10 Semantics for Modal Predicate Logic

10.1 Constant domain semantics

We have met the language $\mathfrak{L}_{MP}$ of (first-order) modal predicate logic. It is time to think about how this language should be interpreted. This will tell us which sentences and inferences in the language are valid.

As in modal propositional logic, we will assume that the box and the diamond are quantifiers over accessible worlds, where "accessibility" is a placeholder whose meaning depends on the application. If we want to reason about knowledge, a world v might be accessible from a world w iff v is compatible with what is known at w. If we're interested in metaphysical modality then a world v might be accessible from a world w iff it is compatible with the nature of things at w. Here we might, for example, read $\Diamond Fa$ as saying that Aristotle could have been a sailor, assuming that a picks out Aristotle and F the property of being a sailor.

Our topic in logic is not whether a particular claim about Aristotle is true. We want to know which statements are *logically true* or *valid*, meaning that they are true in any conceivable scenario, under any interpretation of the non-logical expressions (but holding fixed the meaning of the modal operators).

As always, we use models to represent a scenario together with an interpretation of the non-logical vocabulary. A model for $\mathfrak{L}_{MP}$ contains just enough information about a scenario and an interpretation to determine, for every $\mathfrak{L}_{MP}$-sentence and every world, whether the sentence is true at that world.

The non-logical vocabulary of $\mathfrak{L}_{MP}$ are the names and the predicates (with the exception of the identity predicate '='). Let's assume, for now, that the purpose of a name is simply to pick out an individual. Intuitively, a predicate picks out a property or relation. In non-modal predicate logic, we could represent these properties or relations by their extension – by the sets of individuals (or tuples of individuals) to which they apply. In modal predicate logic, however, we typically want to allow for scenarios in which an individual has different properties at different worlds. In one world, Aristotle might be a sailor, in another he might be a shoemaker. If F expresses the property of being a sailor, then the set of individuals to whom F applies will differ from world to world. To determine the truth-value of Fa at a world, we need to know to which individuals F applies *at that world*. A model's interpretation function will therefore assign a set of (tuples of) individuals to each predicate *relative to each world*.

Consider a model with two worlds w and v. Both worlds, let's assume, are accessible from w and neither is accessible from v. The model's interpretation function tells us that the name a picks out, say, Aristotle. It also tells us that the predicate F applies to Aristotle and Boethius at w and only to Boethius at v. We can write this as follows:

$V(a)$ = Aristotle
$V(F, w)$ = {Aristotle, Boethius}
$V(F, v)$ = {Boethius}

We don't know what property is expressed by F, nor which properties Aristotle and Boethius have at w and v. Nonetheless, we can figure out that Fa is true at w, because the predicate F applies to Aristotle at w. We can also figure out that Fa is false at v, and that $\Box Fa$ is false at w.

To determine the truth-value of arbitrary $\mathfrak{L}_{MP}$-sentences, we need some more information. As it stands, we can't tell whether (say) $\forall x Fx$ is true at w. Informally, $\forall x Fx$ says that every individual is F. We know that Aristotle and Boethius are F at w. But we don't know if there are other individuals besides Aristotle and Boethius. If yes, then $\forall x Fx$ is false at w. If no, the sentence is true. We therefore assume that a model for $\mathfrak{L}_{MP}$ also specifies a domain of individuals.

Definition 10.1

A **constant-domain Kripke model** for $\mathfrak{L}_{MP}$ is a structure M consisting of

1. a non-empty set W (the "worlds"),
2. a binary ("accessibility") relation R on W,
3. a non-empty set D (of "individuals"), and

4. an interpretation function V that assigns
   - to each $\mathfrak{L}_{MP}$-name a member of D, and
   - to each n-place predicate of $\mathfrak{L}_{MP}$ and world $w \in W$ a set of n-tuples from D.

Models of this type are called "constant-domain models" because the domain of individuals is the same for each world. This may seem questionable – and we are soon going to question it – but it simplifies the semantics. Let's stick with it for the moment.

Having defined a concept of a model, we can lay down the rules that determine whether any given $\mathfrak{L}_{MP}$-sentence is true at a world in a model.

