Modal Correspondence Theory

Contents

Part I

Background: Frames, Models, and Frame Properties

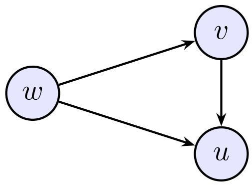
1 Kripke Frames and Models

1.1 Frames

Definition 1.1 (Frame)**.** A **frame** is a pair $\mathcal{F} = \langle W, R\rangle$ where

- W is a non-empty set (the set of *possible worlds*, or *indices*),
- $R \subseteq W \times W$ is a binary relation on W (the *accessibility relation*).

A frame is simply a directed graph: the worlds are the vertices and the accessibility relation gives the edges.



A frame $\langle\{w, v, u\}, R\rangle$ *with* Rwv*,* Rwu*,* Rvu*.*

1.2 Kripke Models

Definition 1.2 (Kripke model)**.** A **Kripke model** is a triple $\mathcal{M} = \langle W, R, I\rangle$ where $\langle W, R\rangle$ is a frame and $I : \mathrm{Prop} \times W \to \{T, F\}$ is an *interpretation function* (or *valuation*) assigning a truth-value to each propositional variable at each world.

Satisfaction of a formula φ at a world w in model $\mathcal{M}$ is written $\mathcal{M}, w \models \varphi$ and is defined recursively:

- $\mathcal{M}, w \models p$ iff $I(p, w) = T$;
- Boolean connectives as usual;
- $\mathcal{M}, w \models \Box\varphi$ iff for every $v \in W$ with Rwv, $\mathcal{M}, v \models \varphi$;
- $\mathcal{M}, w \models \Diamond\varphi$ iff there exists some $v \in W$ with Rwv and $\mathcal{M}, v \models \varphi$.

Think of $\Box\varphi$ as "φ holds in every world reachable from here" and $\Diamond\varphi$ as "φ holds in *some* world reachable from here."

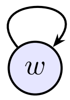
2 Properties of Frames

Since a frame is a directed graph, we can classify frames by the structural properties of R.

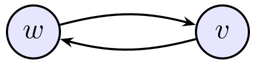
Definition 2.1 (Frame properties)**.** Let $\mathcal{F} = \langle W, R\rangle$ be a frame.

Property	**FOL condition**	**Intuition**
Reflexive	$\forall w\, Rww$	every world sees itself
Transitive	$\forall u\,\forall v\,\forall w\ ((Ruv \wedge Rvw) \to Ruw)$	reachability chains compose
Symmetric	$\forall v\,\forall w\ (Rvw \to Rwv)$	access is bidirectional
Euclidean	$\forall u\,\forall v\,\forall w\ ((Ruv \wedge Ruw) \to Rvw)$	siblings see each other
Serial	$\forall v\,\exists w\, Rvw$	no dead ends

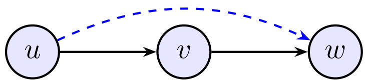
The following diagrams illustrate each property.



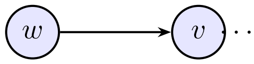
Reflexive: Rww



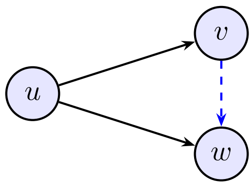
Symmetric: $Rwv \Rightarrow Rvw$



Transitive: $Ruv \wedge Rvw \Rightarrow Ruw$



Serial: $\exists v\, Rwv$



Euclidean: $Ruv \wedge Ruw \Rightarrow Rvw$

Exercise 2.1. Show that a reflexive and euclidean relation is also symmetric and transitive. Conversely, show that a reflexive, symmetric, and transitive relation is euclidean.

Hint: for symmetry from reflexivity + euclideanness, take $u = v$ in the euclidean condition; for transitivity, use symmetry first to swap an edge, then apply euclideanness.

Part II
Modal Correspondence Theory

3 The Central Idea

Definition 3.1 (Modal correspondence)**.** A modal formula (schema) φ **corresponds to** (or **defines**) a frame property P if and only if:

$$\text{for every frame } \mathcal{F}, \quad \mathcal{F} \models \varphi \iff \mathcal{F} \text{ has property } P.$$

Here $\mathcal{F} \models \varphi$ means that φ is valid in $\mathcal{F}$, i.e. $\mathcal{M}, w \models \varphi$ for every model $\mathcal{M}$ built on $\mathcal{F}$ and every world w in $\mathcal{M}$.

In other words, *the class of frames validating the formula is exactly the class of frames having the property.*

Notice the two quantifiers hidden inside “$\mathcal{F} \models \varphi$”:

$$\mathcal{F} \models \varphi \quad \overset{\text{def}}{\iff} \quad \text{for } \textbf{every} \text{ model } \mathcal{M} \text{ on } \mathcal{F}, \text{ for } \textbf{every} \text{ world } w \text{ in } \mathcal{M}, \ \mathcal{M}, w \models \varphi.$$

The formula must hold at every world, under every possible interpretation of the propositional variables. This is a very strong requirement, and confusing it with a weaker one is the most common mistake students make with correspondence theory. The next subsection is devoted entirely to clarifying this point.

3.1 Caution! Model validity is not frame validity

What the definition does NOT say. If a *particular model* $\mathcal{M}$ on a frame $\mathcal{F}$ satisfies

$$\mathcal{M}, w \models \Box\varphi \to \varphi \quad \text{at every world } w,$$

then $\mathcal{F}$ is reflexive. **This is false!**

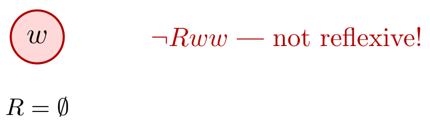
What the definition DOES say. If *every possible model* $\mathcal{M}$ on $\mathcal{F}$ satisfies the formula at every world, then $\mathcal{F}$ is reflexive.

The difference between "a model satisfies the formula" and "a frame validates the formula" is the difference between one particular interpretation working out and *every conceivable* interpretation working out. Let us see a concrete counter-example that makes the distinction vivid.

A counter-example: one world, no accessibility. Consider the simplest possible frame:

$$W = \{w\}, \qquad R = \emptyset.$$

There is a single world w, and the accessibility relation is *empty*: w does not see any world, not even itself. In particular, $\neg Rww$, so **the frame is not reflexive**.



Now build a model on this frame by setting $I(w, p) = T$ for every propositional variable p. Let us check whether $\mathcal{M}, w \models \Box p \to p$:

1. **Is $\mathcal{M}, w \models \Box p$?** By the semantic clause, $\mathcal{M}, w \models \Box p$ iff for every v with Rwv, $\mathcal{M}, v \models p$. But $R = \emptyset$: there is *no* v with Rwv. The universal quantifier ranges over an empty set, so the condition is **vacuously true**. Therefore $\mathcal{M}, w \models \Box p$. (When no world is accessible, *everything* is necessary.)
2. **Is $\mathcal{M}, w \models p$?** Yes, because $I(w, p) = T$.
3. **Conclusion:** $\mathcal{M}, w \models \Box p \to p$. (True $\to$ True $=$ True.)

In fact, more is true: since $I(w, p) = T$ for *every* variable p, the same argument shows that $\mathcal{M}, w \models \Box\varphi \to \varphi$ for *every* formula φ. So this model satisfies axiom M at every world (there is only one world).

And yet the frame is not reflexive!

