (MP), (UG), (Nec) and (Sub*).

So N is standard negative free logic with the addition of the modal principles (K) and (Nec) and the principles for substitution restricted.

Definition 10.5 (Quantified modal logics with substitution)
A *positive quantified modal logic with substitution* is a function that maps each standard language $\mathcal{L}_s$ of quantified modal logic to a subset $L \subseteq \mathcal{L}_s$ that contains P_s and is closed under (MP), (UG) and (Sub$_\mathrm{s}$).

A *negative quantified modal logic with substitution* is a function that maps each standard language $\mathcal{L}_s$ to a subset $L \subseteq \mathcal{L}_s$ that contains N_s and is closed under (MP), (UG) and (Sub$_\mathrm{s}$).

Global and local parts are defined as before.

Definition 10.6 (Negative, positive, and locally classical models)
A counterpart model $\mathcal{M} = \langle \mathcal{S}, V \rangle$ for a language $\mathcal{L}$ on a structure $\mathcal{S} = \langle W, R, U, D, C \rangle$ is

positive if (i) $\mathcal{S}$ is total, and (ii) $V_w(x) \in U_w$ for all worlds $w \in W$ and variables $x \in \mathcal{L}$;

locally classical if $V_w(x) \in D_w$ for all worlds $w \in W$ and variables $x \in \mathcal{L}$.

Moreover, $\mathcal{M}$ is

a P***-model*** if $\mathcal{M}$ is positive,

a CP***-model*** if $\mathcal{M}$ is positive and locally classical,

an N***-model*** if $\mathcal{S}$ is single-domain,

a CN***-model*** if $\mathcal{S}$ is single-domain and locally classical.

Note that e.g. $[(Taut), (UD), (VQ), (UI)]_{\Box}$ does not only include boxed versions of all axioms, but also of all theorems. Otherwise it wouldn't be obvious that we get e.g. $\Box\forall x(Fx \supset Fx)$, or $\Box\forall x(K)$.

> THEOREM 10.7 (SOUNDNESS OF CP AND CP^*)
> Every member of CP and CP^* is valid in every CP-model.

PROOF A CP-model is a P-model with the further constraint that for all variables x and worlds w, $V_w(x) \in D_w$. Hence if something is valid in all P-models it is also valid in all CP-models. So by the previous theorem, every member of P is valid in all CP-models. Since the CP-models satisfy the condition on $\mathcal{X}$ from lemma xxx, it remains to show that (Ex) is valid.

TODO. ■

> THEOREM 10.8 (SOUNDNESS OF CN AND CN^*)
> Every member of CN and CN^* is valid in every CN-model.

PROOF The proof is essentially the same. It remains to show again that (Ex) is valid.

TODO. ■

From CM section:

Things are more complicated in weakly negative logics. Here we want all variables to denote something, if only a member of U_w. But since atomic predicates never apply to members of U_w, we can't define $V_w(x)$ as $\{y : x{=}y \in w\}$, since that would just be the empty set. We could set $V_w(x) = \{x\}$, but then we don't satisfy the redundancy condition that if x has two counterparts (i.e. if $\langle x : y\rangle\Diamond x{\neq}y \in w$), then x has two names. It wouldn't help just to introduce arbitrary new names for x. We would also need it to be the case that if x' is one such name, and $\Box Fx \in w$, then $\Box Fx' \in w$. But since $x{=}x' \notin w$, these constraints would have to be added separately.

The better option is probably to treat weakly negative models as special cases of positive models, as follows. Let '$\doteq$' be a new binary predicate, and let N_p be axiomatised by P + (Neg) +

$$\vdash x \doteq y \leftrightarrow x{=}y \land Ex \land Ey,$$

with (Neg) restricted to non-logical predicates (otherwise it clashes with (=R) in P). So '$\doteq$' is "inner identity" and '$=$' "outer identity". Let $\mathcal{L}'$ be the same language but without '$=$', and let N'_p be N_p restricted to $\mathcal{L}'$. Interpreting '$\doteq$' as identity, N'_p is N. (xxx show!) Now let Γ be any N'_p-consistent set of formulas. Then Γ is also N_p-consistent and therefore verified by some world in the canonical model $\mathcal{M}$ of N_p. So $\mathcal{M}$ can serve as a canonical model for N as well.

Logics with merely local parts. In a positive logic, worlds w as counterfactual must classify things as F and $\neg F$ even if those things don't exist at w. So we'll have both existent and non-existent counterparts. However, if worlds are maximal L-consistent sets and L is classical, then w can't contain $\neg Ex$. Considered as counterfactual, w must change from a classical world to a free world!

In general, if a logic L contains axiom A merely in its local part, we have to allow $\neg A$ at worlds as counterfactual, but not at worlds as actual. I.e., we will have $\neg A^\tau$, but not $\neg A$. However, A^τ is [often??] just a substitution instance [?? no, transformation instance?] of A, so $\vdash_L A$ yields $\vdash_L A^\tau$ as well.

A partial solution is to let τ map $\mathcal{L}$-variables to variables outside $\mathcal{L}$. A world w is a maximal, witnessed, $\Box(L)$-consistent set of formulas such that $w \cap \mathcal{L}$ is L-consistent. Thus in CP, a world can never say $\neg Ex$ for $\mathcal{L}$-variables x, but it can say $\neg Ex^\tau$. (It can of course also say Ex^τ. There is no requirement that the new variables denote non-existents.)

Let's pause to see why that move is OK. We want the canonical model $\mathcal{M}_L$ to be a model for L, otherwise completeness wrt $\mathcal{M}_L$ is rather uninteresting. E.g. to prove that L is characterised by model class $\mathcal{M}$, one would like to show that L is sound on $\mathcal{M}$ and that $\mathcal{M}_L \in \mathcal{M}$ so that L is also complete on $\mathcal{M}$. So L better be sound on $\mathcal{M}_L$. Strictly speaking, model classes and logics are language-relative (for models by way of V), so all this should be prefixed by some choice of a language $\mathcal{L}$. I.e., one would like to show that all members of $L(\mathcal{L})$ are valid on $\mathcal{M}$, and that $\mathcal{M}_{L(\mathcal{L})} \in \mathcal{M}_L$, and so that all members of $L(\mathcal{L})$ are valid in $\mathcal{M}_{L(\mathcal{L})}$. It doesn't matter if other formulas like Ex^τ, that are not members of $L(\mathcal{L})$, are not valid in $\mathcal{M}_{L(\mathcal{L})}$. The constraint that $w \cap \mathcal{L}$ is L-consistent is enough to ensure that w is a model for L in $\mathcal{L}$.

But the solution often doesn't work. Consider the possibility of empty worlds. It would be nice – albeit not essential – to use free logics P/N that allow for empty worlds. But if $\{\Diamond\neg\exists x Ex\}$ is L-consistent, then we need a world w in the CM with $w, V \Vdash \neg\exists x Ex$, and so we need a world w' with $w, V^\tau \Vdash \neg\exists x Ex$ and therefore $w', V \Vdash \neg\exists x Ex$ by transformation lemma and alpha-conversion. But then we can't have $\vdash_L \exists x Ex$!

One response might be to declare worlds like w' as merely counterfactual. We'd then show that every L-consistent set is included in a "proper" world w, and that for proper and improper worlds w, $A^{n+} \in w$ iff $w, V^{n+} \Vdash A$. Unfortunately, a model with improper worlds does not validate certain theorems of L, e.g. $\exists x Ex$. So the "canonical model" for L won't be a model for L at all. So that's no good.

