Part I

Applied Modal Logic

This part contains experimental draft material on some applications of modal logic, such as temporal and epistemic logics.

Chapter 1

Temporal Logics

This chapter covers temporal logics.

1.1 Introduction

aml:tl:int:sec Temporal logics deal with claims about things that will or have been the case. Arthur Prior is credited as the originator of temporal logic, which he called *tense logic*. Our treatment of temporal logic here will largely follow Prior's original modal treatment of introducing temporal operators into the basic framework of propositional logic, which treats claims as generally lacking in tense.

For example, in propositional logic, I might talk about a dog, Beezie, who sometimes sits and sometimes doesn't sit, as dogs are wont to do. It would be contradictory in classical logic to claim that Beezie is sitting and also that Beezie is not sitting. But obviously both can be true, just not at the same time; adding temporal operators to the language can allow us to express that claim relatively easily. The addition of temporal operators also allows us to account for the validity of inferences like the one from "Beezie will get a treat or a ball" to "Beezie will get a treat or Beezie will get a ball."

However, a lot of philosophical issues arise with temporal logic that might lead us to adopt one framework of temporal logic over another. For example, a future contingent is a statement about the future that is neither necessary nor impossible. If we say "Richard will go to the grocery store tomorrow," we are expressing a claim about something that has not yet happened, and whose truth value is contestable. In fact, it is contestable whether that claim can even be *assigned* a truth value in the first place. If we are strict determinists, then perhaps we can be comfortable with the idea that this sentence is in fact true or false, even before the event in question is supposed to take place—it just may be that we do not know its truth value yet. In contrast, we might believe in a genuinely open future, in which the truth values of future contingents are undetermined.

As it turns out, a lot of these commitments about the structure and nature of time are built in to our choices of models and frameworks of temporal logics. For example, we might ask ourselves whether we should construct models in which time is linear, branching or even circular. We might have to make decisions about whether our temporal models will have beginning and end points, and whether time is to be represented using discrete instants or as a continuum.

1.2 Semantics for Temporal Logic

aml:tl:sem:sec

Definition 1.1. The basic language of temporal logic contains

1. The propositional constant for falsity $\bot$.
2. The propositional constant for truth $\top$.
3. A denumerable set of propositional variables: p_0, p_1, p_2, ...
4. The propositional connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional), $\leftrightarrow$ (biconditional).
5. Past operators P and H.
6. Future operators F and G.

Later on, we will discuss the potential addition of other kinds of modal operators.

Definition 1.2. *Formulas* of the temporal language are inductively defined as follows:

1. $\bot$ is an atomic formula.
2. $\top$ is an atomic formula.
3. Every propositional variable p_i is an (atomic) formula.
4. If φ is a formula, then $\neg\varphi$ is a formula.
5. If φ and ψ are formulas, then $(\varphi \wedge \psi)$ is a formula.
6. If φ and ψ are formulas, then $(\varphi \vee \psi)$ is a formula.
7. If φ and ψ are formulas, then $(\varphi \rightarrow \psi)$ is a formula.
8. If φ and ψ are formulas, then $(\varphi \leftrightarrow \psi)$ is a formula.
9. If φ is a formula, then $\mathsf{P}\varphi$, $\mathsf{H}\varphi$, $F\varphi$, $\mathsf{G}\varphi$ are all formulas.
10. Nothing else is a formula.

The semantics of temporal logics are given in terms of relational models, as with other kinds of intensional logics.

Definition 1.3. A *model* for temporal language is a triple $\mathfrak{M} = \langle T, \prec, V \rangle$, where

1. T is a nonempty set, interpreted as points in time.
2. $\prec$ is a binary relation on T.
3. V is a function assigning to each propositional variable p a set $V(p)$ of points in time.

When $t \prec t'$ holds, we say that t *precedes* t'. When $t \in V(p)$ we say p is *true at* t.

For now, you will notice that we do not impose any conditions on our precedence relation $\prec$. This means that at present, there are no restrictions on the structure of our temporal models, so we could have models in which time is linear, branching, circular, or has any structure whatsoever.