A student who confuses "the model satisfies the formula" with "the frame validates the formula" would wrongly conclude that the frame must be reflexive. The correspondence theorem says no such thing.

Why the frame does not validate M. To show that the *frame* validates M, we would need *every* model on the frame to satisfy the formula. Let us build a different model on the same frame:

$$I'(w, p) = F.$$

Check $\mathcal{M}', w \models \Box p \to p$:

1. $\mathcal{M}', w \models \Box p$: still vacuously true (no accessible worlds).
2. $\mathcal{M}', w \not\models p$: because $I'(w,p) = F$.
3. Therefore $\mathcal{M}', w \not\models \Box p \to p$. (True $\to$ False $=$ False.)

So the model $\mathcal{M}'$ **falsifies** axiom M. The frame does not validate M. And indeed, the frame is not reflexive, just as the correspondence theorem predicts.

The moral. The first model ($I(w,p) = T$ everywhere) was a *lucky* interpretation: it happened to make p true at w, which made $\Box p \to p$ true for trivial reasons that had nothing to do with the structure of the accessibility relation. The second model ($I'(w,p) = F$) exposed the real problem: $\Box p$ can be vacuously true while p is false, precisely because w cannot see itself.

This is why the correspondence definition quantifies over *all* models on the frame: it ensures that the validity of the formula is due to the *structural* properties of R (here, reflexivity), not to an accidental choice of interpretation.

Rule of thumb. When working with correspondence theory, always ask:
- Does the formula hold because of the *structure of* R? $\longrightarrow$ frame property.
- Or does it hold because of a *particular choice of* I? $\longrightarrow$ model accident, says nothing about the frame.

The ($\Rightarrow$) direction of each correspondence theorem works by *contrapositive*: if R lacks the structural property, we explicitly *construct* a model (a bad choice of I) that falsifies the formula. That is why the counter-model constructions always define a specific valuation.

4 The Five Correspondences at a Glance

Axiom	**Modal formula**	**Frame property**
D	$\Box\varphi \to \Diamond\varphi$	Serial: $\forall w\, \exists v\; Rwv$
M (or T)	$\Box\varphi \to \varphi$	Reflexive: $\forall w\; Rww$
4	$\Box\varphi \to \Box\Box\varphi$	Transitive: $\forall u, v, w\; (Ruv \wedge Rvw \to Ruw)$
B	$\varphi \to \Box\Diamond\varphi$	Symmetric: $\forall v, w\; (Rvw \to Rwv)$
5	$\Diamond\varphi \to \Box\Diamond\varphi$	Euclidean: $\forall u, v, w\; (Ruv \wedge Ruw \to Rvw)$

For each correspondence, we prove two directions:

($\Leftarrow$) If the frame has the property, then the formula is valid in the frame. (Typically by *reductio*: assume a world falsifies the formula and derive a contradiction using the frame property.)

($\Rightarrow$) If the formula is valid in the frame, then the frame has the property. (Typically by *contrapositive*: assume the frame lacks the property, and construct a counter-model.)

After the model-theoretic proof, we give a proof-theoretic derivation of the axiom in the appropriate extension of K, using the labelled natural deduction system presented in the previous session.

5 D Defines Serial Frames

Theorem 5.1. *For every frame $\mathcal{F}$,*

$$\mathcal{F} \models \Box\varphi \to \Diamond\varphi \quad \Longleftrightarrow \quad \mathcal{F} \text{ is serial.}$$

5.1 Model-theoretic proof

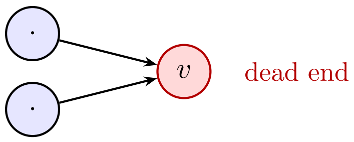
Direction ($\Leftarrow$): Seriality implies validity of D. Suppose $\mathcal{F}$ is serial and, seeking a contradiction, suppose there is a model $\mathcal{M} = \langle W, R, I\rangle$ on $\mathcal{F}$ and a world $w \in W$ such that $\mathcal{M}, w \not\models \Box\varphi \to \Diamond\varphi$. Then:

1. $\mathcal{M}, w \models \Box\varphi$ (the antecedent is true),
2. $\mathcal{M}, w \not\models \Diamond\varphi$ (the consequent is false).

Since R is serial, there exists w' with Rww'. By (1), $\mathcal{M}, w' \models \varphi$. But (2) means that for every v with Rwv, $\mathcal{M}, v \models \neg\varphi$; in particular $\mathcal{M}, w' \models \neg\varphi$. Contradiction. □

Direction ($\Rightarrow$): Validity of D implies seriality. We prove the contrapositive: if $\mathcal{F}$ is *not* serial, then $\mathcal{F} \not\models \Box\varphi \to \Diamond\varphi$.

If $\mathcal{F}$ is not serial, then there exists a "dead-end" world $v \in W$ with no successor: $\forall w \in W,\ \neg Rvw$.



Since no world is accessible from v, we have vacuously $\mathcal{M}, v \models \Box\varphi$ (for any φ and any model $\mathcal{M}$ on $\mathcal{F}$). We also have $\mathcal{M}, v \not\models \Diamond\varphi$ (since there is no successor at all). Therefore $\mathcal{M}, v \not\models \Box\varphi \to \Diamond\varphi$, so $\mathcal{F} \not\models D$. □

Remark 5.1. A dead-end world vacuously satisfies $\Box\varphi$ ("everything is necessary") and falsifies $\Diamond\varphi$ ("nothing is possible"). Axiom D rules out such worlds by requiring seriality.

5.2 Proof-theoretic derivation

In system D, one adds to K a **seriality rule**: from nothing, assert wRv for a fresh v. We can then derive axiom D:

1	$\quad w : \Box\varphi$	H
2	$\quad wRv$ (fresh v, by seriality)	
3	$\quad v : \varphi$	$\Box$E, 1, 2
4	$\quad w : \Diamond\varphi$	$\Diamond$I, 2, 3
5	$w : \Box\varphi \to \Diamond\varphi$	$\to$I, 1—4

Explanation: The key step is line 2: in pure K, we cannot assert the existence of a successor from nothing. The seriality rule (corresponding to $\forall w\, \exists v\ Rwv$) is exactly what is needed. Once we have a successor v, $\Box$E gives φ at v (line 3), and $\Diamond$I gives $\Diamond\varphi$ at w (line 4).

Remark 5.2 (The proof-theoretic derivation only gives one direction). Notice that the natural-deduction derivation above establishes the ($\Leftarrow$) direction of the correspondence: *if the frame is serial, then axiom D is derivable* (and therefore valid, by soundness). We will see the same pattern for every correspondence result: the structural rule for the frame property is added to K, and the axiom is derived.

But the ($\Rightarrow$) direction — *if the axiom is valid in a frame, then the frame has the property* — **cannot be obtained proof-theoretically**. Why not? Because it is a claim about what the validity of a formula *forces* on the structure of R. Its proof works by *contrapositive*: assume R lacks the property, then *construct a specific valuation I* that falsifies the formula. That construction — choosing a particular interpretation that exploits the structural defect of R —

is an inherently *semantic* operation. A proof system manipulates labelled formulae; it does not reason about which valuations exist on a frame.

In short:

- The ND derivation shows that the structural rule is **sufficient** to derive the axiom. $(\Leftarrow)$
- The model-theoretic argument shows that the structural property is **necessary** for the axiom to be valid. $(\Rightarrow)$

The two halves use genuinely different methods, and this asymmetry is not an accident: it reflects a deep fact about the relationship between syntax and semantics.