There's something odd about logics L that rule out emptiness but where $\Box(L)$ doesn't rule out emptiness. A logic should apply from any point in a relational structure. It makes sense that something should exist here that is not C-related to anything over there. But if it makes sense to say that nothing at all exists over there, we shouldn't have a logic that can prove that something exists. So I might as well require that if $L \vdash \exists x Ex$, then $L \vdash \Box\exists x Ex$. I.e. I might as well require that $\exists x Ex \supset \Box\exists x Ex$ is L-provable. With this restriction, we can actually replicate the completeness technique for logics with Ex as a merely local axiom.

Henkin sets would then be defined with the requirement that H be

1. $\Box(L)$-*consistent*: there are no $A_1, \ldots, A_n \in H$ with $\vdash_{\Box(L)(\mathcal{L}^*)} \neg(A_1 \wedge \ldots \wedge A_n)$,
2. L-*consistent wrt* $\mathcal{L}$: there are no $A_1, \ldots, A_n \in H$ with $\vdash_{L(\mathcal{L})} \neg(A_1 \wedge \ldots \wedge A_n)$, (equivalently, there are no $A_1, \ldots, A_n \in H \cap \mathcal{L}$ with $\vdash_{L(\mathcal{L}^*)} \neg(A_1 \wedge \ldots \wedge A_n)$),

instead of L-consistent.

The extensibility lemma becomes:

Lemma 10.9 (Extensibility Lemma)
If Γ is an L-consistent set of $\mathcal{L}_s^*$-sentences whose $\mathcal{L}_s$-fragment is L-consistent, and if infinitely many $\mathcal{L}_s^*$-variables do not occur in Γ, then there is a Henkin set $H \in \mathcal{H}_L$ such that $\Gamma \subseteq H$.

Proof Let $S_1, V_2, \ldots$ be an enumeration of all $\mathcal{L}_s^*$-sentences, and $z_1, z_2, \ldots$ an enumeration of the unused $\mathcal{L}_s^*$-variables such that $z_i \notin \mathit{Var}(S_1 \wedge \ldots \wedge S_i)$. Let $\Gamma_0 = \Gamma$, and define Γ_n for $n \geq 1$ as follows.

(i) If $\Gamma_{n-1} \cup \{S_n\}$ is not L-consistent, then $\Gamma_n = \Gamma_{n-1}$;

(ii) else if S_n is an existential formula $\exists x A$ and $\mathcal{L}_s$ contains substitution, then $\Gamma_n = \Gamma_{n-1} \cup \{\exists x A, \langle z_n : x \rangle A, E z_n\}$;

(ii*) else if S_n is an existential formula $\exists x A$ and $\mathcal{L}_s$ doesn't contain substitution, then $\Gamma_n = \Gamma_{n-1} \cup \{\exists x A, [z_n/x]A, E z_n\}$;

(iii) else if S_n is a substitution formula $\langle y : x \rangle A$, then $\Gamma_n = \Gamma_{n-1} \cup \{y{=}y \supset y{=}z_n\}$;

(iv) else $\Gamma_n = \Gamma_{n-1} \cup \{S_n\}$.

Define w as the union of all Γ_n. We show that w is a Henkin set for L.

1. w is L-consistent. This is shown by proving that Γ_0 is L-consistent and that whenever Γ_{n-1} is L-consistent, then so is Γ_n. It follows that no finite subset of w is L-inconsistent, and hence that w itself is L-consistent.

   a) Γ_0 is L-consistent. This is given by assumption.
   b) If Γ_{n-1} is L-consistent and case (i) in the construction applied, then trivially Γ_n is L-consistent.
   c) Assume case (ii) in the construction applied, and suppose that $\Gamma_n = \Gamma_{n-1} \cup \{\exists x A, \langle z_n : x \rangle A, E z\}$ is L-inconsistent. Then there is a finite subset $\{C_1, \ldots, C_m\} \subseteq \Gamma_{n-1}$ such that

$$\vdash_L \neg(C_1 \wedge \ldots \wedge C_m \wedge \exists x A \wedge \langle z_n : x \rangle A \wedge E z_n).$$

   Thus by (Taut) and (MP),[7]

$$\vdash_L C_1 \wedge \ldots \wedge C_m \wedge \exists x A \supset (E z_n \supset \neg \langle z_n : x \rangle A).$$

   By (UG) and (UD),[8]

$$\vdash_L \forall z_n (C_1 \wedge \ldots \wedge C_m \wedge \exists x A) \supset \forall z_n (E z_n \supset \neg \langle z_n : x \rangle A).$$

7 Need: (Taut), (MP).
8 Need: (UG), (UD).

Since z_n does not occur in Γ, (VQ) yields $\vdash_L C_1 \wedge \ldots \wedge C_m \wedge \exists x A \supset \forall z_n(C_1 \wedge \ldots \wedge C_m \wedge \exists x A)$.[9] So by (Taut) and (MP),

$$\vdash_L C_1 \wedge \ldots \wedge C_m \wedge \exists x A \supset \forall z_n(Ez_n \supset \neg\langle z_n : x\rangle A).$$

But by fiat $\vdash_L \forall z_n(Ez_n \supset \neg\langle z_n : x\rangle A) \supset \forall x \neg A$.[10] So by (Taut) and (MP),

$$\vdash_L C_1 \wedge \ldots \wedge C_m \wedge \exists x A \supset \neg\exists x A.$$

So $\{C_1, \ldots C_m, \exists x A\}$ is not L-consistent, contradicting the assumption that clause (ii) applies and so $\Gamma_n \cup \{\exists x A\}$ is L-consistent.

d) Assume case (ii*) in the construction applied, and suppose that $\Gamma_n = \Gamma_{n-1} \cup \{\exists x A, [z_n/x]A, Ez\}$ is L-inconsistent. Then there is a finite subset $\{C_1, \ldots, C_m\} \subseteq \Gamma_{n-1}$ such that

$$\vdash_L \neg(C_1 \wedge \ldots \wedge C_m \wedge \exists x A \wedge [z_n/x]A \wedge Ez_n).$$

Thus by (Taut) and (MP),[11]

$$\vdash_L C_1 \wedge \ldots \wedge C_m \wedge \exists x A \supset (Ez_n \supset \neg[z_n/x]A).$$

By (UG) and (UD),[12]

$$\vdash_L \forall z_n(C_1 \wedge \ldots \wedge C_m \wedge \exists x A) \supset \forall z_n(Ez_n \supset \neg[z_n/x]A).$$

Since z_n does not occur in Γ, (VQ) yields $\vdash_L C_1 \wedge \ldots \wedge C_m \wedge \exists x A \supset \forall z_n(C_1 \wedge \ldots \wedge C_m \wedge \exists x A)$.[13] So by (Taut) and (MP),

$$\vdash_L C_1 \wedge \ldots \wedge C_m \wedge \exists x A \supset \forall z_n(Ez_n \supset \neg[z_n/x]A).$$

But by fiat $\vdash_L \forall z_n Ez_n$.[14] So by (Taut) and (UD), $\vdash \forall z_n(Ez_n \supset \neg[z_n/x]A) \supset \forall z_n \neg[z_n/x]A$. So by syntactic change of bound variables, $\vdash \forall z_n(Ez_n \supset \neg[z_n/x]A) \supset \forall x \neg A$.[15] So by (Taut) and (MP),

$$\vdash_L C_1 \wedge \ldots \wedge C_m \wedge \exists x A \supset \neg\exists x A.$$