Just as with normal modal logic, every temporal model determines which formulas count as true at which points in it. We use the same notation "model $\mathfrak{M}$ makes formula φ true at point t" for the basic notion of relational semantics. The relation is defined inductively and is identical to the normal modal case for all non-modal operators.

aml:tl:sem: defn:tmodels **Definition 1.4.** *Truth of a formula* φ *at* t in a $\mathfrak{M}$, in symbols: $\mathfrak{M}, t \Vdash \varphi$, is defined inductively as follows:

1. $\varphi \equiv \bot$: Never $\mathfrak{M}, t \Vdash \bot$.
2. $\varphi \equiv \top$: Always $\mathfrak{M}, t \Vdash \top$.
3. $\mathfrak{M}, t \Vdash p$ iff $t \in V(p)$
4. $\varphi \equiv \neg\psi$: $\mathfrak{M}, t \Vdash \varphi$ iff $\mathfrak{M}, t \nVdash \psi$.
5. $\varphi \equiv (\psi \land \chi)$: $\mathfrak{M}, t \Vdash \varphi$ iff $\mathfrak{M}, t \Vdash \psi$ and $\mathfrak{M}, t \Vdash \chi$.
6. $\varphi \equiv (\psi \lor \chi)$: $\mathfrak{M}, t \Vdash \varphi$ iff $\mathfrak{M}, t \Vdash \psi$ or $\mathfrak{M}, t \Vdash \chi$ (or both).
7. $\varphi \equiv (\psi \to \chi)$: $\mathfrak{M}, t \Vdash \varphi$ iff $\mathfrak{M}, t \nVdash \psi$ or $\mathfrak{M}, t \Vdash \chi$.
8. $\varphi \equiv (\psi \leftrightarrow \chi)$: $\mathfrak{M}, t \Vdash \varphi$ iff either both $\mathfrak{M}, t \Vdash \psi$ and $\mathfrak{M}, t \Vdash \chi$ or neither $\mathfrak{M}, t \Vdash \psi$ nor $\mathfrak{M}, t \Vdash \chi$.
9. aml:tl:sem: defn:sub:mmodels-p $\varphi \equiv \mathsf{P}\psi$: $\mathfrak{M}, t \Vdash \varphi$ iff $\mathfrak{M}, t' \Vdash \psi$ for some $t' \in T$ with $t' \prec t$
10. aml:tl:sem: defn:sub:mmodels-h $\varphi \equiv \mathsf{H}\psi$: $\mathfrak{M}, t \Vdash \varphi$ iff $\mathfrak{M}, t' \Vdash \psi$ for every $t' \in T$ with $t' \prec t$
11. aml:tl:sem: defn:sub:mmodels-f $\varphi \equiv \mathsf{F}\psi$: $\mathfrak{M}, t \Vdash \varphi$ iff $\mathfrak{M}, t' \Vdash \psi$ for some $t' \in T$ with $t \prec t'$

If $\prec$ *is* ...	*then* ... *is true in* $\mathfrak{M}$:
transitive: $\forall u \forall v \forall w((u \prec v \land v \prec w) \rightarrow u \prec w)$	$\mathsf{FF}p \rightarrow \mathsf{F}p$
linear: $\forall w \forall v(w \prec v \lor w = v \lor v \prec w)$	$(\mathsf{FP}p \lor \mathsf{PF}p) \rightarrow (\mathsf{P}p \lor p \lor \mathsf{F}p)$
dense: $\forall w \forall v(w \prec v \rightarrow \exists u(w \prec u \land u \prec v))$	$\mathsf{F}p \rightarrow \mathsf{FF}p$
unbounded (past): $\forall w \exists v(v \prec w)$	$\mathsf{H}p \rightarrow \mathsf{P}p$
unbounded (future): $\forall w \exists v(w \prec v)$	$\mathsf{G}p \rightarrow \mathsf{F}p$

Table 1.1: Some temporal frame correspondence properties.

aml:tl:acc:
tab:correspondence
aml:tl:sem:
defn:sub:mmodels-g

12. $\varphi \equiv \mathsf{G}\psi$: $\mathfrak{M}, t \Vdash \varphi$ iff $\mathfrak{M}, t' \Vdash \psi$ for every $t' \in T$ with $t \prec t'$

Based on the semantics, you might be able to see that the operators P and H are duals, as well as the operators F and G, such that we could define $\mathsf{H}\varphi$ as $\neg\mathsf{P}\neg\varphi$, and the same with G and F.