6 M (T) Defines Reflexive Frames

Theorem 6.1. *For every frame $\mathcal{F}$,*

$$\mathcal{F} \models \Box\varphi \to \varphi \quad \iff \quad \mathcal{F} \text{ is reflexive.}$$

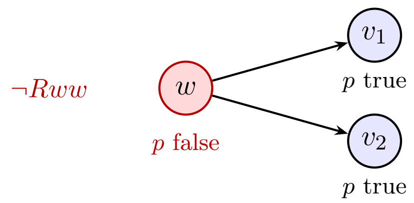
6.1 Model-theoretic proof

Direction $(\Leftarrow)$: Reflexivity implies validity of M. Suppose $\mathcal{F}$ is reflexive. Let $\mathcal{M}$ be any model on $\mathcal{F}$ and $w \in W$. Suppose $\mathcal{M}, w \models \Box\varphi$. Since R is reflexive, Rww. By the semantics of $\Box$, $\mathcal{M}, w \models \varphi$. Therefore $\mathcal{M}, w \models \Box\varphi \to \varphi$. $\square$

Direction $(\Rightarrow)$: Validity of M implies reflexivity. Contrapositive: assume $\mathcal{F}$ is not reflexive. Then $\exists w \in W$, $\neg Rww$.

We construct a counter-model. Choose an interpretation I such that:

- $I(p, w) = F$;
- for every $i \neq w$ with Rwi, $\quad I(p, i) = T$.



Then $\mathcal{M}, w \models \Box p$ (since p is true at every world accessible from w—and w itself is *not* accessible from w). But $\mathcal{M}, w \not\models p$. Therefore $\mathcal{M}, w \not\models \Box p \to p$, and $\mathcal{F} \not\models M$. $\square$

6.2 Proof-theoretic derivation

In system T ($= K +$ reflexivity), we add a **reflexivity rule**: at any world w, assert wRw. This gives us an explicit accessibility link from w to itself, which we can then feed into the standard $\Box$E rule.

Derivation of axiom T:

1	$\quad w : \Box\varphi$	H
2	$\quad wRw$	Refl.
3	$\quad w : \varphi$	$\Box$E, 1, 2
4	$w : \Box\varphi \to \varphi$	$\to$I, 1—3

Explanation: The key step is line 2, where the reflexivity of R gives us wRw. Once we have this accessibility link, ordinary $\Box$E (line 3) extracts φ from $\Box\varphi$ at the same world. In K, line 2 is unavailable: we have no way to assert that a world sees itself, so $\Box\varphi$ at w tells us nothing about what holds at w.

A useful theorem of T: $\varphi \to \Diamond\varphi$. In T, whatever is actual is possible.

1	$w : \varphi$	H
2	$w : \neg\Diamond\varphi$	H
3	$w : \Box\neg\varphi$	Def. $\Diamond$, 2
4	wRw	Refl.
5	$w : \neg\varphi$	$\Box$E, 3, 4
6	$w : \bot$	$\neg$E, 1, 5
7	$w : \neg\neg\Diamond\varphi$	$\neg$I, 2—6
8	$w : \Diamond\varphi$	$\neg\neg$E, 7
9	$w : \varphi \to \Diamond\varphi$	$\to$I, 1—8

Explanation: The key step is line 4: reflexivity gives wRw, which lets $\Box$E extract $\neg\varphi$ from $\Box\neg\varphi$ at the same world (line 5). This contradicts the assumption φ. Without reflexivity, line 4 is unavailable and the proof breaks down.

7 Axiom 4 Defines Transitive Frames ($S4$)

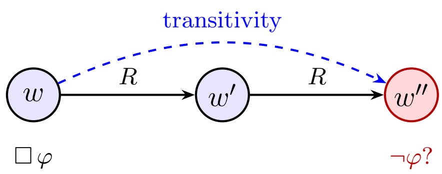
Theorem 7.1. *For every frame $\mathcal{F}$,*

$$\mathcal{F} \models \Box\varphi \to \Box\Box\varphi \quad \Longleftrightarrow \quad \mathcal{F} \textit{ is transitive.}$$

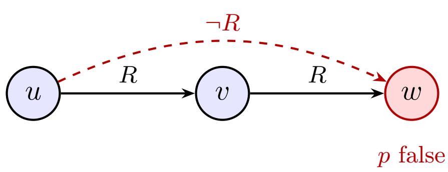
7.1 Model-theoretic proof

Direction ($\Leftarrow$): Transitivity implies validity of axiom 4. Suppose $\mathcal{F}$ is transitive. Let $\mathcal{M}$ be any model on $\mathcal{F}$, let $w \in W$, and suppose $\mathcal{M}, w \models \Box\varphi$. We need to show $\mathcal{M}, w \models \Box\Box\varphi$.

By *reductio*: suppose $\mathcal{M}, w \not\models \Box\Box\varphi$, i.e. $\mathcal{M}, w \models \Diamond\Diamond\neg\varphi$. Then there exist $w', w'' \in W$ with Rww', $Rw'w''$, and $\mathcal{M}, w'' \models \neg\varphi$. By transitivity, Rww''. But $\mathcal{M}, w \models \Box\varphi$ together with Rww'' gives $\mathcal{M}, w'' \models \varphi$. Contradiction. $\Box$



Direction ($\Rightarrow$): Validity of axiom 4 implies transitivity. Contrapositive: suppose R is not transitive. Then there exist $u, v, w \in W$ with Ruv, Rvw, but $\neg Ruw$.



Define I so that $I(p, w) = F$ and for every $x \neq w$ with Rux, $I(p, x) = T$. Then $\mathcal{M}, u \models \Box p$ (since every successor of u satisfies p), but $\mathcal{M}, u \models \Diamond\Diamond\neg p$ (via v and then w). So $\mathcal{M}, u \not\models \Box p \to \Box\Box p$, and $\mathcal{F} \not\models$ axiom 4. $\Box$

7.2 Proof-theoretic derivation

$S4 = T +$ transitivity rule: from wRv and vRu, conclude wRu.

Derivation of $\Box p \vdash_{S4} \Box\Box p$:

1	$w : \Box p$	P
2	$\quad wRv$	fresh v
3	$\quad\quad vRu$	fresh u
4	$\quad\quad wRu$	Trans., 2, 3
5	$\quad\quad u : p$	$\Box$E, 1, 4
6	$\quad v : \Box p$	$\Box$I, 3—5
7	$w : \Box\Box p$	$\Box$I, 2—6

Explanation: The outer sub-proof (lines 2–6) introduces a fresh v accessible from w; the inner sub-proof (lines 3–5) introduces a fresh u accessible from v. The **transitivity rule** (line 4) gives wRu, which lets us apply $\Box$E to the premise $\Box p$ at w and obtain p at u. Closing the inner sub-proof gives $v : \Box p$; closing the outer one gives $w : \Box\Box p$.

Remark 7.1. Without transitivity, line 4 is unavailable: we have wRv and vRu but not wRu, so $\Box$E cannot extract p at u from $\Box p$ at w. This is precisely why $\Box p \to \Box\Box p$ fails in K.