So $\{C_1, \ldots C_m, \exists x A\}$ is not L-consistent, contradicting the assumption that clause (ii) applies and so $\Gamma_n \cup \{\exists x A\}$ is L-consistent.

e) Assume case (iii) in the construction applied, and suppose that $\Gamma_n = \Gamma_{n-1} \cup \{\langle y : x\rangle A, y = y \supset y = z_n\}$ is L-inconsistent, although $\Gamma_{n-1} \cup \{\langle y : x\rangle A\}$ is L-consistent. Then there are $\underline{C} \in \Gamma_{n-1}$ such that

$$\vdash_L \underline{C} \wedge \langle y : x\rangle A \supset y = y \wedge y \neq z_n.$$

9 Need: (VQ).
10 Need: $\vdash_L \forall z_n(Ez_n \supset \neg\langle z_n : x\rangle A) \supset \forall x \neg A$, provided z_n is new.
11 Need: (Taut), (MP).
12 Need: (UG), (UD).
13 Need: (VQ).
14 Need: $\vdash_L \forall z_n Ez_n$.
15 Need: $\vdash \forall z_n \neg[z_n/x]A \supset \forall x A$, provided z_n is new.

But since L is closed under substitution, this means that

$$\vdash_L \langle y : z_n \rangle(\underline{C} \wedge \langle y : x \rangle A \supset y = y \wedge y \neq z_n).$$

So by Subs-distribution,[16]

$$\vdash_L \langle y : z_n \rangle(\underline{C} \wedge \langle y : x \rangle A) \supset \langle y : z_n \rangle y = y \wedge \langle y : z_n \rangle y \neq z_n.$$

Since z_n does not occur in $\Gamma_{n-1}, \langle y : x \rangle A$,[17] and since $\vdash \langle y : z_n \rangle y \neq z_n \supset y \neq y$[18],

$$\vdash_L \underline{C} \wedge \langle y : x \rangle A \supset (y = y \wedge y \neq y).$$

But then $\{\underline{C}, \langle y : x \rangle A\}$ is L-inconsistent.

Old.

By (Taut) and (MP),

$$\Gamma_{n-1}, \langle y : x \rangle A \vdash_L y = z_n \supset \neg[z_n/x]A.$$

Now by (LL),[19]

$$\vdash_L y = z_n \supset (\langle y : x \rangle A \supset \langle z_n : y \rangle \langle y : x \rangle A).$$

And by fiat, $\vdash_L \langle z_n : y \rangle \langle y : x \rangle A \supset \langle z_n : x \rangle A$.[20] Moreover, since z_n does not occur in x, A, $\vdash_L \langle z : x \rangle A \supset [z_n/x]A$.[21] So by (Taut) and (MP),

$$\Gamma_{n-1}, \langle y : x \rangle A \vdash_L y = z_n \supset [z_n/x]A.$$

So

$$\Gamma_{n-1}, \langle y : x \rangle A \vdash_L \neg y = z_n.$$

But since z_n does not occur in $\Gamma_{n-1}, \langle y : x \rangle A$ and L is closed under classical substitution, this means that

$$\Gamma_{n-1}, \langle y : x \rangle A \vdash_L \neg z_n = z_n,$$

and that for all z,

$$\Gamma_{n-1}, \langle y : x \rangle A \vdash_L \neg z = z,$$

and by (UG) and (VQ) that

$$\Gamma_{n-1}, \langle y : x \rangle A \vdash_L \forall z(\neg z = z).$$

Some of these should be impossible xxx.

16 Need: Subs-distribution $\vdash \langle y : x \rangle(A \supset B) \supset \langle y : x \rangle A \supset \langle y : x \rangle B$.
17 Need: if x does not occur in A, then $\vdash A \supset \langle y : x \rangle A$.
18 Need: if x does not occur in A, then $\vdash A \supset \langle y : x \rangle A$.
19 Need: (LL).
20 Need: $\vdash_L \langle z : y \rangle \langle y : x \rangle A \supset \langle z : x \rangle A$.
21 Need: if z_n does not occur in x, A, $\vdash_L \langle z : x \rangle A \supset [z_n/x]A$.

f) Assume case (iv) in the construction applied. Then $\Gamma_n = \Gamma_{n-1} \cup \{S_n\}$ is L-consistent, since otherwise case (i) would have applied.

2. $w \cap \mathcal{L}_s$ is L-consistent. Suppose $w \cap \mathcal{L}_s \vdash \bot$. Then $w \cap \mathcal{L}_s, \{Ex : x \in \mathcal{L}_s^*\} \vdash_L \bot$. But the $x \notin \mathcal{L}_s$ can't matter [explain!]. So $w \cap \mathcal{L}_s, \{Ex : x \in \mathcal{L}_s\} \vdash_L \bot$. But $\{Ex : x \in \mathcal{L}_s\} \subseteq w$. So we'd have $w \vdash_L \bot$, contradicting the fact that w is L-consistent.

3. w is maximal. Assume some formula S_n is not in w. Then $\Gamma_{n-1} \cup \{S_n\}$ is not L-consistent. So there are $C_1, \ldots, C_m \in \Gamma_{n-1}$ such that (by (Taut) and (MP)) $\vdash_L C_1 \land \ldots C_m \supset \neg S_n$. Similarly, if $S_k = \neg S_n$ is not in w, then there are $D_1, \ldots, D_l \in \Gamma_{k-1}$ such that $\vdash_L D_1 \land \ldots D_l \supset \neg S_k$. By (Taut) and (MP), it follows that there are $C_1, \ldots, C_m, D_1, \ldots D_l \in w$ such that

$$\vdash_L C_1 \land \ldots \land C_m \land D_1 \land \ldots \land D_l \supset (\neg S_n \land \neg\neg S_n).$$

   But then w is inconsistent, contradicting what was shown above.

4. w is witnessed. This is guaranteed by clause (ii) of the construction.

5. w is substitutionally witnessed. This is guaranteed by clause (iii) and the fact that the $z_n \notin Var(S_n)$. ■

The only other part that needs adjustment is the existence lemma, which says that for any world w in a canonical model $\mathcal{M}_L$ with $\Diamond A \in w$ there is an accessible world w' in the model with $A^\tau \in w'$. As before, let $\Gamma = \{A^\tau\} \cup \{B^\tau : \Box B \in w\}$. If $L = [[\Lambda]_\Box, Ex]$, then Γ may not be L-consistent, since w might contain $\Diamond\neg Ex$ or $\Box\neg Ex$, although CN can prove Ex (and Ex^τ). Then let $\Gamma' = \Gamma \cup \{Ex : x \in Var(\mathcal{L})\}$. Suppose Γ' is $\Box(L)$-inconsistent. Then $\vdash_{\Box(L)} \underline{C} \supset \neg(Ex_1 \land \ldots \land Ex_n)$ for some $\underline{C} \in \Gamma$ and $x_1 \ldots x_n \in Var(\mathcal{L})$. But Γ contains none of $x_1 \ldots x_n$, and so $\vdash_{\Box(L)} \underline{C} \supset \neg(Ex_1 \land \ldots \land Ex_n)$ only if $\vdash_{\Box(L)} \underline{C} \supset \neg(Ex_1 \land \ldots \land Ex_1)$, by (Sub) [check], and thus only if $\vdash_{\Box(L)} \underline{C} \supset \neg Ez$, for any variable z, and also $\vdash_{\Box(L)} \underline{C} \supset \forall z \neg Ez$. But I have stipulated that if $L = [[\Box(L)]_\Box, Ex]$, then $\vdash_{\Box(L)} \exists x Ex$.

