# 10 Semantics for Modal Predicate Logic

## 10.1 Constant domain semantics

In the previous chapter, we met the basic language $\mathfrak{L}_{MP}$ of (first-order) modal predicate logic. I mentioned that we normally interpret the modal operators of $\mathfrak{L}_{MP}$ as quantifiers over accessible worlds, while ordinary quantifiers $\forall x$ and $\exists y$ range over individuals that are assumed to inhabit these worlds. But more questions need to be answered.

Let's think about what a model for $\mathfrak{L}_{MP}$ might look like. Informally, a model should represent a conceivable scenario together with an interpretation of the non-logical vocabulary. We assume that in any relevant scenario, there are a number of "worlds" and a number of "individuals" that we want to talk about. So we have two domains, $W$ and $D$. As in chapter 3 (and all subsequent chapters), we assume that the modal operators quantify over worlds that are in some sense accessible from the world at which they are interpreted; so we also have an accessibility relation $R$ on $W$.

Next, we need to specify how the non-logical vocabulary of $\mathfrak{L}_{MP}$ is interpreted in a scenario. The non-logical vocabulary of $\mathfrak{L}_{MP}$ are the names and the predicates (except '='). Let's assume that each name picks out some individual. Intuitively, a predicate expresses a property of individuals, or a relation between individuals. In the semantics of $\mathfrak{L}_P$, we could represent properties by sets of individuals. This is no longer plausible in $\mathfrak{L}_{MP}$, because we usually want to allow the same individual to have different properties at different worlds. For example, if $F$ expresses the property of studying logic, then the set of individuals to whom $F$ applies will plausibly vary from world to world. To determine the truth-value of $Fa$ or $\exists xFx$ at a world, we need to know to which individuals $F$ applies *at that world*. So we will assume that the interpretation function assigns tuples of individuals to predicates *relative to any world*.

Here is the resulting definition of a model.

**Definition 10.1**

A **constant domain model** for modal predicate logic is a structure $M$ consisting of

1. a non-empty set $W$ (the "worlds"),
2. an accessibility relation $R$ on $W$,
3. a non-empty set $D$ (the "individuals"), and
4. an interpretation function $V$ that assigns
   - to each name a member of $D$, and
   - to each $n$-place predicate and world $w$ a set of $n$-tuples from $D$.

In constant domain models, the domain of individuals is "constant" insofar as it does not vary from world to world: at each world, the $\mathfrak{L}_P$-quantifiers range over the very same individuals. This may seem questionable – and we are soon going to question it – but it simplifies the semantics, so let's stick with it for the moment.

Having defined what a model looks like, we now have to specify which $\mathfrak{L}_{MP}$-sentences are true at a given world in a given model. Here we face the same problem with quantifiers that we faced in non-modal predicate logic: the truth-value of $\forall x Fx$ (at a world) is not determined by the truth-value of $Fx$ (at the same world, or at other worlds). So we need to relativise truth to three parameters: a model, a world, and an assignment function.

As in the previous chapter, I will use $[\tau]^{M,g}$ for the individual picked out by a term (name or variable) $\tau$ relative to a model $M = (D, W, R, V)$ and an assignment function $g$:

$$[\tau]^{M,g} =_{\text{def}} \begin{cases} V(\tau) & \text{if } \tau \text{ is a name} \\ g(\tau) & \text{if } \tau \text{ is a variable.} \end{cases}$$

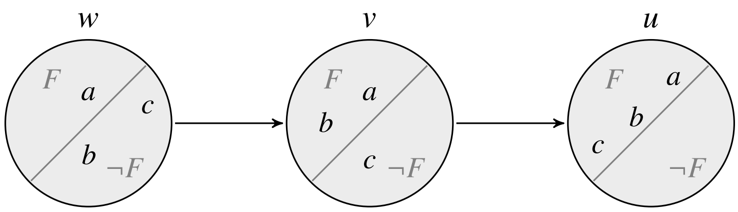
The obvious way of combining Kripke semantics (definition 3.2) with the semantics of classical predicate logic (definition 9.2) yields the following definition of truth at a world in a model relative to an assignment.

**Definition 10.2: Constant domain semantics**

If $M = \langle W, R, D, V \rangle$ is a constant domain model, $w$ is a member of $W$, $\phi$ is an $n$-place predicate (for $n \geq 1$), $\tau_1, \tau_2$ are terms, $\chi$ is a variable, and $g$ is a variable assignment, then

| | | |
|---|---|---|
| (a) | $M, w, g \models \phi\tau_1 \ldots \tau_n$ | iff $\langle [\tau_1]^{M,g}, \ldots, [\tau_n]^{M,g} \rangle \in V(\phi, w)$. |
| (b) | $M, w, g \models \tau_1 = \tau_2$ | iff $[\tau_1]^{M,g} = [\tau_2]^{M,g}$. |
| (c) | $M, w, g \models \neg A$ | iff $M, w, g \not\models A$. |
| (d) | $M, w, g \models A \land B$ | iff $M, w, g \models A$ and $M, w, g \models B$. |
| (e) | $M, w, g \models A \lor B$ | iff $M, w, g \models A$ or $M, w, g \models B$. |
| (f) | $M, w, g \models A \to B$ | iff $M, w, g \models B$ or $M, w, g \not\models A$. |
| (g) | $M, w, g \models A \leftrightarrow B$ | iff $M, w, g \models (A \to B)$ and $M, w, g \models (B \to A)$. |
| (h) | $M, w, g \models \forall\chi A$ | iff $M, w, g' \models A$ for all $\chi$-variants $g'$ of $g$. |
| (i) | $M, w, g \models \exists\chi A$ | iff $M, w, g' \models A$ for some $\chi$-variant $g'$ of $g$. |
| (j) | $M, w, g \models \Box A$ | iff $M, v, g \models A$ for all $v \in W$ such that $wRv$. |
| (k) | $M, w, g \models \Diamond A$ | iff $M, v, g \models A$ for some $v \in W$ such that $wRv$. |

Here is a (partial) picture of a constant domain model, with three worlds and three individuals.