Converse in $S4$: $\Box\Box p \vdash_{S4} \Box p$. This direction uses reflexivity (from T, which is included in $S4$):

1	$w : \Box\Box p$	P
2	$\quad wRv$	fresh v
3	$\quad v : \Box p$	$\Box$E, 1, 2
4	$\quad vRv$	Refl.
5	$\quad v : p$	$\Box$E, 3, 4
6	$w : \Box p$	$\Box$I, 2—5

Explanation: At line 3, ordinary $\Box$E brings $\Box p$ down to v. Then reflexivity gives vRv (line 4), and a second $\Box$E (line 5) strips the remaining box at the same world v.

8 Axiom B Defines Symmetric Frames

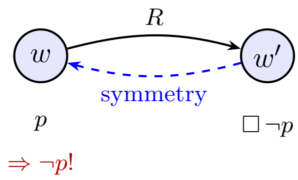
Theorem 8.1. *For every frame $\mathcal{F}$,*

$$\mathcal{F} \models \varphi \to \Box\Diamond\varphi \quad \Longleftrightarrow \quad \mathcal{F} \text{ is symmetric.}$$

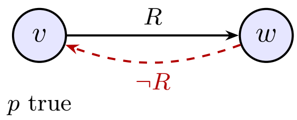
8.1 Model-theoretic proof

Direction ($\Leftarrow$): Symmetry implies validity of B. Suppose $\mathcal{F}$ is symmetric. Let $\mathcal{M}$ be any model on $\mathcal{F}$ and $w \in W$. Suppose $\mathcal{M}, w \models p$. We want to show $\mathcal{M}, w \models \Box\Diamond p$.

By *reductio*: suppose $\mathcal{M}, w \not\models \Box\Diamond p$. Then $\mathcal{M}, w \models \Diamond\Box\neg p$, so there exists w' with Rww' and $\mathcal{M}, w' \models \Box\neg p$. By symmetry, $Rw'w$. Therefore $\mathcal{M}, w \models \neg p$. Contradiction with $\mathcal{M}, w \models p$. $\Box$



Direction (⇒): Validity of B implies symmetry. Contrapositive: suppose R is not symmetric. Then $\exists v, w \in W\ (Rvw \wedge \neg Rwv)$.



Define I such that $I(p, v) = T$ and for every index x accessible from w, $I(p, x) = F$. (This is possible since v is *not* accessible from w.) Then $\mathcal{M}, v \models p$, but $\mathcal{M}, w \models \Box\neg p$ (all successors of w make p false), so $\mathcal{M}, v \models \Diamond\Box\neg p$, i.e. $\mathcal{M}, v \not\models \Box\Diamond p$. Therefore $\mathcal{F} \not\models B$. □

8.2 Proof-theoretic derivation

$B = T +$ symmetry rule: from wRv, conclude vRw.

Derivation of $p \vdash_B \Box\Diamond p$:

1	$w : p$	P
2	$\quad wRv$	fresh v
3	$\quad vRw$	Sym., 2
4	$\quad v : \Diamond p$	$\Diamond$I, 3, 1
5	$w : \Box\Diamond p$	$\Box$I, 2—4

Explanation: The crucial step is line 3, where symmetry gives vRw from wRv. We can then use $\Diamond$I at line 4: since v sees w (line 3) and p holds at w (line 1), we conclude $v : \Diamond p$.

Remark 8.1 (On the use of line 1 inside the sub-proof). At line 4, we use $w : p$ *inside* the $\Box$I sub-proof. Is this not a violation of the rule against importing unboxed formulae? No: $w : p$ is a premise with label w, which already exists *outside* the sub-proof. The restriction says you cannot take an unboxed formula at w and *relabel* it at v. Here, $w : p$ keeps its own label; we use it in combination with vRw to derive $v : \Diamond p$ via $\Diamond$I.

9 Axiom 5 Defines Euclidean Frames ($S5$)

Theorem 9.1. *For every frame $\mathcal{F}$,*

$$\mathcal{F} \models \Diamond\varphi \to \Box\Diamond\varphi \quad \Longleftrightarrow \quad \mathcal{F} \text{ is euclidean.}$$

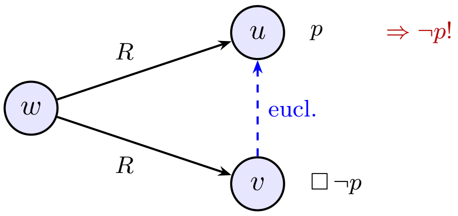
9.1 Model-theoretic proof

Direction (⇐): Euclideanness implies validity of axiom 5. Suppose $\mathcal{F}$ is euclidean. Let $\mathcal{M}$ be a model on $\mathcal{F}$, $w \in W$, and suppose $\mathcal{M}, w \models \Diamond p$. We want $\mathcal{M}, w \models \Box\Diamond p$.

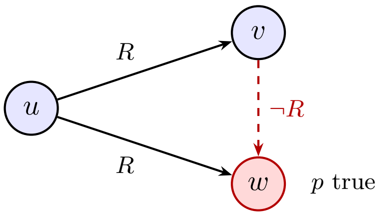
By *reductio*: suppose $\mathcal{M}, w \not\models \Box\Diamond p$, i.e. $\mathcal{M}, w \models \Diamond\Box\neg p$. Then:

- From $\mathcal{M}, w \models \Diamond p$: there exists u with Rwu and $\mathcal{M}, u \models p$.
- From $\mathcal{M}, w \models \Diamond\Box\neg p$: there exists v with Rwv and $\mathcal{M}, v \models \Box\neg p$.

By euclideanness, $Rwu \wedge Rwv \to Rvu$. Therefore $\mathcal{M}, u \models \neg p$. Contradiction. $\square$



Direction ($\Rightarrow$): Validity of axiom 5 implies euclideanness. Contrapositive: suppose R is not euclidean. Then $\exists u, v, w \in W\ (Ruv \wedge Ruw \wedge \neg Rvw)$.



Define I so that $I(p, w) = T$ and for every $x \neq w$ with Rvx, $I(p, x) = F$. (This is consistent since $\neg Rvw$.) Then:

- $\mathcal{M}, u \models \Diamond p$ (via w, where p is true);
- $\mathcal{M}, v \models \square\neg p$ (every successor of v makes p false, and w is not a successor of v);
- hence $\mathcal{M}, u \models \Diamond\square\neg p$, i.e. $\mathcal{M}, u \not\models \square\Diamond p$.

So $\mathcal{M}, u \not\models \Diamond p \to \square\Diamond p$, and $\mathcal{F} \not\models$ axiom 5. $\square$

9.2 Proof-theoretic derivation

$S5 = T +$ euclideanness rule: from wRv and wRu, conclude vRu.

Derivation of $\Diamond p \vdash_{S5} \square\Diamond p$:

1	$w : \Diamond p$	P
2	$\quad wRv,\ \ v : p$	fresh v
3	$\quad\quad wRu$	fresh u
4	$\quad\quad uRv$	Eucl., 3, 2
5	$\quad\quad u : \Diamond p$	$\Diamond$I, 4, 2
6	$\quad w : \square\Diamond p$	$\square$I, 3—5
7	$w : \square\Diamond p$	$\Diamond$E, 1, 2—6

Explanation: The proof nests a $\square$I sub-proof (lines 3–5) inside a $\Diamond$E sub-proof (lines 2–6). The euclideanness rule at line 4 gives uRv: since both v and u are accessible from w (lines 2 and 3), euclideanness yields uRv. Then $\Diamond$I (line 5) uses uRv and $v : p$ to conclude $u : \Diamond p$. Closing the $\square$I sub-proof gives $w : \square\Diamond p$, and this does not mention v, so the $\Diamond$E sub-proof closes validly.