But the solution still generalise. Consider a 2D logic with $p \leftrightarrow \text{ACT}p$ as a local axiom. We want $w, V \Vdash \Diamond(p \land \neg\text{ACT}p)$, so we would need a world w' with $w', V \Vdash p \land \neg\text{ACT}p$ – which is precisely what we don't want! In a proper 2D logic, imaging must do more than just re-interpret variables. When looking at worlds as counterfactual, we must also reconsider the interpretation of ACTp. That's why 2D logics require double-indexing: there's no sensible completion of $w, V \Vdash \text{ACT}p$ iff ... (unless we have a fixed actual world and thus only ever consider worlds as counterfactual!).

xxx traditional 2D logics with "actually" are non-trivial: they assume that "actually" works as an undo-operator, not as a shifter to the actual world. (Alternatively, we could treat "actually" as a shifter, but then we'd have to say how formulas are evaluated at a specific world as counterfactual. See hybrid logics.) Or can we just say that certain traditional principles involving "actually" presuppose functionality, and are thus valid on functional frames only?

11 Roots and comparisons

11.1 Ghilardi: Functor Semantics

(The following is mostly drawn from [Skvortsov and Shehtman 1993] and [van Benthem 1993]. The latter gives an accessible overview of 'functional' and 'categorical' alternatives to standard Kripke semantics, proposed by Ohlbach and Ghilardi [1991]. [Skvortsov and Shehtman 1993] is much harder, but still easier than Ghilardi's own presentation.) Following [van Benthem 1993], I will translate Ghilardi's account from category theory into set theory.

A *functional frame* (or $\mathcal{C}$*-set*) is a family of domains $\mathcal{D} = \{D_w : w \in W\}$ together with a family $\mathcal{F} = \{f_\mu : D_w \xrightarrow{\mu} D_{w'}\}$ of maps between such domains, indexed by morphisms between objects in $\mathcal{D}$.

An *interpretation* V is a function that assigns an extension to predicates and function symbols relative to points $w \in W$. In the recursive definition of truth, the clause for the box says that $\Box\Phi$ is true at w under assignment σ, for short $\mathcal{M}, w \Vdash \Box\Phi[\sigma]$, iff for all maps f from D_w to $D_{w'}$, $\mathcal{M}, w' \Vdash \Phi[f \circ \sigma]$.

Van Benthem mentions a few correspondence results. For example, $\Box Ax \supset Ax$ is valid iff for each $w \in W$ and $d \in D_w$ there is a map $f : D_w \to D_w$ with $f(d) = d$. Similarly, $\Box Ax \supset \Box\Box Ax$ is valid iff for all $d \in D_w$ and maps $f : D_w \to D_{w'}, g : D_{w'} \to D_{w''}$ there is a map $h : D_w \to D_{w''}$ with $g(f(d)) = h(d)$. These are *local* conditions insofar as different maps f or h may be chosen for different individuals. The stronger global conditions cannot be expressed in the language, but they could be enforced by stipulating that the mappings in frame must be closed under 'patching', to the effect that all consistent unions of functions in $\mathcal{F}$ must themselves be in $\mathcal{F}$. The Converse Barcan Formula is valid in all frames as long as the functions in $\mathcal{F}$ are total. The Barcan Formula defines a kind of surjectivity: for all $d \in D_w, d' \in D_{w'}$, if $w \xrightarrow{f} w'$ then there is a $g : D_w \to D_{w'}$ with $g(d) = d'$.

Ghilardi notes an interesting connection between functional models and standard propositional models: Given a functional model $\mathcal{M}$, let $F(\mathcal{M}) = \langle W, R\rangle$ be the propositional Kripke frame whose points W are mappings in $\mathcal{M}$, ordered by the relation R defined by Rfg iff there is a map h such that $g = h \circ f$. It turns out that $F(\mathcal{M}) \Vdash \Phi$ for a propositional formula Φ iff $\mathcal{M} \Vdash \Phi'$ for every predicate logical substitution instance Φ' of Φ. Ghilardi uses this fact to show that in the standard semantics, all modal predicate logics between S4.3 and S5 are incomplete.

Our counterpart models are obvious generalizations of Ghilardi's models. Instead of a family of functions we have a family of relations linking individual domains. This does complicate the semantics because $[R \circ \sigma]$ is a *set* of assignment functions, rather than a single function.

11.2 Metaframes: Skvortsov and Shehtman

11.3 Kutz 2000 and Kracht and Kutz 2002

[Kutz 2000] describes a model theory similar to mine. The main ideas and results are summarized in [Kracht and Kutz 2002].

1. Kutz's language excludes individual constants. He presents this as a limitation that is announced to be lifted at some other point [35] ([2002: 3]).

2. Non-modal (positive) free logic is axiomatised by (Taut), (VQ), (UD), (∀Ex), (=R),

   (FUI$^{\mathrm{K}}$) $\forall x \Phi(x, \bar{z}) \supset (Ey \supset \Phi(y, \bar{z})$, provided y is free for x in $\Phi(x, \bar{z})$ and $y \notin \bar{z}$;

   (LL$^{\mathrm{K}-}$) $x = y \supset (\Phi \supset \Phi(y//x))$, provided x is free in Φ and y is free for x in $\Phi(x)$.

   with the rules (MP) and (UG). [39] The requirement that $y \notin \bar{z}$ means that (unless y is x) y is not free in Φ at all.

   When free logic is merged with the basic modal logic K into FK ("Free K"), (LL$^{\mathrm{K}-}$) is revised as follows:

   (LL$^{\mathrm{K}}$) $x = y \supset (\Phi \supset \Phi(y//x))$, provided x is free in Φ, y is free for x in Φ, y does not occur freely in the scope of a modal operator in Φ, and $\Phi(y//x)$ is Φ with some free occurrences of x replaced by x such that in the scope of a modal operator, either all or no occurrences of x are replaced. [43]

   Arbitrary modal predicate logics are defined as sets $\mathsf{FK} \subseteq L \subseteq \mathcal{L}$ closed under (Nec), (UG), (MP) and closed under first and second-order substitution,

   (Sub$^{\mathrm{K}}$) if $\vdash_L \Phi(x, \bar{z})$, then $\vdash_L \Phi(y, \bar{z})$, provided x is free in Φ and y free for x;

   (Sub2$^{\mathrm{K}}$) if $\vdash_L \Phi(P(\bar{y}))$, then $\vdash_L \Phi(\Psi(\bar{x}))$,

   where the latter "may require some restrictions" e.g. to avoid turning $x = y \supset Gxx \supset Gxy$ into $x = y \supset \Box(x = x) \supset \Box(x = y)$ [43f.]. In [2002: 6f.], first-order closed modal predicate logics are explicitly distinguished from second-order closed logics, and the latter is not generally assumed.

   Note that (FUI$^{\mathrm{K}}$) and (LL$^{\mathrm{K}}$) are too restrictive (too weak), while (Sub$^{\mathrm{K}}$) is not restrictive enough (too strong). This last fact is not noticed because the soundness proof for FK doesn't consider (Sub$^{\mathrm{K}}$).