Each world is inhabited by $a$, $b$, and $c$. At world $w$, only $a$ is $F$; at $v$, $a$ and $b$ are $F$; at $u$, all three individuals are $F$. So $Fa$ is true at $w$. $Fx$ is true at $w$ relative to an assignment $g$ that maps $x$ to $a$. By definition 10.2, this means that $\exists xFx$ is true at $w$. Along the same lines, we can figure out that $\exists xFx$ is true at $v$. Since $w$ can see $v$, it follows that $\Diamond\exists xFx$ is true at $w$. The *de re* sentence $\exists x\Diamond Fx$ is also true at $w$, because $\Diamond Fx$ is true at $w$ relative to an assignment $g$ that maps $g$ to (say) $b$.

**Exercise 10.1**

Which of the following statements about the above model $M$ are true?

(a) $M, w \models \Diamond Fc$
(b) $M, w \models \Box\exists x Fx$
(c) $M, w \models \forall x\Box Fx \to \Box\forall x Fx$
(d) $M, v \models \neg\forall x\Box\neg Fx$
(e) $M, u \models \Box\forall x Fx$
(f) $M, w \models \forall x(Fx \leftrightarrow \Box Fx)$
(g) $M, w \models \Box\forall x(\Diamond Fx \to Fx)$

Like in the previous chapter, we can define validity (and truth at a world) by quantifying over assignments: A sentence $A$ is **true at a world** $w$ in a constant-domain model $M$ iff $M, w, g \models A$ for all assignments $g$. A sentence is **CK-valid** iff it is true at all worlds in all models. (A *schema* is CK-valid if all its instances are CK-valid.) 'C' comes from 'constant domains'; 'K' indicates that we have put no constraints on the accessibility relation.

It is not hard to see that every $\mathfrak{L}_{MP}$-instance of a schema that is valid in classical predicate logic is CK-valid. Similarly, every $\mathfrak{L}_{MP}$-instance of a K-valid schema is CK-valid. But we also get some new interaction principles between modal operators and quantifiers. For example, the following schema, known as the *Barcan Formula* (after Ruth Barcan Marcus) is CK-valid.

(**BF**) $\quad \forall x\Box A \to \Box\forall x A$

**Observation 10.1: BF** is CK-valid.

*Proof.* Suppose (some instance of) $\forall x\Box A$ is true at some world $w$ in some constant domain model relative to some assignment $g$. By clause (h) of definition 10.2, it follows that $\Box A$ is true at $w$ relative to every $x$-alternative $g'$ of $g$. By clause (j) of definition 10.2, it follows that $A$ is true at every world $v$ accessibility from $w$ relative to every $x$-alternative $g'$ of $g$. By clause (h), this means that $\forall x A$ is true relative to $g$ at every world $v$ accessible from $w$. So by clause (j), $\Box\forall x A$ is true at $w$ relative to $g$. This shows that whenever $\forall x\Box A$ is true at some world $w$ relative some assignment $g$, then $\Box A\forall xA$ is also true at $w$ relative to $g$. By clause (f) of definition 10.2, it follows that $\forall x\Box A \to \Box A\forall xA$ is also true at every world relative to every assignment. □

Instead of working through definition 10.2, we can use trees to test if a sentence is CK-valid. The tree rules for CK are all the rules for K (from chapter 3) together with all the rules for standard predicate logic, with an added world parameter on each node that is held fixed when applying a rule from standard predicate logic. To get a complete proof system, we need one further identity rule, reflecting the fact that in constant-domain semantics, the reference of a name does not vary from world to world:

Identity Invariance

$$\begin{array}{cc} \eta_1 = \eta_2 & (\omega) \\ \vdots & \\ \eta_1 = \eta_2 & (v) \\ & \uparrow \\ & \text{old} \end{array}$$

Here is a tree proof for a simple instance of the Barcan Formula, $\forall x\Box Fx \to \Box\forall xFx$.

| | | | |
|---|---|---|---|
| 1. | $\neg(\forall x\Box Fx \to \Box\forall xFx)$ | $(w)$ | (Ass.) |
| 2. | $\forall x\Box Fx$ | $(w)$ | (1) |
| 3. | $\neg\Box\forall xFx$ | $(w)$ | (1) |
| 4. | $wRv$ | | (3) |
| 5. | $\neg\forall xFx$ | $(v)$ | (3) |
| 6. | $\neg Fa$ | $(v)$ | (5) |
| 7. | $\Box Fa$ | $(w)$ | (2) |
| 8. | $Fa$ | $(v)$ | (7,4) |
| | x | | |

And here is a proof of $\forall x\forall y(x = y \to \Box(x = y))$, the "necessity of identity":

| | | | |
|---|---|---|---|
| 1. | $\neg\forall x\forall y(x=y\to\Box(x=y))$ | $(w)$ | (Ass.) |
| 2. | $\neg\forall y(a=y\to\Box(a=y))$ | $(w)$ | (1) |
| 3. | $\neg(a=b\to\Box(a=b))$ | $(w)$ | (2) |
| 4. | $a=b$ | $(w)$ | (3) |
| 5. | $\neg\Box(a=b)$ | $(w)$ | (3) |
| 6. | $\neg\Box(b=b)$ | $(w)$ | (4, 5, LL) |
| 7. | $wRv$ | | (6) |
| 8. | $b\neq b$ | $(v)$ | (6) |
| 9. | $b=b$ | $(v)$ | (Ex) |
| | x | | |

**Exercise 10.2**

Use the tree method to show that the following schemas and sentences are CK-valid.