Remark 9.1. Without euclideanness, line 4 is unavailable. We would know wRv and wRu, but nothing about the relationship between v and u—so u might not see any p-world, and $\Diamond$I could not fire.

Remark 9.2 ($S5$ and equivalence relations). In $S5 = T +$ axiom 5, the accessibility relation is reflexive and euclidean. A reflexive euclidean relation is also symmetric and transitive (cf. Exercise 2.1), hence it is an *equivalence relation*. This means:

- Iterated modalities collapse: $\Box\Box\,\varphi \leftrightarrow \Box\,\varphi$ and $\Diamond\Diamond\,\varphi \leftrightarrow \Diamond\,\varphi$.
- Mixed modalities simplify: $\Box\Diamond\,\varphi \leftrightarrow \Diamond\,\varphi$ and $\Diamond\Box\,\varphi \leftrightarrow \Box\,\varphi$.
- Every string of $\Box$'s and $\Diamond$'s reduces to a single $\Box$ or $\Diamond$.

10 Summary of Modal Systems and Frame Conditions

System	Axiom	Frame condition	ND rule added to K
K	$\Box(\varphi \to \psi) \to (\Box\,\varphi \to \Box\,\psi)$	none	(base system)
D	$\Box\,\varphi \to \Diamond\,\varphi$	seriality	seriality rule
T (M)	$\Box\,\varphi \to \varphi$	reflexivity	Refl. rule
$S4$	$\Box\,\varphi \to \Box\Box\,\varphi$	refl. + transitivity	Refl. + Trans. rule
B	$\varphi \to \Box\Diamond\,\varphi$	refl. + symmetry	Refl. + Sym. rule
$S5$	$\Diamond\,\varphi \to \Box\Diamond\,\varphi$	equivalence relation	Refl. + Eucl. rule

Part III
Applications

11 Deontic Logic: The System D

New notation. We introduce a deontic operator O ("it is obligatory that") which satisfies the same rules as $\Box$ in K, plus one additional axiom. The dual operators are:

- $P\varphi =_{\text{Def}} \neg O\neg\varphi$: "it is *permitted* that φ".
- $F\varphi =_{\text{Def}} O\neg\varphi$: "it is *forbidden* that φ".

The accessibility relation is read as: Rwv means "v is a world in which all the norms holding at w are satisfied." The worlds accessible from w are the *deontically ideal* alternatives to w.

Axiom D.

$$O\varphi \to P\varphi \qquad \text{equivalently:} \qquad O\varphi \to \neg O\neg\varphi$$

What is obligatory is permitted. (Remember Kant: "ought implies can.") This corresponds to **seriality**: every world sees at least one deontically ideal world.

Why not axiom T? The schema $O\varphi \to \varphi$ ("what is obligatory is actual") would say that obligations are always fulfilled. This is plainly false. Deontic logic uses D, not T.

Necessitation in deontic logic. Because of the necessitation rule, all logical theorems are obligatory: $\vdash_D O(p \vee \neg p)$, for instance. Is this correct? One might argue: logical truths are trivially fulfilled and cost nothing to obey, so their obligatory status is harmless. Others find it philosophically suspect that logical necessities should count as moral obligations.

12 Tense Logic: A Case Study in Expressive Power and Its Limits

Tense logic provides the richest illustration of modal correspondence theory at work. It is also where the *limits* of modal definability become most visible and most philosophically significant.

In this section, we ask a single question:

Question. The standard temporal ordering ("occurring before") is a *dense, strict linear order without endpoints*. Can tense logic axiomatize all of these properties?

12.1 The operators

In tense logic, we introduce two pairs of modal operators:

- **Future**: $G\varphi$ ("it is always going to be the case that φ") and $F\varphi$ ("it will be the case that φ").
- **Past**: $H\varphi$ ("it has always been the case that φ") and $P\varphi$ ("it has been the case that φ").

As with $\Box$ and $\Diamond$:

$$F\varphi =_{\text{Def}} \neg G \neg \varphi \qquad\qquad P\varphi =_{\text{Def}} \neg H \neg \varphi$$

Here, G and H each have their own accessibility relation: R_F ("strictly later than") and R_P ("strictly earlier than"), where R_P is the converse of R_F. Both operators satisfy the rules of K.

The past and future operators are linked by two **interaction axioms** expressing that the past is the converse of the future:

$$(GP) \qquad \varphi \to GP\varphi \quad \text{"If } \varphi \text{ is true now, it will always have been true."} \tag{1}$$

$$(HF) \qquad \varphi \to HF\varphi \quad \text{"If } \varphi \text{ is true now, it always was going to be true."} \tag{2}$$

12.2 The target: properties of temporal order

What should a temporal ordering look like? The classical answer, going back to Russell and Walker, is that the "occurring before" relation $<$ on the set of instants should satisfy the following properties:

Property	**FOL condition**	**Intuition**
Transitivity	$a < b \wedge b < c \Rightarrow a < c$	temporal order chains compose
Irreflexivity	$\neg(a < a)$	no instant is before itself
Asymmetry	$a < b \Rightarrow \neg(b < a)$	time doesn't run in circles
Seriality	$\forall a\, \exists b\, (a < b)$ and $\forall a\, \exists b\, (b < a)$	no first, no last instant
Connectivity	$a \neq b \Rightarrow (a < b \vee b < a)$	time is a line, not branching
Density	$a < b \Rightarrow \exists c\, (a < c \wedge c < b)$	between any two instants, a third

Together, these properties axiomatize a *dense strict linear order without endpoints*. The standard model is $(\mathbb{Q}, <)$ — the rational numbers with their usual ordering. (Remarkably, Cantor proved that any countable dense linear order without endpoints is isomorphic to $(\mathbb{Q}, <)$.)

Our question becomes: *which of these six properties can be defined by tense-logical formulas, and which cannot?*

12.3 What tense logic can capture

We now go through each definable property, explaining not just the axiom but *why* it works.

Transitivity.

$$G\varphi \to GG\varphi \qquad \text{and} \qquad H\varphi \to HH\varphi$$

These are instances of axiom 4 for G and H respectively. The argument is exactly the one from Section 7: if φ holds at every future instant, and the future-of-the-future is still in the future (by transitivity), then φ holds at every future-of-the-future instant.

Seriality: no last instant, no first instant.

$$G\varphi \to F\varphi \qquad \text{(no last instant)} \qquad\qquad H\varphi \to P\varphi \qquad \text{(no first instant)}$$

These are instances of axiom D for G and H respectively. "If φ will always be the case, then φ will be the case" — this presupposes that the future is non-empty, i.e. every instant has a successor. Similarly for the past.

Connectivity (linearity).

$$(PF\varphi \vee FP\varphi) \to (P\varphi \vee \varphi \vee F\varphi)$$

This axiom forces the ordering to be linear: time is a single thread, not a branching tree. The left side says "at some past instant, φ will hold in the future" or "at some future instant, φ held in the past" — i.e., φ holds at *some* instant reachable by a zig-zag through the ordering. If the ordering is linear, any such instant must be in the past, present, or future of the current instant.

Density.

$$Fp \to FFp \qquad \text{and} \qquad Pp \to PPp$$

If something will be the case, then it will be that it will be the case — there must be an intermediate instant between now and the instant at which p holds. This forces density: between any two instants there is a third. (The reasoning is: if Fp holds at a via some $b > a$ with $b \models p$, then FFp requires some c with $a < c$ and Fp holding at c, i.e. $c < b$. So $a < c < b$.)