3. A *modal structure* is defined as a pair $\langle \mathcal{W}, \mathcal{C} \rangle$, where $\mathcal{W}$ is a set of free logic models $\langle U, D, I \rangle$ and $\mathcal{C}$ is a set of relations between pairs U_1, U_2 of individual domains in $\mathcal{W}$ such that all things in U_2 are related to something in U_2. (This is analogous to the common requirement of increasing domains.) Accessibility is defined so that wRw' iff at least one member C of $\mathcal{C}$ links w and w'; in this case it is said that w "sees" w' via C, for short: $w \xrightarrow{C} w'$. [46f.]

   Observe that a structure, for Kutz, includes an interpretation function for each world (but not an assignment function). Structures without interpretations are called *frames*.

   The fact that structures include a *set* of counterpart relations, each defined only for a given pair of domains, means that e.g. x at w can be its own counterpart at w, but not at w', because $\mathcal{C}$ contains the identity relation on U_w as well as some other relation between U_w and $U_{w'}$. (Kutz explicitly allows for numerical identity across worlds [37, fn.47].) OTOH, if $U_{w'} = U_{w''}$, then it looks like individuals at w cannot have different counterparts at w' than at w''. My models allow for this by having the relations in $\mathcal{C}$ tagged by the relevant structures (e.g., w and w'). It might be worth checking if this creates problems in Kutz's completeness proofs.

   There can be several counterpart relations between two given structures. As [Kracht and Kutz 2002: 11] remark, "this feature is not eliminable and actually one of the basic ingredients of all generalizations of standard Kripke-semantics. This is underlined by the fact that one can easily construct second-order closed modal predicate logics that are frame-complete only with respect to frames having at least two counterpart-relations between any two worlds".

4. A *modal model* is defined as a pair of a structure and an assignment function mapping each variable at each world to a member of the domain of that world. (So variables are local and non-rigid.) [47]

5. Satisfaction relative to a structure, an assignment function and a world is defined in the obvious way. The clause for the box is

   $S, w, v \Vdash \Box A(\bar{y})$ iff $S, w', v' \Vdash A$ for all w', v' such that there is a C with $w \xrightarrow{C} w'$ and v' is a $\bar{y}$-variant of v with $C(v_w(y_i), v'_w(y_i))$. [48]

   So $\Box A$ requires that A is true at all accessible worlds w' under all w'-images of v, via all counterpart relations linking the two worlds. (To make real use of the different counterpart relations, one could subscript the box by the relevant (type of) relation.)

   Kutz's rule is effectively the same as mine, and determines a uniform interpretation of variables in the scope of modal formulas, so that $x = x \land \Diamond x \neq x$ is unsatisfiable (see [37, fn.46]).

6. In the construction of canonical models for arbitrary logics L, worlds ("Freie Henkin Typen") are defined as sets H of formulas which are maximal, consistent, and witnessed (i.e. whenever H contains $\exists x \Phi(x) \in H$, then it contains both $\Phi(y)$ and Ey for some y free for x in $\Phi(x)$.) [57] The language is not extended by new variables.

7. For transformations τ (called "faithful substitutions"), Φ^τ is defined as Φ with all variables, including bound ones, replaced according to τ. [57]

   The extensibility lemma is then expressed as follows: if Δ is consistent and $\bar{y}, \bar{z}$ are two lists of m pairwise distinct variables, then there is a transformation τ with $\tau(y_i) = z_i$ and a free Henkin set H such that $\Delta^\tau \subseteq H$. [60] The main use of τ is to make sure that infinitely many variables do not occur in Δ^τ, which can then be used as witnesses. (In fact, τ is fixed so that for $x_i \notin \bar{y}$, $\tau(x_i) = x_{k+2i+i}$, where k is a sufficiently large index.) Our extensibility lemma, that Δ itself can be extended to a Henkin set provided infinitely many variables do not occur in Δ, is stated as corollary 7.1. [64]

8. To define the canonical model, each Henkin set H must be turned into a structure $\langle U_H, D_H, I_H \rangle$, in the obvious way: $U_H = \{[x]_H : x \in Var\}$, $D_H = \{[x]_H : Ex \in H\}$, and $I_H(P) = \{[\bar{x}]_H : P\bar{x} \in H\}$. [66f.] Moreover, the canonical assignment function obviously maps each x at w to $[x]_w$. [67f.]

   The counterpart relations for a canonical model are defined in two steps. First, for any Henkin sets w, w' and transformation τ, let C_τ be the relation on $U_w \times U_{w'}$ such that $C_\tau([x]_w, [y]_{w'})$ iff $\tau(x) = y$. Relations of this kind are called *pseudo-canonical counterpart relations* for w and w'. [68] Some of these relations are functional or injective (e.g. the ones induced by the identity transformation), others are not. [69f.]

   Not all pseudo-canonical counterpart relations C_τ between w and w' are actual links $w \xrightarrow{C_\tau} w'$ in the canonical model. To qualify as such a link, w' must contain Φ^τ whenever

w contains $\Box\Phi$. [70] Finally, the class $\mathcal{C}$ of canonical counterpart relations is defined as $\{C_\tau : \exists w \exists w', w \xrightarrow{C_\tau} w'\}$.

Here the worry arises that $\mathcal{C}$ might contain a counterpart relation C such that $w \xrightarrow{C_\tau} w'$ although w' does *not* contain Φ^τ whenever w contains $\Box\Phi$, where C_τ got into $\mathcal{C}$ by linking some other worlds v, v' with the same individuals.

9. Kutz shows that whenever a Henkin set w in a canonical model contains $\Diamond\Phi$, then there is a canonical counterpart relation C_τ such that some other set w' in the model contains Φ^τ. [71f.]

10. The proof of the truth lemma ("Fundamentallemma", or "Fundamental Theorem"), however, is invalid.

Here is part (vi) of the induction ((ii) in [2002: 14]), for formulas of type $\Diamond A(\bar{x})$. Assume that $w, V \Vdash \Diamond A(\bar{x})$. Then there is a set w' and an assignment V' such that $S', V' \Vdash A(\bar{x})$, where $V'(x_i)$ is a counterpart of $V(x_i)$, i.e. there is a counterpart substitution τ and variables u_i, v_i such that $u_i \in V(x_i)$, $v_i \in V'(x_i)$ and $\tau(u_i) = v_i$. In the following argument, it is implicitly assumed that the $\bar{u}$ and $\bar{v}$ are pairwise distinct. But that is not guaranteed. Suppose the individual $\{x_1, x_2\}$ at w has multiple τ-counterparts $\{v_1\}, \{v_2\}$ at w', with $\tau(x_i) = v_i$. Then $w, V \Vdash \Diamond(x_1 \neq x_2)$ as well as $w, V \Vdash \Diamond(x_1 = x_2)$. For the latter, one relevant V' is such that $V'_{w'}(x_1) = V'_{w'}(x_2) = \{v_1\}$. In this case, $u_1 = u_2 = x_1$. Hence there is no transformation h (or g in [2002: 14]) with $h(x_i) = u_i$, and one cannot infer via Lemma 7.11 that