(a) $\Box\forall xA\to\forall x\Box A$
(b) $\exists x\Box A\to\Box\exists xA$
(c) $\forall x\Box(A\wedge B)\to\Box\forall xA$
(d) $\Box\Diamond\exists xA\to\Box\exists x\Diamond(A\vee B)$
(e) $\forall x\Box\exists y\,(y=x)$
(f) $\forall x\forall y(x\neq y\to\Box x\neq y)$

**Exercise 10.3**

The following sentences are CK-invalid. Can you describe a countermodel for each? (It may help to construct a tree and inspect its open branches.)

(a) $\Diamond\exists xFx\to\Diamond\exists x(Fx\wedge Gx)$
(b) $\Box\exists xFx\to\exists x\Box Fx$
(c) $\Box\forall x\forall y((Fx\wedge\neg Fy)\to x\neq y)$
(d) $\forall x\Box(Px\to Qx)\to\forall x(Px\to\Box Qx)$

We can get a sound and complete axiomatic calculus for CK by combining all the axioms and rules of classical predicate logic with those of K, and adding two new axioms: the Barcan Formula **BF** and the “necessity of distinctness”,

(**ND**) $\forall x \forall y (x \neq y \to \Box x \neq y)$.

Like in propositional modal logic, stronger logics can be defined by putting constraints on the accessibility relation. For example, we can define a system CT as the set of $\mathfrak{L}_{MP}$-sentences that are valid in the class of CK-models with a reflexive accessibility relation. CS4 could be defined as the set of $\mathfrak{L}_{MP}$-sentences that are valid in the class of reflexive and transitive CK-models. And so on.

Properties of the accessibility relation still correspond to modal schemas, just like in chapter 3: **T** corresponds to reflexivity, **4** to transitivity, **G** to convergence, etc. Officially, correspondence is defined in terms of frames. A *frame* is a model without an accessibility relation. In constant domain semantics, a frame therefore consists of two sets $W$ and $D$ and a relation $R$ on $W$. To say that **T** corresponds to reflexivity means that **T** is valid on all and only the reflexive frames.

Proof methods for CK are usually easy to adapt to stronger logics. If we want to use trees to test for CT-validity, we can simply add the Reflexivity rule to the rules for CK-validity. To test for CS4-validity, we would add the Reflexivity and Transitivity rules. To get a complete axiomatic calculus for CT we can similarly add the **T** schema to the calculus for CK; for CS4, we can add **T** and **4**. And so on.

But there are exceptions. Recall that the propositional system S4.2 is the set of $\mathfrak{L}_M$-sentences valid in the class of reflexive, transitive, and convergent Kripke models. Reflexivity corresponds to **T**, transitivity to **4**, and convergence to **G**. Accordingly, if we add these schemas to the axiomatic calculus for system K, we get a sound and complete calculus for S4.2. But if we add these schemas to the calculus for CK, we get a calculus that is *not* complete for CS4.2: there are $\mathfrak{L}_{MP}$-sentences that are valid in the class of reflexive, transitive, and convergent constant domain models that can’t be derived.

## 10.2 Quantification and existence

Let’s return to our assumption that $\forall x$ and $\exists x$ range over a fixed set of individuals, no matter at which world they are interpreted. For many applications, this seems problematic.

Consider the loaf of bread I baked today. Let's call it Loafy. Intuitively, Loafy could have failed to exist. I could have decided not to bake bread. Even if determinism is true, we can consider worlds at which the laws of nature or the origin of the universe are different. In many of these worlds, there are no humans, and no loafs of bread. So we should allow for worlds at which Loafy doesn't exist.

If we use $b$ as a name for Loafy, we can express Loafy's existence as

$$\exists x(x = b).$$

To see why this expresses Loafy's existence, consider a scenario in which Loafy exists. In that scenario, there is some thing $x$ which is identical to Loafy (namely, Loafy). Conversely, consider a scenario in which Loafy doesn't exist. In that scenario, there is no thing $x$ which is identical to Loafy. So $\exists x(x = b)$ is true in all and only the scenarios in which Loafy exists.

Now we can sharpen the above problem. Intuitively, it could have been the case that Loafy doesn't exist. So $\Diamond\neg\exists x(x = b)$ is true, on a suitable understanding of the diamond. But in constant domain semantics, this sentence is a contradiction: it is false at every world in every model.

A converse problem arises if we want to allow that something could have existed that doesn't actually exist. For example, let's assume that there could have been unicorns. If we interpret the predicate $U$ as '– is a unicorn' and the box as a suitable kind of circumstantial necessity, $\Box\forall x\neg Ux$ should then be false. But let's also assume that no individual in our world could have been a unicorn. So $\forall x\Box\neg Ux$ is true. We then have a counterexample to the Barcan Formula $\forall x\Box A \to \Box\forall xA$. And the Barcan Formula is valid in constant domain semantics.

**Exercise 10.4**

The **Converse Barcan Formula** is the schema $\Box\forall xA \to \forall x\Box A$. Explain why Loafy's possible non-existence seems to provide a counterexample to the Converse Barcan Formula.

**Exercise 10.5**

Consider the following four schemas.

(1) $\Diamond\exists xA \to \exists x\Diamond A$
(2) $\Box\exists xA \to \exists x\Box A$
(3) $\exists x\Box A \to \Box\exists xA$
(4) $\exists x\Diamond A \to \Diamond\exists xA$

(a) Are any of (1)–(4) equivalent to the Barcan Formula or the Converse Barcan Formula (given the duality of $\Box$ and $\Diamond$, of $\forall x$ and $\exists x$, and the standard truth-tables for propositional connectives)?
(b) Which of these schemas do you think are intuitively valid on a circumstantial interpretation of the box and the diamond?

One obvious response to these problems is to replace constant domain semantics by a different semantics in which the domain of individuals can vary from world to world. We will explore this option in the following section. First I want to mention two other lines of response.