Taking stock. At this point, tense logic has captured a remarkable amount of structure. The axiom system consisting of:

K axioms for G and H	(basic tense logic)
GP and HF	(past = converse of future)
$G\varphi \to GG\varphi$, $\quad H\varphi \to HH\varphi$	(transitivity)
$G\varphi \to F\varphi$, $\quad H\varphi \to P\varphi$	(no endpoints)
$(PF\varphi \vee FP\varphi) \to (P\varphi \vee \varphi \vee F\varphi)$	(linearity)
$Fp \to FFp$, $\quad Pp \to PPp$	(density)

characterizes exactly the class of *dense linear orders without endpoints*. For a purely *propositional* language with just two pairs of modal operators, this is a striking degree of expressive power. It captures the topology of $(\mathbb{Q}, <)$.

12.4 What tense logic cannot capture

Two properties remain on our checklist: **irreflexivity** and **asymmetry**. These are precisely the properties that make the temporal order *strict* ("strictly before" as opposed to "before or simultaneous with").

Irreflexivity.

Theorem (Goldblatt–Thomason). Irreflexivity ($\forall w\, \neg Rww$) **cannot be defined by any modal formula**.

This means: there is no tense-logical formula φ such that $\mathcal{F} \models \varphi$ iff $\mathcal{F}$ is irreflexive. No matter how cleverly we combine G, H, F, P, and the propositional connectives, we cannot express "no instant is before itself."

Asymmetry. Similarly, asymmetry ($\forall v \forall w\,(Rvw \to \neg Rwv)$) is not modally definable.

Why? The deep reason is that modal formulas are *bisimulation-invariant*: if two models are bisimilar, they satisfy the same modal formulas. But irreflexivity is not preserved by bisimulations. More precisely, the Goldblatt–Thomason theorem states that a first-order property of frames is definable by a set of modal formulas if and only if the corresponding class of frames is closed under generated subframes, disjoint unions, and *ultrafilter extensions* (and its complement is closed under ultrapowers). The class of irreflexive frames fails the ultrafilter-extension condition.

Irreflexive: $\neg Rww$

No modal formula defines this!

12.5 The philosophical moral

The situation is both impressive and frustrating:

- Tense logic can capture the **topology** of time: dense, linear, no endpoints, transitive.
- But it **cannot capture the strictness** of temporal precedence: irreflexivity and asymmetry escape modal definability.

This means that tense logic *cannot distinguish a strict order ($<$) from a non-strict order ($\leq$).* A dense linear preorder without endpoints (where instants can be "simultaneous" with themselves) validates exactly the same tense-logical formulas as a dense strict linear order without endpoints. From the perspective of tense logic, the two are indistinguishable.

The very property that makes temporal ordering *temporal* — that no instant precedes itself — is precisely what escapes the expressive power of modal logic. This is arguably the most concrete and philosophically significant illustration of the limits of modal definability. It also motivates the development of richer formalisms — such as *hybrid logic* (which adds nominals naming individual instants) or *first-order modal logic* — where irreflexivity becomes expressible.

12.6 Circular time and the eternal recurrence

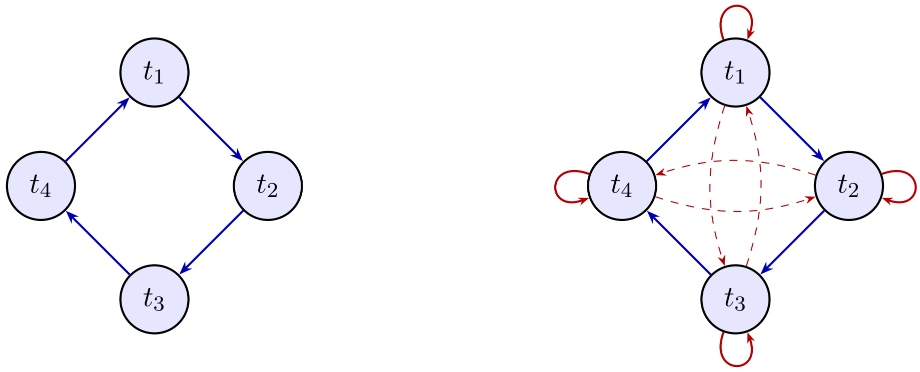
The failure of modal definability for irreflexivity and asymmetry has a striking philosophical consequence. In this section we make it concrete by asking a single question:

> **Question.** Can tense logic exclude *circular time* — a temporal structure in which the future eventually loops back to the past?

The answer is no, and the argument is both technically clean and philosophically instructive.

12.6.1 The circular model

Consider a set of instants $\{t_1, t_2, t_3, t_4\}$ arranged in a cycle: t_1 precedes t_2, which precedes t_3, which precedes t_4, which precedes t_1 again. We take the accessibility relation R to be the "strictly later than" relation, interpreted cyclically.



(a) The successor relation *(b) The transitive closure: a complete graph*

What transitivity does. Diagram (a) shows the bare cycle: t_1Rt_2, t_2Rt_3, t_3Rt_4, t_4Rt_1. Now recall that we have established transitivity as a tense-logical axiom ($G\varphi \to GG\varphi$). Taking the transitive closure of the cycle produces new edges: from t_1Rt_2 and t_2Rt_3 we get t_1Rt_3; from t_1Rt_3 and t_3Rt_4 we get t_1Rt_4; and from t_1Rt_4 and t_4Rt_1 we get t_1Rt_1. Repeating for every starting point, the transitive closure is the *complete relation*: every instant sees every instant, including itself. Diagram (b) shows the result. The self-loops (in red) are the critical point: transitivity has forced reflexivity.

Key observation. In any transitive frame, cycles force reflexivity: if $t_1Rt_2R\cdots Rt_nRt_1$, then by iterated transitivity t_1Rt_1. A transitive cycle of length n collapses into a single $S5$ equivalence class — a complete graph in which every world sees every other.

The Nietzschean structure. The resulting frame is an equivalence class: reflexive, symmetric, and transitive. It is therefore an $S5$ frame. In this frame, every instant has every other instant in both its future and its past. This is precisely the structure of Nietzsche's *eternal recurrence*: "This life as you now live it and have lived it, you will have to live once more and innumerable times more; and there will be nothing new in it, but every pain and every joy ... must return to you, all in the same succession and sequence." (*The Gay Science*, §341.)

12.6.2 Why no tense-logical axiom can exclude circular time

We now explain, in three layers of increasing depth, why no formula of propositional tense logic can distinguish linear time from circular time.

Layer 1: the direct argument from non-definability. The argument is immediate from what we have already established. To exclude cycles, we would need to enforce **asymmetry**: $aRb \to \neg(bRa)$. But asymmetry is not modally definable (Section 12.4). Therefore no tense-logical formula can rule out circular temporal structures. Equivalently, we would need to enforce **irreflexivity**: $\forall w\, \neg Rww$ (and we have just seen that in the presence of transitivity, cycles entail reflexivity). But irreflexivity is not modally definable either.

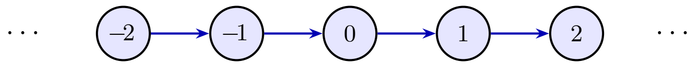
This argument is quick but appeals to the Goldblatt–Thomason theorem as a black box. The next layer opens it.