1.3 Properties of Temporal Frames

aml:tl:acc:
sec

Given that our temporal models do not impose any conditions on the relation $\prec$, the only one of our familiar axioms that holds in all models is K, or its analogues K_{G} and K_{H}:

$$\mathsf{G}(p \rightarrow q) \rightarrow (\mathsf{G}p \rightarrow \mathsf{G}q) \qquad (K_{\mathsf{G}})$$

$$\mathsf{H}(p \rightarrow q) \rightarrow (\mathsf{H}p \rightarrow \mathsf{H}q) \qquad (K_{\mathsf{H}})$$

However, if we want our models to impose stricter conditions on how time is represented, for instance by ensuring that $\prec$ is a linear order, then we will end up with other validities in our models.

Several of the properties from Table 1.1 might seem like desirable features for a model that is intended to represent time. However, it is worth noting that, even though we can impose whichever conditions we like on the $\prec$ relation, not all conditions correspond to formulas that can be expressed in the language of temporal logic. For example, irreflexivity, or the idea that $\forall w \neg(w \prec w)$, does not have a corresponding formula in temporal logic.

1.4 Additional Operators for Temporal Logic

aml:tl:ext:
sec

In addition to the unary operators for past and future, temporal logics also sometimes include binary operators S and U, intended to symbolize "since" and "until". This means adding S and U into the language of temporal logic and adding the following clause into the definition of a temporal formula:

If φ and ψ are formulas, then $(\mathsf{S}\varphi\psi)$ and $(\mathsf{U}\varphi\psi)$ are both formulas.

The semantics for these operators are then given as follows:

aml:tl:ext: defn:since-until **Definition 1.5.** *Truth of a formula φ at t* in a $\mathfrak{M}$:

aml:tl:ext: defn:sub:mmodels-since
1. $\varphi \equiv \mathsf{S}\psi\chi$: $\mathfrak{M}, t \Vdash \varphi$ iff $\mathfrak{M}, t' \Vdash \psi$ for some $t' \in T$ with $t' \prec t$, and for all s with $t' \prec s \prec t$, $\mathfrak{M}, s \Vdash \chi$

aml:tl:ext: defn:sub:mmodels-until
2. $\varphi \equiv \mathsf{U}\psi\chi$: $\mathfrak{M}, t \Vdash \varphi$ iff $\mathfrak{M}, t' \Vdash \psi$ for some $t' \in T$ with $t \prec t'$, and for all s with $t \prec s \prec t'$, $\mathfrak{M}, s \Vdash \chi$

The intuitive reading of $\mathsf{S}\psi\chi$ is "Since ψ was the case, χ has been the case." And the intuitive reading of $\mathsf{U}\psi\chi$ is "Until ψ will be the case, χ will be the case."

1.5 Possible Histories

aml:tl:poss: sec The relational models of temporal logic that we have been using are extremely flexible, since we do not have to place any restrictions on the accessibility relation. This means that temporal models can branch in the past and in the future, but we might want to consider a more "modal" conception of branching, in which we consider sequences of events as possible histories. This does not necessarily require changing our language, though we might also add our "ordinary" modal operators $\Box$ and $\Diamond$, and we could also consider adding epistemic accessibility relations to represent changes in agents' knowledge over time.

Definition 1.6. A *possible histories model* for the temporal language is a triple $\mathfrak{M} = \langle T, C, V\rangle$, where

1. T is a nonempty set, interpreted as states in time.

2. C is a set of computational paths, or *possible histories* of a system. In other words, C is a set of sequences σ of states s_1, s_2, s_3, ..., where every $s_i \in T$.

3. V is a function assigning to each propositional variable p a set $V(p)$ of points in time.

To make things simpler, we will also generally assume that when a history is in C, then so are all of its suffixes. For example, if s_1, s_2, s_3 is a sequence in C, then so are s_2, s_3 and s_3. Also, when two states s_i and s_j appear in a sequence σ, we say that $s_i \prec_\sigma s_j$ when $i < j$. When $t \in V(p)$ we say p is *true at* t.