Some philosophers have argued that we should bite the bullet: we are simply mistaken when we judge that Loafy could have failed to exist, or that anything could have existed that doesn't actually exist.

Of course, the problems we have reviewed don't just affect circumstantial modality. In temporal logic, biting the bullet means to accept that anything that has ever existed still exists today, and that anything that exists today has always existed and is always going to exist. In epistemic logic, biting the bullet means to accept that nobody can be unsure or ignorant about which individuals exists: if something exists, nobody can fail to know that it exists, nor can anyone believe that an individual exists that doesn't really exist.

A different response is to break the link between quantification and existence. $\exists x$ is traditionally called an "existential" quantifier, and pronounced 'there is an $x$' or 'there exists an $x$'. But $\mathfrak{L}_{MP}$ is a made-up language. We can make its symbols mean whatever we want. We can give a different interpretation of $\exists x$ on which 'Loafy exists' can't be translated as $\exists x(x = b)$.

One alternative to the standard interpretation of quantifiers goes back to the

Austrian philosopher Alexius Meinong. Meinong observed that when we describe beliefs, plans, hopes, or fears, we often seem to refer to non-existent objects. We might say that someone is afraid of *a ghost*, or that they are searching for *a golden mountain* – even though there are no ghosts or golden mountains. According to Meinong, people who are searching for a golden mountain are really searching for *something*. That something is a golden mountain. But it is not an existent golden mountain. Meinong concluded that besides existent mountains, there are also non-existent mountains. When we say that there are no golden mountains, we are quantifying only over existent individuals. If we include non-existent individuals in the domain of quantification, 'there are golden mountains' is true.

Quantifiers that range over both existent and non-existent individuals are called *Meinongian*. If the $\mathfrak{L}_{MP}$-quantifiers are Meinongian, then clearly $\exists x(x = b)$ does not translate 'Loafy exists'.

Meinong's postulation of non-existent individuals is widely rejected as incoherent. It certainly raises some difficult questions. Suppose you search for a golden mountain. You probably don't have any firm views about the mountain's height. You are not looking for a mountain that is exactly 2000 meters tall, nor are you looking for a mountain that is exactly 2100 meters tall. However, according to Meinong, the thing you are looking for is a genuine mountain. So it is a mountain that is not 2000 meters tall, not 2100 meters tall, and doesn't have any other particular height either. But isn't it incoherent to assume that there are mountains without a particular height? Besides, it also doesn't seem right to say that you are looking for a peculiar "mountain" that doesn't have any height and doesn't exist. Intuitively, you are looking for an *existent* mountain that *does* have a height.

Even if we accept Meinongian quantification as coherent, it is not clear whether it fully avoids the problem of constant domains. For example, couldn't I be unsure about how many things there are, even in Meinong's extended sense of 'there are'? To model my uncertainty in terms of accessible worlds, the domain of the Meinongian quantifier would have to vary from world to world.

A more straightforward alternative to the standard interpretation of quantifiers is the *possibilist* interpretation. Here we assume that $\forall x$ and $\exists x$ range not only over things that exist at the world at which the quantifiers are interpreted, but over everything that exists at any possible world. On this interpretation, too, $\exists x(x = b)$ no longer states that Loafy exists. It merely states that Loafy could have existed, in an unrestricted sense of 'could'. Constant domain semantics then only assumes that the set of individuals that exist at some world or other does not vary from world to world.

A downside of the possibilist interpretation is that it goes against the "internalist" spirit of modal logic. As we saw in section 9.2, one of the key features of modal logic is that it looks at the structure of worlds from the inside, from the perspective of a particular world, with only the modal operators providing (incomplete) access to other worlds. Possibilist quantifiers would provide unrestricted access to the inhabitants of other worlds.

Let's set aside these alternatives and see how constant domain semantics could be changed to allow for variable domains.

## 10.3 Variable domain semantics

In variable domain models, every world $w$ is associated with its own individual domain $D_w$. If $w$ and $v$ are different worlds, then Loafy the bread may be a the individuals in the local domain of the world at which they are interpreted: $\exists x Fx$ is true at $w$ iff $Fx$ is true (at $w$) of some individual in $D_w$.

So here is our revised definition of an $\mathfrak{L}_{MP}$-model.

**Definition 10.3**

A **variable domain model** for modal predicate logic is a structure $M$ consisting of

1. a non-empty set $W$ (the "worlds"),
2. an accessibility relation $R$ on $W$,
3. for each world $w$, a non-empty set $D_w$ (of "individuals"), and
4. an interpretation function $V$ that assigns
   - to each name a member of some domain $D_w$, and
   - to each $n$-place predicate and world $w$ a set of $n$-tuples from $D_w$.

To complete the semantics, we need explain how $\mathfrak{L}_{MP}$-sentences are interpreted relative to any given world in any model. This is not entirely straightforward.

To respect the intuition that Loafy could have failed to exist, we want $\Diamond\neg\exists x(x = b)$ to be true at some world $w$ in some model. This means that $\neg\exists x(x = b)$ should be true at some world $v$ accessible from $w$. Intuitively, $v$ will be a world at which Loafy doesn't exist. So we need to explain how a sentence that contains a name (here, $b$) should be interpreted at a world (here, $v$) where the thing that's picked out by the name doesn't exist.

In the case of $\neg\exists x(x = b)$, the sentence should come out true. But other cases are less clear. Consider $b = b$. Is Loafy identical to Loafy at $v$, where Loafy doesn't exist? Or consider $Fb$, $\neg Fb$, or $Fb \vee \neg Fb$. Is Loafy delicious at $v$? Is Loafy not delicious at $v$? Is Loafy either delicious or not delicious at $v$?