Layer 2: the bisimulation argument. The reason modal formulas cannot distinguish linear from circular time is that they are **bisimulation-invariant**: if two models are bisimilar (i.e. there is a bisimulation linking them), they satisfy exactly the same modal formulas.

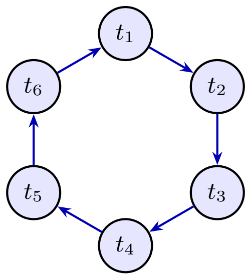
The key insight is that modal logic is fundamentally *local*: from any given instant, it can explore the accessibility relation outward — what is accessible, what is accessible from what is accessible, and so on — but it can only explore finitely many steps in any single formula. A formula of modal depth n can "see" at most n steps ahead or behind.

Now consider two frames:

- $\mathcal{F}_{\text{lin}} = (\mathbb{Z}, <)$: the integers with their usual strict ordering (a linear, transitive, serial frame).
- $\mathcal{F}^n_{\text{circ}}$: a cycle of length n with the transitive closure taken (a complete graph on n nodes).



$\mathcal{F}_{\text{lin}}$: *linear time* $(\mathbb{Z}, <)$



$\mathcal{F}^6_{\text{circ}}$: *circular time*

For any fixed modal depth k, if we choose n large enough (specifically $n > k$), then a world in the circle cannot "tell" that it is on a circle: looking k steps forward and k steps backward, it sees a chain that looks exactly like a segment of $\mathbb{Z}$. More precisely, for any formula φ of modal depth at most k, and any model $\mathcal{M}_{\text{circ}}$ on $\mathcal{F}^n_{\text{circ}}$ with $n > k$, one can build a model $\mathcal{M}_{\text{lin}}$ on $\mathcal{F}_{\text{lin}}$ such that $\mathcal{M}_{\text{circ}}, t_i \models \varphi$ iff $\mathcal{M}_{\text{lin}}, m \models \varphi$ for a suitable m. The two models are *k-bisimilar*: they are indistinguishable by any formula of depth $\leq k$. Since k was arbitrary, no single formula (which always has a finite depth) can separate the two.

Remark 12.1. The argument works even more directly at the level of frames. The frame $\mathcal{F}^n_{\text{circ}}$ (with transitive closure) is a complete graph on n nodes — a single $S5$ equivalence class. The frame $(\mathbb{Z}, <)$ is also serial, transitive, dense (if we pass to $\mathbb{Q}$), and linear. Yet no tense-logical formula can distinguish them, because every *local* pattern of accessibility that appears in one also appears in the other, provided n is large enough.

Layer 3: the philosophical diagnosis. The root cause is now clear: modal logic is a *local* formalism. From any given instant, a modal formula can inspect the structure of accessibility within a bounded neighbourhood. It asks questions of the form "is there an accessible world where . . . ?" and "do all accessible worlds satisfy . . . ?", iterated finitely many times. But the question "does the chain of accessible worlds eventually loop back to where I started?" is a *global* property of the frame. It requires tracing a path through the entire graph and checking whether it returns to its origin. No finite nesting of G, H, F, P, and propositional connectives can do this.

Irreflexivity ($\forall w\, \neg Rww$) and asymmetry ($Rwv \to \neg Rvw$) are both *negative universal* conditions on the accessibility relation: they say that certain edges *do not exist*. Modal formulas, by contrast, are good at asserting the *existence* of accessible worlds (via $\Diamond$, F, P) and propagating information along *existing* edges (via $\Box$, G, H). They have no mechanism for asserting the *absence* of an edge that would complete a cycle.

The moral. The difference between *Aristotelian time* — a strict linear order flowing irreversibly from past to future — and *Nietzschean time* — an eternal recurrence in which every moment eventually returns — is real, but **modally invisible**. Both structures validate exactly the same tense-logical formulas (given a sufficiently large cycle). Propositional modal logic lacks the resources to express what makes time *directional.*

12.6.3 What would be needed to exclude circular time

The limitation is not that the distinction between linear and circular time is inexpressible in principle, but that *propositional* modal language lacks the resources to formulate it. Richer formalisms can:

Hybrid logic. Hybrid logic extends modal logic with *nominals* — propositional symbols that are true at exactly one world — and a *satisfaction operator* $@_i\varphi$ ("φ holds at the world named i"). In hybrid tense logic, irreflexivity is expressible:

$$@_i \neg Fi$$

reads: "at instant i, it is not the case that i is in the future of i." This directly asserts that i does not precede itself. Since i names a specific instant, the formula can "look back" and check whether the path has returned.

First-order modal logic. With quantification over instants, one can simply write:

$$\forall x\, \neg(x < x)$$

which is a straightforward first-order sentence, not a modal formula.

In both cases, the key is the ability to *name* or *quantify over* individual points in the frame and to compare a world with itself. Propositional modal logic, which treats worlds as anonymous evaluation points, cannot do this — and that is precisely why circular time escapes its reach.

13 Logics for Belief and Knowledge

Belief operators. We introduce a family of modal operators B_i for each agent i: $B_i\varphi$ is read "agent i believes that φ." Each B_i satisfies the rules of K; in particular, the necessitation rule applies.

Logical omniscience. Since necessitation holds, if φ is a logical truth, then $B_i\varphi$ for every agent i: every agent believes every logical truth. This is the **principle of logical omniscience**. It is descriptively unrealistic (human agents are not logically omniscient), but may be defensible as a normative idealisation (a rational agent *should* believe logical truths).

13.1 Axiom D for Beliefs

$$B_i\varphi \to \neg B_i\neg\varphi$$

If an agent accepts a proposition, the agent does not also accept its negation. This is a *consistency* requirement on beliefs. It corresponds to seriality: for every doxastic state, there is at least one world compatible with the agent's beliefs.

13.2 Axiom T/M and Epistemic Logic

$$B_i\varphi \to \varphi \qquad ???$$

This principle is *not* appropriate for belief: as Plato famously remarked in the *Gorgias*, beliefs can be false. But it does hold for **knowledge**, because knowledge is a *factive* concept: if an agent knows that φ, then φ is true. Writing K_i for "agent i knows that":

$$K_i\varphi \to \varphi$$

13.3 Positive Introspection

$$K_i\varphi \to K_iK_i\varphi \qquad\qquad B_i\varphi \to B_iB_i\varphi$$

If an agent knows (or believes) something, the agent knows (or believes) that they know (or believe) it. This corresponds to axiom 4 (transitivity).

13.4 Negative Introspection

$$\neg K_i\varphi \to K_i\neg K_i\varphi \qquad\qquad \neg B_i\varphi \to B_i\neg B_i\varphi$$

If an agent does not know (or believe) something, the agent knows (or believes) that they do not. This corresponds to axiom 5 (euclideanness). Adding both positive and negative introspection to epistemic logic yields the system $S5$ for knowledge—the standard framework for epistemic logic in game theory, computer science, and formal epistemology.

Part IV
Conclusion: Why Modal Logic?

These lecture notes have traced an arc from the semantics of the minimal modal logic K, through its proof theory, to the correspondence between modal axioms and frame properties, and finally to the limits of modal definability. The tense-logic case study, in particular, ended on a striking negative result: propositional modal logic cannot express irreflexivity or asymmetry, and therefore cannot distinguish linear time from circular time. This raises a natural and important question, which deserves to be addressed explicitly.