The one relevant change is that when we evaluate the truth of a formula at a point in time t in a model $\mathfrak{M}$, we do so relative to a history σ, in which t appears as a state. We do not need to change any of the semantics for propositional variables or for truth-functional connectives, though. All of those are exactly as they were in Definition 1.4, since none of those will make reference to σ. However, we now redefine our future operator F and add our $\Diamond$ operator with respect to these histories.

Definition 1.7. *Truth of a formula φ at t, σ* in $\mathfrak{M}$, in symbols: $\mathfrak{M}, t, \sigma \Vdash \varphi$: aml:tl:poss: defn:phmodels

1. $\varphi \equiv \mathsf{F}\psi$: $\mathfrak{M}, t, \sigma \Vdash \varphi$ iff $\mathfrak{M}, t', \sigma \Vdash \psi$ for some $t' \in T$ such that $t \prec_\sigma t'$. aml:tl:poss: defn:sub:phmodels-f
2. $\varphi \equiv \Diamond\psi$: $\mathfrak{M}, t, \sigma \Vdash \varphi$ iff $\mathfrak{M}, t, \sigma' \Vdash \psi$ for some $\sigma' \in C$ in which t occurs. aml:tl:poss: defn:sub:phmodels-diamond

Other temporal and modal operators can be defined similarly. However, we can now represent claims that combine tense and modality. For example, we might symbolize "p will not occur, but it might have occurred" using the formula $\neg\mathsf{F}p \wedge \Diamond\mathsf{F}p$. This would hold at a point and a history at which p does not become true at a successor state, but there is an alternative history at which p will become true.

Chapter 2

Epistemic Logics

This chapter covers the metatheory of epistemic logics. It is structured in a similar way to Aldo Antonelli's notes on classical basic modal logic, but has been rewritten by Audrey Yap in order to add material on bisimulation and dynamic epistemic logics.

2.1 Introduction

aml:el:int:sec Just as modal logic deals with *modal propositions* and the entailment relations among them, epistemic logic deals with *epistemic propositions* and the entailment relations among them. Rather than interpreting the modal operators as representing possibility and necessity, the unary connectives are interpreted in epistemic or doxastic ways, to model knowledge and belief. For example, we might want to express claims like the following:

1. Richard knows that Calgary is in Alberta.
2. Audrey thinks it is possible that a dog is on the couch.
3. Richard knows that Audrey knows that her class is on Tuesdays.
4. Everyone knows that a year has 12 months.

Contemporary epistemic logic is often traced to Jaako Hintikka's *Knowledge and Belief*, from 1962, and it was written at a time when possible worlds semantics were becoming increasingly more used in logic. In fact, epistemic logics use most of the same semantic tools as other modal logics, but will interpret them differently. The main change is in what we take the *accessibility relation* to represent. In epistemic logics, they represent some form of *epistemic possibility.* We'll see that the epistemic notion that we're modelling will affect the constraints that we want to place on the accessibility relation. And we'll also see what happens to correspondence theory when it is given an epistemic interpretation. You'll notice that the examples above mention two agents: Richard and Audrey, and the relationship between the things that each one knows. The epistemic logics we'll consider will be multi-agent logics, in which such things can be expressed. In contrast, a single-agent epistemic logic would only talk about what one individual knows or believes.

2.2 The Language of Epistemic Logic

aml:el:lan:sec

Definition 2.1. Let G be a set of agent-symbols. The basic language of multi-agent epistemic logic contains

1. The propositional constant for falsity $\bot$.
2. The propositional constant for truth $\top$.
3. A denumerable set of propositional variables: p_0, p_1, p_2, ...
4. The propositional connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional), $\leftrightarrow$ (biconditional).
5. The knowledge operator K_a where $a \in G$.

If we are only concerned with the knowledge of a single agent in our system, we can drop the reference to the set G, and individual agents. In that case, we only have the basic operator K.