These questions are discussed not just in modal logic, but also in a branch of non-modal logic called **free logic**. Free logic differs from classical predicate logic by dropping the assumption that every name refers to an individual. The assumption is, after all, not true for names in natural language.

Consider the story of 'Vulcan'. Astronomers of the 19th century noticed that Mercury's path around the Sun only conforms to Newton's laws if there is another, smaller planet between Mercury and the Sun. With the help of Newton's laws, they calculated the size and position of that planet, and called it Vulcan. But Vulcan was never discovered. Eventually, Mercury's path was explained by Einstein's theory of relativity, without assuming any new planets. The name 'Vulcan' turned out to be "empty": it doesn't refer to anything.

So how should we formalize reasoning with empty names? The orthodox answer is that we shouldn't: the function of a name is to pick out an individual; if there is no individual to be picked out, we shouldn't use a name. Proponents of free logic disagree. They hold that we can perfectly well reason with empty names. We then need to answer the same questions I posed above: if $b$ is an empty name, how should we interpret $Fb$, or $\neg Fb$, or $Fb \vee \neg Fb$?

Within free logic, there are broadly three approaches.

The first is Meinongian. It assumes that apparently empty names are not really empty; they merely pick out a non-existent individual. Statements with such names are then interpreted as usual: $Fb$ may be true or false, depending on whether the (non-existent) individual picked out by $b$ has the property expressed by $F$.

Non-Meinongian versions of free logic usually assume that *atomic* sentences with empty names are never true: if $b$ is empty, then $Fb$ can't be true. Loosely speaking, the idea is that predicates express properties, and if something doesn't exist, then it doesn't have any properties. For example, it is not true that Vulcan is a planet – as you can see from the fact that Vulcan would not occur on a list of all planets. Nor is true that Vulcan orbits the sun, or that Vulcan has any particular mass.

What shall we say about $\neg Fb$ if $b$ is an empty name? In some versions of free logic, the standard semantic rules for complex sentences are applied: since $Fb$ is not true, $\neg Fb$ is true, and so is $Fb \vee \neg Fb$. Other versions of free logic assume that $\neg Fb$ is not true either. $Fb$ and $\neg Fb$ are treated as neither true nor false. This leads to a three-valued semantics that divides into several sub-flavours, with different verdicts on sentences like $Fb \vee \neg Fb$.

Each version of free logic can be used to give a semantics for modal predicate logic with variable domains. I am going to use the two-valued non-Meinongian approach, mainly because it is the simplest. So we'll assume that at worlds where Loafy doesn't exist, every atomic sentence involving a name for Loafy is false: $b = b$ is false, $Fb$ is also false, but $\neg Fb$ and $Fb \vee \neg Fb$ are true.

**Definition 10.4: Variable domain semantics**

If $M = \langle W, R, D, V \rangle$ is a variable domain model, $w$ is a member of $W$, $\phi$ is an $n$-place predicate (for $n \geq 1$), $\tau_1, \tau_2$ are terms, $\chi$ is a variable, and $g$ is a variable assignment, then

| | | |
|---|---|---|
| (a) | $M, w, g \models \phi\tau_1 \ldots \tau_n$ | iff $\langle [\tau_1]^{M,g}, \ldots, [\tau_n]^{M,g} \rangle \in V(\phi, w)$. |
| (b) | $M, w, g \models \tau_1 = \tau_2$ | iff $[\tau_1]^{M,g} = [\tau_2]^{M,g}$ and $[\tau_1]^{M,g} \in D_w$. |
| (c) | $M, w, g \models \neg A$ | iff $M, w, g \not\models A$. |
| (d) | $M, w, g \models A \wedge B$ | iff $M, w, g \models A$ and $M, w, g \models B$. |
| (e) | $M, w, g \models A \vee B$ | iff $M, w, g \models A$ or $M, w, g \models B$. |
| (f) | $M, w, g \models A \to B$ | iff $M, w, g \models B$ or $M, w, g \not\models A$. |
| (g) | $M, w, g \models A \leftrightarrow B$ | iff $M, w, g \models (A \to B)$ and $M, w, g \models (B \to A)$. |
| (h) | $M, w, g \models \forall\chi A$ | iff $M, w, g' \models A$ for all $\chi$-variants $g'$ of $g$ for which $g'(\chi) \in D_w$. |
| (i) | $M, w, g \models \exists\chi A$ | iff $M, w, g' \models A$ for some $\chi$-variant $g'$ of $g$ for which $g'(\chi) \in D_w$. |
| (j) | $M, w, g \models \Box A$ | iff $M, v, g \models A$ for all $v \in W$ such that $wRv$. |
| (k) | $M, w, g \models \Diamond A$ | iff $M, v, g \models A$ for some $v \in W$ such that $wRv$. |

As in constant domain semantics, we say that a sentence $A$ is **true at a world** $w$ in a model $M$ iff $M, w, g \models A$ for all assignments $g$. A sentence is **VK-valid** iff it is true at all worlds in all variable domain models.

The system VK is weaker than classical predicate logic: not everything that is valid in classical predicate logic is CK-valid. For example, both $b = b$ and $\exists x(x = b)$ are valid in classical predicate logic, but they are not true at every world in every variable domain model: if $V(b)$ is not a member of $D_w$, then $b = b$ and $\exists x(x = b)$ are false at $w$.

On the other hand, you can check that $\forall x(x = x)$ is VK-valid. So we don't just have to revise the rules for identity. We also need to revise the rule of "universal instantiation": from the fact that a universal generalisation like $\forall x(x = x)$ is true at a world, we can't infer that all its instances of are true (at the world): $b = b$ may be false. For another kind of example, consider a world $w$ where everything is made of chocolate. Let $F$ express the property of being made of chocolate. $\forall xFx$ is true at $w$. But we can't infer that Loafy the bread is made of chocolate ($Fb$) at $w$, for Loafy may not exist at $w$.