Question. Since first-order logic can express everything modal logic can express and more, why is modal logic still considered philosophically relevant? Why not simply use FOL?

The answer has several layers. Taken together, they show that modal logic is not a deficient version of first-order logic, but a self-standing formalism with its own identity, its own virtues, and its own way of generating philosophical insight.

14 Modal Logic as a Mathematical Object: The van Benthem Theorem

The most precise answer to the question "what *is* modal logic?" comes from a theorem due to Johan van Benthem.

Theorem 14.1 (van Benthem, 1976)**.** *A first-order formula (with one free variable) is equivalent to the standard translation of a modal formula if and only if it is invariant under* ***bisimulations****.*

In other words, modal logic is not an arbitrary fragment of FOL: it is the *bisimulation-invariant fragment*. Bisimulations are structure-preserving maps between Kripke models that match worlds with the same local accessibility patterns. Two models that are bisimilar offer the "same view" from each world: whatever is accessible from here, whatever is accessible from there, and so on, to any finite depth.

What the theorem says is this: *modal logic is exactly the part of FOL that cannot tell bisimilar models apart.* Its limitations are not defects or historical accidents — they are the **defining characteristic** of the formalism. Modal logic is the logic of *structural indistinguishability up to bisimulation.*

This gives modal logic a clean mathematical identity, independent of first-order logic. It is not "FOL with some resources removed"; it is the largest fragment of FOL that respects a natural and well-motivated equivalence relation on models.

15 Reasoning from a Standpoint

Why would bisimulation invariance matter philosophically? Because it captures a specific *mode of reasoning*: reasoning **from a standpoint**.

A modal formula evaluates the world from the perspective of a particular point: what is accessible from here, what holds at those accessible points, what is accessible from *those*, and so on. It never steps outside this perspective to ask "am I the same point I started from?" or "how many points are there in total?" This is precisely what bisimulation invariance means: if two models offer the same perspective from each point, they are modally indistinguishable.

This turns out to be the *right* level of abstraction for many philosophical questions:

- When we ask "is it possible that it rains tomorrow?", we are reasoning from our current standpoint about what is accessible. We are not naming individual possible worlds or counting them.
- When we ask "does the agent know that p?", we are asking whether p holds at every world compatible with the agent's information — not whether the epistemic accessibility relation has particular graph-theoretic properties.
- When we reason about obligation ("it ought to be that q"), we are quantifying over deontically ideal alternatives to the current situation — not naming them individually.

The **anonymity of worlds** in modal logic is a *feature*, not a limitation. It reflects the fact that what matters about possible worlds, for most philosophical purposes, is their relational structure, not their individual identities.

16 Decidability: A Logical Virtue

Propositional modal logics — K, T, $S4$, $S5$, and many others — are **decidable**: there exists an algorithm that determines, for any formula, whether it is valid. First-order logic, by Church's theorem (1936), is *not*: there is no algorithm that decides, for an arbitrary first-order sentence, whether it is valid.

This is not a minor technical point. It means that modal logic can serve as a framework in which philosophical questions about validity can be *settled mechanically*, at least in principle. The tableau method we studied is one such decision procedure for K: it always terminates, and it either produces a proof (if the formula is valid) or a finite counter-model (if it is not).

The weaker expressivity of modal logic is exactly what buys this. By restricting what can be said, modal logic ensures that the question "does this follow from that?" always has a computable answer. This trade-off — less expressivity in exchange for decidability — is one of the central themes of mathematical logic, and modal logic occupies a particularly elegant point in the space of possible trade-offs: it is expressive enough to formalise substantial philosophical reasoning, yet simple enough to remain decidable.

17 The Limitations Are Themselves Informative

Perhaps the deepest reason for the philosophical relevance of modal logic is that its *limitations generate genuine insights.*

The fact that tense logic cannot capture irreflexivity (Section 12.4) is not just a negative result. It tells us something substantive about the *concept* of temporal reasoning: what makes time flow forward — its strict directedness, the impossibility of returning to the same instant — is not something that can be grasped by reasoning *from within time* about what lies ahead and what lies behind. The directedness of time is a *global structural fact* about the temporal order, not a *perspectival fact* accessible from any particular moment.

This is a genuine philosophical insight, and it emerged *precisely from the limitations of modal logic.* If we had simply used FOL from the start, we would have written $\forall x\, \neg(x < x)$ and moved on, without ever realising that this simple sentence expresses something fundamentally beyond the reach of perspectival temporal reasoning.

More generally, correspondence theory — the systematic study of which modal axioms correspond to which frame properties — would not exist without modal logic. It is a theory about *what can be said* in a restricted language, and its richness depends on the restriction. Consider what we have established in these notes:

- *Reflexivity* corresponds to *factivity*: what is necessary is actual ($\Box\varphi \to \varphi$). The match between a simple structural property of R and a philosophically fundamental feature of knowledge (factivity) is a non-trivial and illuminating fact.

- *Transitivity* corresponds to *positive introspection*: what is necessary is necessarily necessary ($\Box\varphi \to \Box\Box\varphi$). That the question "can we know that we know?" reduces to a structural property of accessibility is a striking connection.

- *Euclideanness* corresponds to *negative introspection*: what is possible is necessarily possible ($\Diamond\varphi \to \Box\Diamond\varphi$). That the question "can we know what we don't know?" reduces to *a different* structural property of the same relation is equally striking.

- *Irreflexivity* has *no modal correspondent at all.* The very property that makes time *temporal* escapes the reach of modal logic entirely.

FOL can express all of these properties trivially, so there is no interesting correspondence theory for FOL. The philosophical depth of the correspondence results comes precisely from the fact that modal logic is weaker: the match between simple, meaningful axioms and elegant structural properties is a non-trivial discovery, not a tautology.

18 Appropriate Expressiveness

A final, more pragmatic point. When philosophers reason about necessity, knowledge, obligation, or time, they do not typically reason in first-order logic. They reason *modally*: "necessarily p", "the agent knows that q", "it ought to be that r", "it will always be that s." Modal logic formalises this practice directly, at the level at which philosophical arguments are actually conducted.

First-order logic *can* express the same things, by quantifying explicitly over worlds, agents, or times. But the translation is cumbersome and obscures the conceptual structure. Compare:

Modal	First-order translation
$\Box\,\varphi \to \varphi$	$\forall w\,\big[(\forall v\,(Rwv \to \varphi(v))) \to \varphi(w)\big]$
$\Box(\varphi \to \psi) \to (\Box\,\varphi \to \Box\,\psi)$	$\forall w\,\big[(\forall v\,(Rwv \to (\varphi(v) \to \psi(v))))$ $\to ((\forall v\,(Rwv \to \varphi(v))) \to (\forall v\,(Rwv \to \psi(v))))\big]$

The modal formulations are transparent; the first-order translations, while logically equivalent, bury the philosophical content under layers of quantification and variable-binding. Modal logic earns its philosophical relevance not by being maximally expressive, but by being **appropriately expressive** — by capturing exactly the reasoning patterns that matter for the questions at hand, and no more.

Summary. Modal logic is not a poor approximation to first-order logic. It is a self-standing formalism with its own mathematical identity (the bisimulation-invariant fragment of FOL), its own virtues (decidability, perspectival reasoning, a rich correspondence theory), and its own way of generating philosophical insight — including, crucially, insight into its own limitations. The question "why not just use FOL?" has a clear answer: because what modal logic *cannot* say is often as philosophically revealing as what it can.