Definition 2.2. *Formulas* of the epistemic language are inductively defined as follows:

1. $\bot$ is an atomic formula.
2. $\top$ is an atomic formula.
3. Every propositional variable p_i is an (atomic) formula.
4. If φ is a formula, then $\neg\varphi$ is a formula.
5. If φ and ψ are formulas, then $(\varphi \wedge \psi)$ is a formula.
6. If φ and ψ are formulas, then $(\varphi \vee \psi)$ is a formula.
7. If φ and ψ are formulas, then $(\varphi \rightarrow \psi)$ is a formula.
8. If φ and ψ are formulas, then $(\varphi \leftrightarrow \psi)$ is a formula.
9. If φ is a formula and $a \in G$, then $\mathsf{K}_a\varphi$ is a formula.
10. Nothing else is a formula.

If a formula φ does not contain K_a, we say it is *modal-free*.

Definition 2.3. While the K operator is intended to symbolize individual knowledge, E, often read as "everybody knows," symbolizes group knowledge. Where $G' \subseteq G$, we define $\mathsf{E}_{G'}\varphi$ as an abbreviation for

$$\bigwedge_{b \in G'} \mathsf{K}_b \varphi.$$

We can also define an even stronger sense of knowledge, namely *common knowledge* among a group of agents G. When a piece of information is common knowledge among a group of agents, it means that for every combination of agents in that group, they all know that each other knows that each other knows ... ad infinitum. This is significantly stronger than group knowledge, and it is easy to come up with relational models in which a formula is group knowledge, but not common knowledge. We will use $\mathsf{C}_G\varphi$ to symbolize "it is common knowledge among G that φ."

2.3 Relational Models

aml:el:rel:sec The basic semantic concept for epistemic logics is the same as that of ordinary modal logics. Relational models still consist of a set of worlds, and an assignment that determines which propositional variables count as "true" at which worlds. And if we are only dealing with a single agent, we have a single accessibility relation as usual. However, if we have a multi-agent epistemic logic, then our single accessibility relation becomes a set of accessibility relations, one for each a in our set of agent symbols G.

A *relational model* consists of a set of worlds, which are related by binary accessibility relations—one for each agent—together with an assignment which determines which propositional variables are true at which worlds.

Definition 2.4. A *model* for the multi-agent epistemic language is a triple $\mathfrak{M} = \langle W, R, V \rangle$, where

1. W is a nonempty set of "worlds,"
2. For each $a \in G$, R_a is a binary accessibility relation on W, and
3. V is a function assigning to each propositional variable p a set $V(p)$ of possible worlds.

When $R_a ww'$ holds, we say that w' is *accessible by a from w*. When $w \in V(p)$ we say p is *true at w*.

The mechanics are just like the mechanics for normal modal logic, just with more accessibility relations added in. For a given agent, we will generally interpret their accessibility relation as representing something about their informational states. For example, we often treat $R_a ww'$, as expressing that w' is consistent with a's information at w. Or to put it another way, at w, they cannot tell the difference between world w and world w'.

2.4 Truth at a World

Just as with normal modal logic, every epistemic model determines which formulas count as true at which worlds in it. We use the same notation "model $\mathfrak{M}$ makes formula φ true at world w" for the basic notion of relational semantics. The relation is defined inductively and is identical to the normal modal case for all non-modal operators. aml:el:trw:sec

Definition 2.5. *Truth of a formula φ at w* in a $\mathfrak{M}$, in symbols: $\mathfrak{M}, w \Vdash \varphi$, is defined inductively as follows: aml:el:trw:defn:mmodels