In proof systems for variable domain semantics, universal instantiation requires another premise: from $\forall xA$ we can infer $A[b/x]$ only if we also knows that $b$ exists – which can be expressed as $\exists x(x = b)$, or even simpler as $b = b$, given our assumption that atomic sentences with empty names are always false.

Here are the revised tree rules for VK. I give the quantifier rules for $\forall\chi A$ and $\exists\chi A$. You can find the rules for $\neg\forall\chi A$ and $\neg\exists\chi A$ by converting these into $\exists\chi\neg A$ and $\forall\chi\neg A$, respectively.

$$\begin{array}{ccc}
 & \forall\chi A \quad (\omega) & \\
 & \vdots & \\
\eta \neq \eta \quad (\omega) & & A[\eta/\chi] \quad (\omega)
\end{array}$$

↑ old

$$\begin{array}{c}
\exists\chi A \quad (\omega) \\
\vdots \\
\eta = \eta \quad (\omega) \\
A[\eta/\chi] \quad (\omega)
\end{array}$$

↑ new

Leibniz's Law remains the same. The Existence and Identity Invariance rules are changed as below. We also have a new rule for expanding atomic nodes. It allows us to infer that $b$ exists at a world from the assumption that $Fb$ is true at the world.

Existence

$$\begin{array}{c} \vdots \\ \eta = \eta \\ \uparrow \\ \text{new} \end{array}$$

Identity Invariance

$$\begin{array}{ll} \eta_1 = \eta_2 & (\omega) \\ \eta_1 = \eta_1 & (\nu) \\ \vdots & \\ \eta_1 = \eta_2 & (\nu) \end{array}$$

$$\begin{array}{ll} \Phi\eta_1 \dots \eta_n & (\omega) \\ \vdots & \\ \eta_1 = \eta_1 & (\omega) \\ \eta_2 = \eta_2 & (\omega) \\ \vdots & \\ \eta_n = \eta_n & (\omega) \end{array}$$

**Exercise 10.6**

Use the tree method to show that the following schemas and sentences are VK-valid.

(a) $\exists x \Box A \to \Box \exists x A$
(b) $\Box \forall x (A \to B) \to (\Box \forall x A \to \Box \forall x B)$
(c) $\Box \exists x (x = x)$
(d) $\Diamond Fa \to \Diamond \exists x Fx$
(e) $a = b \to \Box (a = a \to a = b)$

It is easy to check that the Barcan Formula $\forall x \Box A \to \Box \forall x A$ and its converse $\Box \forall x A \to \forall x \Box A$ are invalid in variable domain semantics. In fact, we can now prove that the Barcan formula corresponds to the assumption that whatever exists at an accessible world also exists at the original world. Similarly, the Converse Barcan Formula corresponds to the assumption that whatever exists at a world also exists at all accessible worlds.

**Observation 10.2:**

(i) **CBF** is valid on a variable domain frame iff the frame has *increasing domains*, meaning that whenever $wRv$, then $D_w \subseteq D_v$.

(ii) **BF** is valid on a variable domain frame iff the frame has *decreasing domains*, meaning that whenever $wRv$ then $D_v \subseteq D_w$.

*Proof of (i).* Recall that a frame is a model without an interpretation function. We'll proof the two directions separately.

Suppose some variable domain frame $F$ does not have increasing domains. Then $F$ contains a world $w$ in whose domain $D_w$ lies an individual $d$ which does not exist at some $w$-accessible world $v$. Let $V$ be an interpretation function on $F$ so that $V(F, w) = D_w$ and $V(F, v) = D_v$. In the model composed of $F$ and $V$, $\Box\forall x Fx$ is true at $w$, but $\forall x \Box Fx$ is false, since $d$ is not in $V(F, v)$. So **CBF** is not true at all worlds in all models based on $F$.

In the other direction, suppose **CBF** is not valid on a frame $F$. This means that there is a world $w$ in some model $M$ based on $F$ at which some instance of $\Box\forall x A$ is true while $\forall x \Box A$ is false. If $\forall x \Box A$ is false at $w$, then there is some $w$-accessible world $v$ at which $A$ is false for some individual $d$ in $D_w$. But since $\Box\forall x A$ is true at $w$, $A$ is true of all members of $D_v$. So $d$ is not in $D_v$. And so $F$ does not have increasing domains.

The proof of (ii) is similar. □

In modal propositional logic, we saw that many interesting modal schemas corresponded to properties of the accessibility relation. In variable domain semantics, interaction principles like **BF** and **CBF** correspond to frame conditions that link the accessibility relation with the domain of individuals.

**Exercise 10.7**

Definition 10.3 requires that every name in every model picks out a possible individual. In that sense, it does not allow for genuinely empty names. How could we change definitions 10.3 and 10.4 if we wanted to allow for names that don't pick out anything?

## 10.4 Trans-world identity

In section 9.5 I mentioned a problem for Leibniz' Law. In its traditional form, the Law allows us to reason from $\Box Fa$ and $a = b$ to $\Box Fb$. On some interpretations of the box, however, that inference looks invalid. Lois Lane knows that Superman can fly, and Superman is identical to Clark Kent. Does it follow that Lois knows that Clark Kent can fly?

If it does, we seem forced to conclude that Lois Lane has inconsistent beliefs, since she also believes that Clark Kent *cannot* fly. She would believe that Clark Kent can't fly, but also that he can fly. Intuitively, however, Lois's beliefs are perfectly consistent. What she lacks is information, not logical acumen. Her belief worlds are not worlds at which someone can both fly and not fly. Rather, they are worlds at which one person plays the Superman role and a different person plays the Clark Kent role.