$$\Diamond A(\bar{x}) \in w \text{ iff } \Diamond A^\tau(\bar{u}) \in w.$$

In the example, that would be

$$\Diamond(x_1 = x_2) \in w \text{ iff } \Diamond(x_1 = x_1) \in w.$$

What is needed here is a strengthening of (LL$^{\mathrm{K}}$) so that

$$\vdash x_1 = x_2 \supset \Diamond(x_1 = x_1) \supset \Diamond(x_1 = x_2).$$

(The coincidence lemma 7.3 is not my coincidence lemma; rather it is a lemma about substitution: it says that if $\bar{z}$ are pairwise distinct and free for $\bar{y}$ in $\Phi(\bar{y})$, and $v(y_i) = v^*(z_i)$, then $w, v \Vdash \Phi(\bar{y})$ iff $w, v^* \Vdash \Phi(\bar{z})$. It is crucial that the $\bar{z}$ are pairwise distinct. For instance $w, v \Vdash \Diamond x \neq y$ does not entail $w, v' \Diamond y \neq y$ even if $v(x) = v'(y)$. (Here $\bar{y} = \langle x, y \rangle$ and $\bar{z} = \langle y, y \rangle$.) Hence the step from 3 to 4 is only valid if $z \notin \bar{y}$. For example, if $\exists x \Phi(x, \bar{y})$ is $\exists x \Diamond x \neq y$, and the only member of the domain is $[y]_w = \{x, y, \ldots\}$, then it is true that $w, v^{x \mapsto [y]_w} \Vdash \Diamond x \neq y$, but we can't infer that $w, v \Vdash \Diamond y \neq y$.)

In part (v) there is a similar problem. Consider the formula $\exists x \Box x = y$. The argument goes as follows.

$$\begin{aligned}
& w, V \Vdash \exists\Box(x{=}y) \\
\Leftrightarrow\quad & w, V' \Vdash \Box(x{=}y) \text{ for some exist. } x\text{-var. } V' \text{ of } V \\
\Leftrightarrow\quad & w, V^{x\mapsto[z]} \Vdash \Box(x{=}y) \text{ for some } [z] \in D_w \\
\Leftrightarrow\quad & w, V \Vdash \Box(z{=}y) \text{ for some } [z] \in D_w \text{ (Lemma 7.3)} \\
\Leftrightarrow\quad & Ez \in w \text{ and } \Box(z{=}y) \in w \text{ for some } [z] \in D_w \text{ (induction)} \\
\Rightarrow\quad & \exists z\Box(z{=}y) \in w \text{ by (FUI}^{\mathrm{K}}) \\
\Leftrightarrow\quad & \exists x\Box(x{=}y) \in w \text{ by Lemma 7.11.}
\end{aligned}$$

This assumes that the variable z is distinct from y. Otherwise neither Lemma 7.3 nor (FUI$^{\mathrm{K}}$) nor Lemma 7.11 applies. But z and y may well coincide. For instance, assume $V_w(y) = \{y\}$. Then $V_w(y)$ has a unique counterpart at every accessible w' and we have $w, V \Vdash \exists x\Box(x = y)$. The relevant individual $[z] \in D_w$ is $\{y\}$. In the third step, we'd then need e.g.

$$w, V^{x\mapsto[y]} \Vdash \Box(x = y) \Leftrightarrow w, V \Vdash \Box(y = y)$$

which is not an application of Lemma 7.3 and also not generally valid. What's actually needed here is a strengthening of (FUI$^{\mathrm{K}}$) so that e.g.

$$\vdash \forall x\Box(Fxy \supset Ey \supset \Box Fyy)$$

Or, to stay in the example: if $V(y) = \{y\} \in D_w$, then $w, V \Vdash \exists x\Box(x = y)$; i.e., if w contains $\forall x\Diamond(x \neq y)$ and Ey, then $V(y)$ must not be $\{y\}$. This should be verified by the fact that $\exists v(v{=}y \wedge \Diamond(v{\neq}\, y)) \in w$. But this isn't derivable from $\forall x\Diamond(x{\neq}y)$ by (FUI$^{\mathrm{K}}$).

But the problem isn't just that the logic FK is too weak. Consider the case where w contains $x{=}y$, $\Diamond x{\neq}y$ and $\Box\Diamond x{\neq}y$. So $[x]_w = \{x, y, \ldots\}$ and there is a τ such that for some accessible w', $[x^\tau]_{w'} \neq [y^\tau]_{w'}$; in addition, w' contains $\Diamond x^\tau{\neq}y^\tau$ (for every τ). (Note that both $[x^\tau]_{w'}$ and $[y^\tau]_{w'}$ are τ-counterparts of $[x]_w$, since for each of them, there is some member of $[x]_w$ that is τ-mapped to one of their members.) To verify that $w, V \Vdash \Box\Diamond x{\neq}y$, we need to ensure that $w', V' \Vdash \Diamond x{\neq}y$ for all V'. Consider the V' that assigns $[y^\tau]_{w'}$ to both x and y. Now $w', V' \Vdash \Diamond x{\neq}y$ requires that $[y^\tau]_{w'}$ has two counterparts at some world w accessible from w – and we have no guarantee for that.

11. The canonical model lemma entails that every *categorical* logic (i.e. every logic that is valid on the frame of its canonical model) is frame-complete. Kutz gives some examples of logics and corresponding classes of frames. For example, he shows that the canonical frame of $FK + NI$ is *functional* in the sense that all $C \in \mathcal{C}$ are functional, and thus that $FK + NI$ is frame-complete wrt. the class of functional frames. (It is easy to show that NI is also valid on such frames.) [78f.] Similarly, $FK + NNI$ is frame-complete wrt. the class of *injective* frames where all $C \in \mathcal{C}$ are injective. [79] Next, $FK + CBF$ and $FK + NE$ are shown to be frame-complete wrt. the class of *existentially faithful* frames such that whenever $a \in D_w$, $w \xrightarrow{C} w'$ and aCb, then $b \in D_{w'}$. [79f.]

    A frame is *locally reflexive* if for every w and tuple $\langle a_1, \ldots, a_n\rangle \in U_w$ there is a $w \xrightarrow{C} w$ that contains $\langle a_1, a_1\rangle, \ldots, \langle a_n, a_n\rangle$ (i.e., if all finite sequences at all worlds are their own counterparts at that world). A frame is *reflexive* if the same relation can be chosen for all tuples, so that for all w there is a $w \xrightarrow{C} w$ such that aCa for all $a \in U_w$. (This corresponds to reflexivity of R^ω.) Kutz shows that the T-schema is valid in locally reflexivity frames, and that FK+T is complete wrt. this class. [80f.] Similarly, FK+4 is proved complete wrt. the class of locally transitive frames, and FK+B wrt. the class of locally symmetrical frames [81ff.]

12. Kutz points out that by an argument similar to one in [Skvortsov/Shehtman 1993, p.92] one can show in general that whenever L is a canonical propositional modal logic, then $FK + L$ is also canonical. [83]

13. Finally, Kutz presents a notion of *generalized frames* to prove (generalized) frame-completeness of all modal predicate logics.

11.4 Skvortsov and Shehtman

11.5 Goldblatt 2011

12 Individualistic semantics

Lewis's own semantics validates NE, and invalidates K. This is because it is individualistic. Let's look at some recent developments of this approach.