In fact, truth will be defined relative to three parameters: a model, a world, and an assignment function. The assignment function plays the same role as in non-modal predicate logic. $\forall x \Diamond Fx$, for example, is true at a world w in a model iff there is some assignment of an individual to x that renders $\Diamond Fx$ true at w. We continue to use $[\tau]^{M,g}$ for the individual picked out by a term (name or variable) τ relative to a model $M = \langle D, W, R, V \rangle$ and an assignment function g:

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

Definition 10.2: Constant-domain Kripke semantics

If $M = \langle W, R, D, V \rangle$ is a constant-domain Kripke model, w is a member of W, ϕ is an n-place predicate (for $n \geq 0$), $\tau_1, \tau_2, \ldots, \tau_n$ are terms, χ is a variable, and g is a variable assignment, then

(a)	$M, w, g \models \phi\tau_1 \dots \tau_n$	iff $\langle [\tau_1]^{M,g}, \dots, [\tau_n]^{M,g} \rangle \in V(\phi, w)$.
(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)	$M, w, g \models A \wedge B$	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)	$M, w, g \models A \rightarrow B$	iff $M, w, g \not\models A$ or $M, w, g \models B$.
(g)	$M, w, g \models A \leftrightarrow B$	iff $M, w, g \models (A \rightarrow B)$ and $M, w, g \models (B \rightarrow A)$.
(h)	$M, w, g \models \forall\chi A$	iff $M, w, g' \models A$ for all χ-variants g' of g.
(i)	$M, w, g \models \exists\chi A$	iff $M, w, g' \models A$ for some χ-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.

A is **true at** *w* **in** *M* iff $M, w, g \models A$ for every assignment function g for M.

Let's return to the model from above, and let's add the information that the domain of individuals consists of just Aristotle and Boethius. That is, let M be the following model:

$$W = \{w, v\}$$
$$R = \{\langle w, w \rangle, \langle w, v \rangle\}$$
$$D = \{\text{Aristotle, Boethius}\}$$
$$V(a) = \text{Aristotle}$$
$$V(F, w) = \{\text{Aristotle, Boethius}\}$$
$$V(F, v) = \{\text{Boethius}\}$$

This isn't a complete specification of a model because I haven't assigned a meaning to names and predicates other than a and F, but we have enough information to determine the truth-value of any $\mathfrak{L}_{MP}$-sentence whose only non-logical vocabulary are a and F.

We can, for example, verify that Fa is true at w in M. A sentence is true at w in M iff it is true at w in M relative to every assignment function g. By clause (a) of definition 10.2, Fa is true at w in M relative to g iff $[a]^{M,g}$ is a member of $V(F, w)$. Since a is a name, $[a]^{M,g}$ is $V(a)$. And $V(a)$ is Aristotle. So Fa is true at w relative to g iff Aristotle is a member of $V(F, w)$. We know that $V(F, w)$ is {Aristotle, Boethius}. Aristotle evidently is a member of {Aristotle, Boethius}. So Fa is true at w in M, relative to any assignment g.

We can also verify that $\Box Fa$ is false at w. By clause (j) of definition 10.2, $\Box Fa$ is true at w (in M relative to g) iff Fa is true (in M relative to g) at all worlds accessible from w. And Fa is false at v because Aristotle is not a member of {Boethius}.

Exercise 10.1

Which of the following sentences are true at w in M?

(a) $\neg Fa \to Fa$
(b) $\Box \exists x Fx$
(c) $\Box \forall x Fx$
(d) $\exists x \Box Fx$
(e) $\forall x \Box Fx$
(f) $\forall x(\Box Fx \to \Box\Box Fx)$

Validity is truth at all worlds in all models of a certain kind. A sentence is **CK-valid** iff it is true at all worlds in all constant-domain Kripke models. 'C' comes from 'constant domains'; 'K' indicates that we have put no constraints on the accessibility relation. We get stronger concepts of validity – stronger logics – if we require the accessibility relation to be reflexive, or transitive, or euclidean, etc.

It is not hard to see that every sentence that is valid in classical predicate logic is CK-valid. Similarly, every K-valid sentence is CK-valid. We also get some new interaction principles between modal operators and quantifiers. For example, consider the following schema, known as the *Barcan Formula*, after Ruth Barcan Marcus.

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

Observation 10.1: All instances of (BF) are CK-valid.

Proof. Suppose a sentence $\forall x \Box A$ is true at some world w in some constant-domain model M 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-variant 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-variant 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.