Consider also the case of Julius. When we introduce the name 'Julius' for whoever invented the zip, we can be sure that Julius invented the zip. But it would be absurd to think that we found out who invented the zip merely by making a linguistic stipulation. If before introducing the name 'Julius', we were unsure whether the zip was invented by Benjamin Franklin or Whitcomb L. Judson, the introduction of the new name does nothing to remove our ignorance. There are still epistemically accessible worlds at which the zip was invented by Franklin and others at which it was invented by Judson. So knowing that Julius invented the zip is not the same thing as knowing that Judson invented the zip, even though Julius = Judson.

Similar problems have been argued to arise in the logic of circumstantial modality. Imagine a clay statue, standing on a shelf. Let's call it Goliath. Since Goliath is made of clay, there is also a piece of clay on the shelf, at the exact same spot as the statue. Let's call that piece of clay Lumpl. How is Lumpl related to Goliath? We might want to say that they are one and the same thing: Lumpl = Goliath. After all, there is only *one* statue-shaped object on the shelf, not two. But we might also want to say that Lumpl could have had the shape of a bowl, while Goliath could not: if the clay had been formed into a bowl rather than a statue, then Lumpl would have been a bowl, but Goliath, the statue, would not have existed. Goliath is necessarily not a bowl, but Lumpl is not necessarily not a bowl. We have $\Box\neg Bg$ but not $\Box\neg Bl$, even though $l = g$.

**Exercise 10.8**

Explain why the three examples I just presented also cast doubt on the "necessity of identity", the principle $a = b \to \Box(a = b)$.

Semantically, Leibniz' Law corresponds to the assumption that names are **directly referential**, meaning that the only contribution a name makes to the truth-value of a sentence is its referent. If two names have the same referent, it then makes no difference which of them we use: replacing one by the other never affects the truth-value of a sentence. Above, we have assumed direct reference in both constant and variable domain semantics. On either account, names are interpreted as simply picking out an individual; names that pick out the same individual are interchangeable.

It is a matter of debate whether names in ordinary language are directly referential. Some hold that Lois Lane really has inconsistent beliefs. Others hold that Lois neither believes that Superman can fly nor that Clark Kent cannot fly, because the objects of belief or knowledge are never adequately represented by statements involving ordinary names. (This also gets around the Julius problem.) With respect to Lumpl and Goliath, some simply deny that Lumpl is identical to Goliath.

We will not descend into these debates. Instead, let's explore how we could change our semantics for $\mathfrak{L}_{MP}$ to block the relevant applications of Leibniz' Law.

Here's a simple way to achieve this. So far, we have assumed that names are **rigid**: they pick out the same individual relative to any possible world. No matter at which world in a model the name $a$ is interpreted, it always picks out $V(a)$ – the individual assigned to $a$ by the model's interpretation function. A name like 'Julius', however, seems to be non-rigid. It seems to pick out different individuals relative different (epistemically) possible worlds. At any world, 'Julius' picks out whoever happens to be the inventor of the zip; at some worlds, that is Benjamin Franklin, at others it is Whitcomb L. Judson.

So let's assume that a model's interpretation function assigns individuals to names *relative to a world*. Equivalently, let's assume that each name is interpreted as expressing a *function from worlds to individuals*, telling us which individual the name picks out relative to any given world. Functions from worlds to individuals are known as **individual concepts**, which is why the present approach is often called **individual concept semantics**.

To motivate this label, return to Lois Lane. When Lois is thinking about Superman, she is thinking of the bold hero whose superpowers she has witnessed on several occasions. When she is thinking about Clark Kent, she is thinking of her shy and awkward colleague. Lois has distinct "concepts" for Superman and Clark Kent. The two concepts actually pick out the same person, but they do so via different roles. We can model a role as a function from worlds to individuals. The Superman role corresponds to a function that maps every world to whoever plays the Superman role at that world. The Clark Kent role corresponds to a function that maps every world to whoever plays the Clark Kent role at that world. For the world of the Superman stories, both functions return the same individual. For Lois Lane's belief worlds, they return different individuals.

Formally, both constant and variable domain semantics are easily changed into an individual concept semantics. We first change the definition of a model, so that $V$ assigns individual concepts to names. (In variable domain semantics, we might stipulate that the function maps worlds to individuals that exist at the relevant world.) It is advisable to give a parallel treatment for names and variables, so we'll also assume that assignment functions $g$ interpret variables as expressing individual concepts. In the truth definition, we replace $[\tau]^{M,g}$, by $[\tau]^{M,w,g}$, which is defined as the referent of $\tau$ in $M$ *at* $w$, relative to $g$. (That is, if $\tau$ is a name, then $[\tau]^{M,w,g} = V(\tau)(w)$; if $\tau$ is a variable, then $[\tau]^{M,w,g} = g(\tau)(w)$.) Finally, we adjust the definition of an $x$-variant so that $g'$ is an $x$-variant of $g$ iff $g'$ differs from $g$ at most in the individual concept it assigns to $x$.

The resulting logic of individual concepts has some unexpected features. For example, the following schema becomes valid:

$$\Box\exists xA \rightarrow \exists x\Box A$$

To see why, consider the instance $\Box\exists xFx \rightarrow \exists x\Box Fx$. Suppose the antecedent is true at some world in some model. This means that at every accessible world $v$, there is at least one $F$-individual. So there are functions that map every accessible world to some $F$-individual. Let $g'(x)$ be some such function. Relative to $g'$, $\Box Fx$ is true at $w$. So $\exists x\Box Fx$ is true at $w$.

This is widely regarded as problematic. It would suggest that the two readings of 'something necessarily exists' are actually equivalent: it is necessary that something or other exists just in case there is something that necessarily exists.

Another problematic feature of individual concept semantics is that the resulting logic has no sound and complete proof procedure. There are no tree rules, or natural deduction rules, or axioms and inference rules that would allow proving all and only the sentences that are true at all worlds in all models of individual concept semantics (no matter if we assume constant or variable domains). It's not just that no-one has yet found a suitable proof method. One can prove that no such method exists.