In a Ghilardi-style system, we would have both non-rigid, de dicto formulas like $\Box Fx$ and de re formulas like $\langle x : v\rangle\Box Fv$. Intuitively, $\langle x : v\rangle\Box Fv$ says that the x-individual has the property of being such that at all accessible worlds, it is F. By contrast, $\Box Fx$ says that at all accessible worlds, the x-individual at that world is F. The quantifier $\langle x : v\rangle$, just like $\forall v$ and $\exists v$, rigidifies the variable v.

To keep track of rigidified variables, formulas are interpreted relative to a triple of a world w, an interpretation function V, and a partial assignment function $\sigma : Var \to D_w$. The crucial clauses in the semantics are

$w, V, \sigma \Vdash \langle x : v\rangle A$	iff $w, V, \sigma' \Vdash A$ with σ' being like σ except that $\sigma'(v)$ is either $\sigma(x)$, if that is defined, or else $V_w(x)$.
$w, V, \sigma \Vdash \Box A$	iff $w', V, \sigma \cdot \tau \Vdash A$ for all worlds w' with wRw' and all $\tau : D_w \to D_{w'}$ that map individuals of w to one of their counterparts at w' (or to nothing if they don't have a counterpart there).

xxx no, this doesn't allow for contingent identity! need to use images!

Given an interpretation function V, a world w, and a partial function $\sigma : Var \to D_w$, I write V^σ for the function that assigns to any world w and variable x the value $\sigma(x)$ if that is defined, and otherwise $V_w(x)$. I will sometimes write w, V^σ instead of w, V, σ. By default, σ is the empty assignment. Non-empty assignments are only considered in the compositional evaluation of complex formulas containing dereifiers.

The functions σ play a similar role to the sequences in the Lawvere/Ghilardi framework. On this account, formulas are typed with a number, and express properties of corresponding sequences; e.g. '$w, V, \langle d_1, d_2, d_3\rangle \Vdash Fx_2 : 3$' iff the second member d_2 of the sequence $\langle d_1, d_2, d_3\rangle$ satisfies F.

The length of the relevant sequence is indicated by the postfix ': 3'; the index of the relevant individuals from the sequence by the choice of variable: 'x_1' picks out the first member, 'x_2' the second, and so on. In my semantics, the choice of variables is less restricted, since the functions σ register which variables denotes which element of a sequence. Moreover, the sequential character becomes unimportant: while one could order the elements of D_w and say that σ picks out the sequence of its range under that order, the order plays no role in my semantics. Most importantly, the "length" of σ (the cardinality of its domain) is rather unimportant. This is because the evaluation of the box is not restricted by the elements of σ. Whether or not $V_w(x)$ exists at all accessible worlds, $\langle x : v\rangle\Box Fv$ is true at w under V iff Fv is true at *all* accessible worlds w' under all variations $V^{\sigma\cdot\tau}$ of V, where $\sigma = \{\langle v, V_w(x)\rangle\}$ and τ is any counterpart function between w and w'.

Here is the complete semantics.

Definition 12.1 (Satisfaction in Ghilardi G)
The satisfaction relation $w, V, \sigma \Vdash A$ between a world w, an interpretation function V, a partial assignment σ, and a sentence A is defined by induction:

$w, V, \sigma \Vdash Px_1 \ldots x_n$	iff $\langle V^\sigma_w(x_1), \ldots, V^\sigma_w(x_n)\rangle \in V_w(P)$.
$w, V, \sigma \Vdash \neg A$	iff $w, V, \sigma \nVdash A$.
$w, V, \sigma \Vdash A \land B$	iff $w, V, \sigma \Vdash A$ and $w, V, \sigma \Vdash B$.
$w, V, \sigma \Vdash \forall xA$	iff $w, V, \sigma' \Vdash A$ for all x-variants σ' of σ with $\sigma'(x) \in D_w$.
$w, V, \sigma \Vdash \langle x : v\rangle A$	iff $w, V, \sigma' \Vdash A$ where σ' is the v-variant of σ such that $\sigma'(v) = V^\sigma_w(x)$.
$w, V, \sigma \Vdash \Box A$	iff $w', V, \sigma \cdot \tau \Vdash A$ for all worlds w' with wRw' and all $\tau : D_w \to D_{w'}$ that map individuals of w to one of their counterparts at w' (or to nothing if they don't have a counterpart there).

In my own systems, all occurrences of variables are interpreted de re – as if every formula was prefixed by substitution quantifiers. Since the domain of σ can be redundant, we may as well use σ to interpret *all* variables, and thus merge it into V.

For any interpretation function V and map τ, I write V^τ_w for the interpretation function V' that is like V except that for all worlds w' and variables x, $V'_{w'}(x) = \tau(V_w(x))$. So V^τ_w is V rigidified at w. (Note that V^τ_w may assign to x relative w' an individual that is not in $D_{w'}$, and it may assign to x relative to w' no individual at all if τ is partial; this first is an unimportant artifact, the second is important. Strictly speaking, V^τ is not an interpretation function in the sense defined above. Maybe I should distinguish "proper" from "improper" interpretation functions.)

The Geach quantifier plays the role of an *explicit substitution operator*. As such, it plays a prominent role in [**?**]'s study on counterpart semantics. In the counterpart logics of Ghilardi et al. (see [**?**], [**?**]), modal formulas function syntactically like predicates that can be applied to terms. So $(\Diamond Gxy)$ is treated like a binary predicate, and $(\Diamond Gxy)yy$ says that the pair y, y are things x, y such that some x counterpart is G-related to some y counterpart. This is just what $\langle y : x\rangle\langle y : y\rangle\Diamond Gxy$ says in the Geach notation. The modal predicates are immune to substitution. Hence applying UI to $\forall x(\Diamond Gxy)xy$ can only replace the x outside the modal predicate, resulting e.g. in $(\Diamond Gxy)yy$. (Corsi and Ghilardi et al. also use a *typed* language; I will return to that below.)

The revised version of (UI) only gives us

$$\forall x \Box Gxy \supset \langle y : x \rangle \Box Gxy.$$

To reach $\Box Gyy$, we might use one direction of the commutation principle for the substitution quantifier with the Box. This is known as the *continuity principle*:

(G5) $\quad \langle y : x \rangle \Box A \supset \Box \langle y : x \rangle A.$

[?] proves that (G1)–(G5), together with (UI), (LL) and the rest of QK, are complete for a certain class of models. Ghilardi et al. prove the same for (G5) alone, having (G1)–(G4) implicit in their definition of substitution. However, Corsi and Ghilardi also use types, and their logics validate CBF and NE (due to individualistim).

More importantly at this point, Corsi and Ghilardi do not use counterparts to evaluate ordinary modal formulas like $\Box Fx$. On their account, $\Box Fx$ is true relative to w, V iff for all accessible worlds, the individual that V assigns to x at w' satisfies F. On this reading, it is not at all clear that e.g.

$$\forall x \Box Gxy \supset \Box Gyy$$

is valid: it might be that at all worlds, all counterparts of actual things x are G-related to whatever y denotes there, while this y-thing is not G-related to itself; this only means that at some accessible world, y does not denote any counterpart of some actual thing. Ghilardi and Corsi typically add the "rigidity" constraint that $V_{w'}(x)$ must be some counterpart of $V_w(x)$. This makes the formula valid, but still for quite different reasons; and the motivation of the rigidity constraint remains unclear.