We've shown that whenever $\forall x \Box A$ is true at some world w in some model M

relative some assignment g, then $\Box A \forall x A$ is also true at w in M relative to g. By clause (f) of definition 10.2, it follows that $\forall x \Box A \rightarrow \Box A \forall x A$ is true at every world in every model 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. (In the predicate logic rules, a name counts as 'old' if it already occurs on the relevant branch, no matter at which world.)

To get a complete proof system, we need one further identity rule, reflecting the fact that the reference of a name does not vary from world to world:

Identity Invariance

$$\eta_1 = \eta_2 \quad (\omega)$$

$$\eta_1 = \eta_2 \quad (\nu)$$

↑
old

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

1.	$\neg(\forall x \Box Fx \rightarrow \Box \forall x Fx)$	(w)	(Ass.)
2.	$\forall x \Box Fx$	(w)	(1)
3.	$\neg \Box \forall x Fx$	(w)	(1)
4.	wRv		(3)
5.	$\neg \forall x Fx$	(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)	(SI)
	x		

Exercise 10.2

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

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

There are also axiomatic calculi for CK. We can, for example, combine the axiom schemas and rules of classical predicate logic with those of K, and add two new schemas: the Barcan Formula (BF) and the "necessity of distinctness",

(ND) $\forall x \forall y (x \neq y \rightarrow \Box x \neq y)$.

As I mentioned above, stronger logics can be defined by putting constraints on the accessibility relation. For example, the system **CT** is the set of $\mathfrak{L}_{MP}$-sentences that are valid in the class of constant-domain Kripke models with a reflexive accessibility relation. **CS4** is the set of $\mathfrak{L}_{MP}$-sentences that are valid in the class of constant-domain Kripke models with a reflexive and transitive accessibility relation. And so on.

Properties of the accessibility relation still correspond to modal schemas, just as in chapter 3: (T) corresponds to reflexivity, (4) to transitivity, (G) to convergence, etc. Recall that a schema *corresponds* to a property of the accessibility relation if the schema is valid in all and only the frames in which the accessibility relation has that property. A *frame* is a model without an interpretation function. In the present context, a frame therefore consists of two non-empty sets W and D and a relation R on W.

We can still use the tree method or the axiomatic method to test for validity in logics stronger than CK. To test for CT-validity, for example, we would add the Reflexivity rule to the tree rules for CK. To test for CS4-validity, we would add the Reflexivity and Transitivity rules. We can get an axiomatic calculus for CT by adding the (T)-schema to the calculus for CK; for CS4, we can add (T) and (4). And so on for other systems.

But there are exceptions. Remember S4.2 – 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). 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 the schemas to the calculus for CK, the resulting calculus 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

We have assumed that the domain of individuals is the same for every world. This may seem problematic.

Earlier today I was baking bread. Let's call the loaf of bread that I made 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 arguably express Loafy's existence as

$$\exists x\, x = b.$$

Why might this express that Loafy exists? Consider a scenario in which Loafy does exist. In that scenario, there is some thing x which is identical to Loafy (namely, Loafy). Conversely, consider a scenario in which Loafy does not 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 worry. 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, that sentence is a contradiction: it is false at every world in every model.

A converse problem arises if we think 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 all instances of the Barcan Formula are valid in constant-domain semantics.

Exercise 10.4

The **Converse Barcan Formula** is the schema $\Box\forall xA \to \forall x\Box A$. All instances of the Converse Barcan Formula are CK-valid. 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) $\Diamond \exists x A \to \exists x \Diamond A$
(2) $\Box \exists x A \to \exists x \Box A$
(3) $\exists x \Box A \to \Box \exists x A$
(4) $\exists x \Diamond A \to \Diamond \exists x A$

(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 metaphysical interpretation of the box and the diamond?

An obvious response to these problems is to replace constant-domain semantics with a 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. 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 $\exists x$ so that 'Loafy exists' can't be translated as $\exists x\, x = b$.

One alternative to the standard interpretation of quantifiers is associated with 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.

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 difficult questions. Suppose you are searching 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. On the Meinongian account, there is a genuine mountain that you are looking for. It is a mountain that is not 2000 meters tall, not 2100 meters tall, and doesn't have any other particular height either. But how could there be a mountain without any 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.

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.

One 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. Loafy the bread may be a member of D_w but not of D_v. Quantifiers range over 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.