Both of these problems can be avoided by putting further constraints on models. So far, we have assumed that any function from worlds to individuals is a candidate interpretation for a name or a variable. Relative to a given assignment function, a variable may pick out Donald Trump in one world, the Eiffel tower in another, a fried egg in a third, and so on. Ordinary concepts are not that gerrymandered. So we might identify a certain subset of all individual concepts as "eligible" for being expressed by names or variables. If this is done sensibly, $\Box\exists xA \to \exists x\Box A$ becomes invalid, and complete proof methods become available.

**Exercise 10.9**

In individual concept semantics, both the necessity of identity and the necessity of distinctness are invalid. How could we change the semantics to make the necessity of identity valid, but not the necessity of distinctness? (Assume constant domains.)

**Exercise 10.10**

Translate the following sentence into $\mathfrak{L}_{MP}$:

> Some ticket will win, but I don't know if it will win.

Explain why the sentence poses a problem for accounts on which variables are treated as directly referential. Then explain how individual concept semantics avoids the problem.

An alternative to individual concept semantics is **counterpart semantics**. Here the guiding idea is that when a name occurs in the scope of a modal operator, then at the relevant accessible worlds it picks out whichever individual resembles its actual referent in certain respects. Consider Lois Lane's belief worlds – the worlds in which things are how Lois believes they are. In these worlds, Lois has a shy colleague called 'Clark Kent' who can't fly; there is also a superhero called 'Superman' who can fly; the two are different people. The shy colleague in Lois's belief worlds resembles the actual Clark Kent (by which I mean the Clark Kent of the Superman stories) in certain respects, which is why he is picked out by the name 'Clark Kent' when that name is interpreted relative to Lois's belief worlds. The superhero in Lois's belief worlds resembles the actual Clark Kent in other respects, respects that are associated with the name 'Superman'. The difference between the two names is that they invoke different criteria for trans-world resemblance.

If an individual at some world $v$ sufficiently resembles an individual at $w$ in relevant respects, then the first individual is called a **counterpart** of the second (relative to $v$ and $w$, and relative to the given resemblance criteria). The shy colleague in each of Lois Lane's belief worlds is a counterpart of Clark Kent relative to some resemblance criteria; the superhero in her belief worlds is a counterpart of Clark Kent relative to other criteria.

In counterpart semantics, we now assume that names are associated with an individual, but also with a counterpart relation which determines how the name should be interpreted when modal operators shift the point of evaluation to other worlds. $\Box Fa$ is true at a world $w$ iff at every accessible world $v$, every individual that stands in the $a$-relevant counterpart relation to $a$ at $w$ is $F$.

Counterpart relations play a similar role as individual concept, but they are more liberal. For example, we can allow for cases in which an individual has multiple counterparts at an accessible world, relative to the same counterpart relation. Suppose Bob lives next door to Alice, and has sometimes met her on the stairwell. For some reason, Bob believes that Alice's flat is inhabited by identical twins, so when he sees Alice, he thinks he sees one of the twins. Alice has two counterparts in Bob's belief worlds, but the two counterparts don't correspond to different roles or different similarity standards.

Counterpart relations also needn't be transitive: a counterpart of a counterpart of an individual need not itself be a counterpart of the individual. This might help with the following puzzle about metaphysical modality.

Intuitively, my bicycle could have been composed of somewhat different parts. If you exchange the lights or the seatposts on a bike, you aren't destroying the old bike and creating a new one. So my bike could have had different lights, or a different seatpost. On the other hand, arguably my bike could not have been composed of *entirely* different parts. A bike composed of entirely different parts would be a different bike. Now let $b$ denote my actual bike; let $F_1$ be a predicate that gives a detailed description of my bike as it actually is. Let $F_2$ give a detailed description of a bike that is just like mine except for the seatpost. We want to say that my bike could have fit that description. So $\Diamond F_2 b$ should be true. But as we make more and more changes to $F_1$, we reach a point – say $F_{32}$ – where the description could no longer have applied to my bike. So $\Diamond F_{32} b$ is false. But now consider what would have been the case if $F_{31} b$ had been true. In that case, it seems that $F_{32} b$ could have been true. After all, an $F_{32}$ bike differs from an $F_{31}$ bike only by a single part. It would be strange if the bike in a world where $F_{31} b$ is true could not possibly have had any different parts. So while $\Diamond F_{32}b$ is false at the actual world, it seems to be true at any world where $F_{31}b$ is true. Since $\Diamond F_{31}b$ is true at the actual world, it follows that $\Diamond\Diamond F_{32}b$ is true as well. So we have $\Diamond\Diamond F_{32}b$ but not $\Diamond F_{32}b$.

This is puzzling, because metaphysical possibility is often assumed to be an absolute kind of possibility, with a universal accessibility relation. One would then expect the logic of metaphysical possibility to be S5. Yet in S5, $\Diamond\Diamond A$ entails $\Diamond A$.

In counterpart semantics, modal schemas like $\Diamond\Diamond A \rightarrow \Diamond A$ correspond not just to properties of the accessibility relation but to combined properties of the accessibility relation and the counterpart relations. So one can explain what's going on in the puzzle of the bike without giving up the assumption that the accessibility relation for metaphysical modality is universal. $\Diamond\Diamond A \rightarrow \Diamond A$ is invalid because a counterpart of a counterpart of my bike need not be a counterpart of my bike.

**Exercise 10.11**

In the year 2030, Tim the time traveller will be visiting his younger self. He will enter his old house and find himself asleep. So the following seems true:

At some point in time, Tim will be asleep and Tim will be awake.

Translate this sentence into the language of temporal predicate logic. Explain why it poses a problem for accounts on which names express (temporal) individual concepts. Then explain how counterpart semantics avoids the problem.