Let's call the Ghilardi reading of $\Box Gx$ *de dicto with respect to* x. To get my *de re* readings, every formula must be prefixed by substitution operators that bind all their free variables, which must be renamed. E.g. $\Box Fx$ becomes $\langle x : y \rangle \Box Fy$.

$$\forall x \Box Gxy \supset \Box Gyy$$

becomes

$$\langle y : z \rangle \forall x \Box Gxz \supset \langle y : z \rangle \Box Gzz.$$

To proof this, we can use (UI) and (G1) to get

$$\langle y : z \rangle \forall x \Box Gxz \supset \langle y : x \rangle \langle y : z \rangle \Box Gxz.$$

The required continuity principle for the last step leads from $\langle y : x \rangle \langle y : z \rangle \Box Gxz$ to $\langle y : z \rangle \Box Gzz$. The general principle is not

$$\langle x : y \rangle \langle x : z \rangle \Box A \supset \langle x : y \rangle \Box [z/y],$$

which goes wrong for $A = \Diamond Gyz$, but

(G5') $\quad \langle x : y \rangle \langle x : z \rangle \Box A \supset \langle x : y \rangle \Box \langle y : z \rangle A.$

13 Counterparts and intensional objects

14 Quantified hybrid logic

15 Multiple counterpart relations

Modal logic allows for multiple accessibility relations. Can we also incorporate multiple counterpart relations? There are philosophical motivations, but it's also clear that we're still missing a lot if we can only talk about one cross-world relation.

When we add multiple accessibility relations, we change the syntax to allow for multiple boxes, one for each R. Similarly, the perhaps most clearcut use of multiple counterpart relations is to allow for multiple *kinds of variables*, one for each C. Thus x_{C1} would be x-qua-$C1$, and x_{C2} would be x-qua-$C2$. The sorts of the variables would affect the definition of images: a w'-image V' of V at w must assign to each variable x_C a C-counterpart of its original referent.

Another option is to further multiply boxes: have $\Box_{C1}Fx$ say that under the $C1$-relation, all x-counterparts are F. But this is less perspicuous, and less expressive because we can't switch the counterpart relation mid-sentence.

Lumpl and Goliath. Qua-objects.

Remember the limitation we met earlier: we would like to say that x at w_1 *causes* y at w_2. To some extend, we could build this into the box (and the colons), by stipulating that it maps each individual to its causal successors. The box would then carry a doubly relational meaning: it would express a relation on W and also a (counterpart) relation on $W \times D$, and the two could be completely unrelated. We might write $\Box_{R,C}$ to indicate that this box traces worlds by R and individuals by C.

An alternative, inspired by Lumpl and Goliath, is to build the counterpart relations into the names. We would say $\Box_R F x_C$ to say that x causes an F at all R-accessible worlds.

Maybe this could be rendered more perspicuous by introducing a qua-operator that combines with names and associates it with a counterpart relation: $\langle Cx : y\rangle$ – "x under C is a y such that". Not clear why this is better than x^C ......

References

J. L. Bell and M. Machover [1977]: *A Course in Mathematical Logic*. Noth-Holland Publishing Company

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

Patrick Blackburn and Johan van Benthem [2007]: "Modal Logic: A Semantic Perspective". In [Blackburn et al. 2007], 1–85

Patrick Blackburn, Johan van Benthem and Frank Wolter (Eds.) [2007]: *Handbook of Modal Logic*. Dordrecht: Elsevier

David Bostock [1997]: *Intermediate Logic*. Oxford: Clarendon Press

Torben Braüner and Silvio Ghilardi [2007]: "First-order Modal Logic". In [Blackburn et al. 2007], 549–620

Giovanna Corsi [2002]: "Counterpart Semantics. A Foundational Study on Quantified Modal Logics". Manuscript

Melvin Fitting and Richard L. Mendelsohn [1998]: *First-Order Modal Logic*. Dordrecht: Kluwer

Dov Gabbay, Valentin Shehtman and Dmitrij Skvortsov [2009]: *Quantification in Nonclassical Logic*, vol Vol. 1. Dordrecht: Elsevier

James W. Garson [1984]: "Quantification in Modal Logic". In D. Gabbay and F. Guenthner (Eds.) *Handbook of Philosophical Logic,* vol 2. Dordrecht: Reidel, 249–307

S. Ghilardi [1991]: "Incompleteness results in Kripke semantics". *Journal of Symbolic Logic*, 56: 517–538

Silvio Ghilardi [2001]: "Substitution, quantifiers and identity in modal logic". In E. Morscher and A. Hieke (Eds.) *New essays in free logic,* Dordrecht: Kluwer, 87–115

Silvio Ghilardi and Giancarlo Meloni [1988]: "Modal and tense predicate logic: Models in presheaves and categorical conceptualization". In *Categorical algebra and its applications,* 130–142

— [1991]: "Philosophical and mathematical investigations in first-order modal logic". In *Problemi fondazionali in teoria del significato. Atti del convegno di Pontignano,* Firenze: Olsckhi, 77–107

Allen P. Hazen [1979]: "Counterpart-Theoretic Semantics for Modal Logic". *Journal of Philosophy*, 76: 319–338

G. E. Hughes and Max J. Cresswell [1996]: *A New Introduction to Modal Logic*. London and New York: Routledge

Markus Kracht and Oliver Kutz [2002]: "The semantics of modal predicate logic I. Counterpart frames". In Frank Wolter, Heinrich Wansing, Maarten de Rijke and Michael Zakharayaschev (Eds.) *Advances in Modal Logic,* vol 3. World Scientific Publishing Company

— [2005]: "The semantics of modal predicate logic II. Modal individuals revisited". In Reinhard Kahle (Ed.) *Intensionality,* Los Angeles: A. K. Peters

— [2007]: “Logically possible worlds and counterpart semantics for modal logic”. In Dale Jacquette (Ed.) *Philosophy of Logic,* Amsterdam: North Holland, 943–995

Saul Kripke [1963]: “Semantical considerations on modal logics”. *Acta Philosophica Fennica*, 16: 83–94

Steven T Kuhn [1980]: “Quantifiers as modal operators”. *Studia Logica*, 39(2-3): 145–158

Oliver Kutz [2000]: “Kripke-Typ-Semantiken für die modale Prädikatenlogik”. Diplomarbeit, Humboldt University Berlin

David Lewis [1968]: “Counterpart Theory and Quantified Modal Logic”. *Journal of Philosophy*, 65: 113–126

— [1983]: “Postscript to [Lewis 1968]”. In *Philosophical Papers I,* Oxford: Oxford University Press, 39–46

— [1986]: *On the Plurality of Worlds*. Malden (Mass.): Blackwell

Gerhard Schurz [2011]: “Combinations and Completeness Transfer for Modal Quantificational Logics”. *Logic Journal of the IGPL*, 19(4): 598–616

Theodore Sider [2001]: *Four-Dimensionalism.* Oxford: Clarendon Press

D. P. Skvortsov and V. B. Shehtman [1993]: “Maximal Kripke–Type Semantics for Modal and Superintuitionistic Predicate Logics”. *Annals of Pure and Applied Logic*, 63: 69–101

Robert Stalnaker and Richmond Thomason [1968]: “Abstraction in first-order modal logic”. *Theoria*, 34: 203–207

Johan van Benthem [1993]: “Beyond Accessibility”. In M. de Rijke (Ed.) *Diamonds and Defaults,* Dordrecht: Kluwer, 1–18

Lin Woollaston [1994]: “Counterpart Theory as a Semantics for Modal Logic”. *Logique et Analyse*, 147–148: 255–263