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

Definition 10.3

A **variable-domain Kripke model** for $\mathfrak{L}_{MP}$ is a structure M consisting of

1. a non-empty set W (the "worlds"),
2. a binary ("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 to explain how $\mathfrak{L}_{MP}$-sentences are interpreted relative to any given world in a variable-domain model. This raises a problem.

Since Loafy could have failed to exist, we want to have models in which $\Diamond\neg\exists x\, x = b$ is true at some world w. It follows that $\neg\exists x\, x = b$ is true at some world v accessible from w. Intuitively, v is a world at which Loafy doesn't exist. The problem is that 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. Other cases are less clear. What about $b = b$? Is Loafy identical to Loafy at v, where Loafy doesn't exist? What about 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 has a referent. The assumption is, after all, not true for names in natural language.

Consider the story of 'Vulcan'. In the 19th century, it was observed that Mercury's path around the Sun conforms to Newton's laws only if there is another, smaller planet between Mercury and the Sun. With the help of Newton's laws, astronomers 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.

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 that I posed above: if b is an empty name, how should we interpret $b = b$, Fb, $\neg Fb$, and $Fb \vee \neg Fb$?

Within free logic, there are broadly three approaches.

The first is Meinongian. It assumes that apparently empty names are not really empty after all; 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. 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 it true that Vulcan orbits the sun, or that Vulcan has any particular mass.

What shall we say about $\neg Fb$ then, 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 if b doesn't refer then neither Fb nor $\neg Fb$ is true. Since a sentence is called false iff its negation is true, this means that Fb and $\neg Fb$ are neither true nor false. We get a three-valued semantics that can be spelled out in different ways, 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. We will 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 Kripke semantics

If $M = \langle W, R, D, V \rangle$ is a variable-domain Kripke model, w is a member of W, ϕ is an n-place predicate (for $n \geq 0$), $\tau_1, \ldots, \tau_n$ are terms, χ 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)	$M, w, g \models A \rightarrow B$	iff $M, w, g \not\models A$ or $M, w, g \models B$.
(g)	$M, w, g \models A \leftrightarrow B$	iff $M, w, g \models (A \rightarrow B)$ and $M, w, g \models (B \rightarrow A)$.
(h)	$M, w, g \models \forall\chi A$	iff $M, w, g' \models A$ for all χ-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 χ-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.

A is **true at** w **in** M iff $M, w, g \models A$ for all assignments g for M.

A sentence is **VK-valid** ('V' for 'variable-domain') 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, or at all worlds), we can't infer that all its instances are true: $b = b$ may be false. For another example, consider a world w where everything is made of chocolate. Let F express the property of being made of chocolate. $\forall x Fx$ 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 the type of free logic we have adopted, the rule of universal instantiation requires another premise: from $\forall x A$ we can infer $A[b/x]$ only if we also know 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 only 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.

$$\forall \chi A \quad (\omega)$$

$$\eta \neq \eta \quad (\omega) \qquad\qquad A[\eta/\chi] \quad (\omega)$$

↑ old

$$\exists \chi A \quad (\omega)$$

$$\eta = \eta \quad (\omega)$$
$$A[\eta/\chi] \quad (\omega)$$

↑ new

We keep the rule for Leibniz's Law. But we replace the Self-Identity and Identity Invariance rules by the following three rules.

Existence

$$\eta = \eta \quad (\omega)$$

↑ new

Identity Invariance

$$\eta_1 = \eta_2 \quad (\omega)$$
$$\eta_1 = \eta_1 \quad (\nu)$$

$$\eta_1 = \eta_2 \quad (\nu)$$

$$\Phi \eta_1 \dots \eta_n \quad (\omega)$$

$$\eta_1 = \eta_1 \quad (\omega)$$
$$\eta_2 = \eta_2 \quad (\omega)$$
$$\vdots$$
$$\eta_n = \eta_n \quad (\omega)$$

The Existence rule reflects our assumption that the domain of individuals is never empty. The unnamed last rule is a rule for expanding atomic nodes. From the assumption that Fb is true at a world, for example, the rule allows us to infer that b exists at that world, which can be expressed as $b = b$. We then don't need a separate rule of Self-Identity.

Exercise 10.6

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

(a) $\exists x \Box Fx \to \Box \exists x Fx$

(b) $\Box \forall x(Fx \to Gx) \to (\Box \forall x Fx \to \Box \forall x Gx)$

(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. (By this I mean that not all their instances are valid.) 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, while its converse 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). Suppose some variable-domain frame F does not have increasing domains. Then F has a world w whose domain D_w contains an individual d that 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 of 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. □

Exercise 10.7

Definition 10.3 requires that every name in every model picks out a possible individual. In that sense, the definition 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 an apparent problem with Leibniz' Law. The Law allows us to reason from $\Box Fa$ and $a = b$ to $\Box Fb$. On some interpretations of the box, however, the inference looks problematic. In the Superman stories, Lois Lane knows that Superman can fly, and Superman is identical to Clark Kent. Can we infer that Lois knows that Clark Kent can fly?

If we can, we would have 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 have 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. Knowing that Julius invented the zip is not the same thing as knowing that Judson invented the zip, even if in fact Julius = Judson.

Similar problems have been argued to arise in the logic of metaphysical 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", $\forall x \forall y (x = y \rightarrow \Box x = y)$.

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 names are directly referential, and two names have the same referent, then it makes no difference which of them we use: replacing one by the other never affects the truth-value of a sentence.

So far, we have assumed direct reference in both constant-domain and variable-domain semantics. On either account, names are interpreted as simply picking out an individual. 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. There are several ways to achieve this. We will only look at one.

The approach we will explore drops the assumption that names are rigid. A name is **rigid** if it picks out the same individual relative to any possible world. Earlier, we assumed that no matter at which world the sentence Fa is interpreted, the name a always picks out the same individual, $V(a)$. A name like 'Julius', however, seems to be non-rigid. It picks out different individuals relative to different (epistemically) possible worlds. Relative to a world where Benjamin Franklin invented the zip, ‘Julius’ picks out Benjamin Franklin. Relative to a world where Whitcomb L. Judson invented the zip, the name picks out Whitcomb L. Judson.

Let’s assume, then, that a model’s interpretation function assigns an individual to each name *relative to each world*. This is equivalent to assuming 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 about the audacious hero whose superhuman powers she has witnessed on several occasions. When she is thinking about Clark Kent, she is thinking about her shy and awkward colleague. Lois has distinct “concepts” for Superman and Clark Kent, one associated with the Superman role, the other with the Clark Kent role. The two concepts actually pick out the same person because one and the same person plays both the Superman role and the Clark Kent role. We can model each of these roles as a function from worlds to individuals. The Superman role is represented by a function that maps every world to whoever plays the Superman role at that world. The Clark Kent role is represented by 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.

Exercise 10.9

What individual concepts might be associated with the names ‘Lumpl’ and ‘Goliath’?

We can easily convert our earlier constant-domain and variable-domain semantics into an individual concept semantics. We first need to change the definition of a model, so that V assigns individual concepts to names. In variable-domain semantics, we might stipulate that an individual concept never maps a world to an individual that doesn’t exist at the world. We might also want to allow for “partial concepts”: individual concepts that don’t return any value for certain worlds.

It is advisable to give a parallel treatment for names and variables. So we'll also assume that an assignment function g interprets each variable as expressing an individual concept. In the truth definition, we replace $[\tau]^{M,g}$, by $[\tau]^{M,w,g}$, which is defined as the referent of τ in M *at* w, relative to g. (That is, if τ is a name, then $[\tau]^{M,w,g} = V(\tau)(w)$; if τ 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, all instances of the following schema become valid:

$$\Box\exists xA \to \exists x\Box A$$

To see why, consider the instance $\Box\exists xFx \to \exists x\Box Fx$. Suppose the antecedent is true at some world in some model. This means that at every accessible world v, there is at least one individual that is F. In this case, there are functions that map every accessible world to some individual that is F. 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. 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. We might therefore 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 10.10

The following line of thought may be attributed to Descartes. "I am certain that I exist, but not that my body exists. [After all, it could turn out that I am a disembodied soul.] Therefore: I am not my body." Translate the argument into $\mathfrak{L}_{MP}$. Is it CK-valid? Is it VK-valid? Do you find it convincing?

Exercise 10.11

The following sentence sounds contradictory.

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

Translate the sentence into $\mathfrak{L}_{MP}$. Explain why its apparent contradictoriness poses a problem for accounts on which variables are treated as directly referential.

Exercise 10.12

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.)