1. $\varphi \equiv \bot$: Never $\mathfrak{M}, w \Vdash \bot$.
2. $\varphi \equiv \top$: Always $\mathfrak{M}, w \Vdash \top$.
3. $\mathfrak{M}, w \Vdash p$ iff $w \in V(p)$
4. $\varphi \equiv \neg\psi$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w \nVdash \psi$.
5. $\varphi \equiv (\psi \wedge \chi)$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w \Vdash \psi$ and $\mathfrak{M}, w \Vdash \chi$.
6. $\varphi \equiv (\psi \vee \chi)$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w \Vdash \psi$ or $\mathfrak{M}, w \Vdash \chi$ (or both).
7. $\varphi \equiv (\psi \to \chi)$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w \nVdash \psi$ or $\mathfrak{M}, w \Vdash \chi$.
8. $\varphi \equiv (\psi \leftrightarrow \chi)$: $\mathfrak{M}, w \Vdash \varphi$ iff either both $\mathfrak{M}, w \Vdash \psi$ and $\mathfrak{M}, w \Vdash \chi$ or neither $\mathfrak{M}, w \Vdash \psi$ nor $\mathfrak{M}, w \Vdash \chi$.
9. $\varphi \equiv \mathsf{K}_a\psi$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w' \Vdash \psi$ for all $w' \in W$ with $R_a ww'$ aml:el:trw:defn:sub:mmodels-box

Here's where we need to think about restrictions on our accessibility relations, though. After all, by clause (9), a formula $\mathsf{K}_a\psi$ is true at w whenever there are no w' with $R_a ww'$. This is the same clause as in normal modal logic; when a world has no successors, all $\Box$-formulas are vacuously true there. This seems extremely counterintuitive if we think about K as representing *knowledge*. After all, we tend to think that there are *no* circumstances under which an agent might know both φ and $\neg\varphi$ at the same time.

One solution is to ensure that our accessibility relation in epistemic logic will always be *reflexive*. This roughly corresponds to the idea that the actual world is consistent with an agent's information. In fact, epistemic logics typically use S5, but others might use weaker systems depending on what exactly they want the K_a relation to represent.

Problem 2.1. Consider which of the following hold in Figure 2.1:

1. $\mathfrak{M}, w_1 \Vdash \neg q$;
2. $\mathfrak{M}, w_1 \Vdash \mathsf{K}_a\neg q$;
3. $\mathfrak{M}, w_1 \Vdash \mathsf{K}_b\neg q$;
4. $\mathfrak{M}, w_2 \Vdash \mathsf{K}_b q \vee \mathsf{K}_b\neg q$;

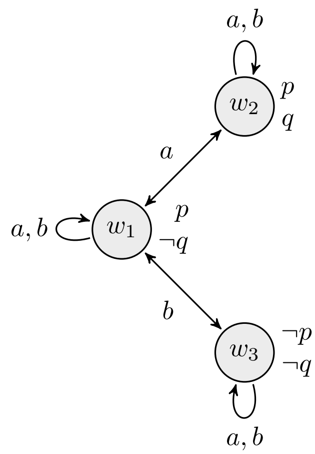

Figure 2.1: A simple epistemic model.

aml:el:trw: fig:simple

5. $\mathfrak{M}, w_2 \Vdash \mathsf{K}_a(\mathsf{K}_b q \lor \mathsf{K}_b \neg q)$;

6. $\mathfrak{M}, w_3 \Vdash \mathsf{E}_{\{a,b\}} \neg q$;

Now that we have given our basic definition of truth at a world, the other semantic concepts from normal modal logic, such as modal validity and entailment, simply carry over, applied to this new way of thinking about the interpretation for the modal operators.

We are now also in a position to give truth conditions for the common knowledge operator C_G. Recall from **??** that the *transitive closure* R^+ of a relation R is defined as

$$R^+ = \bigcup_{n \in \mathbb{N}} R^n,$$

where

$$\begin{aligned} R^0 &= R \text{ and} \\ R^{n+1} &= \{\langle x, z \rangle : \exists y (R^n xy \land Ryz)\}. \end{aligned}$$

Then, where G is a group of agents, we define $R_G = (\bigcup_{b \in G} R_b)^+$ to be the transitive closure of the union of all agents' accessibility relations.

Definition 2.6. If $G' \subseteq G$, we let $\mathfrak{M}, w \Vdash \mathsf{C}_{G'}\varphi$ iff for every w' such that $R_{G'}ww'$, $\mathfrak{M}, w' \Vdash \varphi$.

If R is ...	*then ... is true in $\mathfrak{M}$:*
	$\mathsf{K}(p \to q) \to (\mathsf{K}p \to \mathsf{K}q)$ (Closure)
reflexive: $\forall w Rww$	$\mathsf{K}p \to p$ (Veridicality)
transitive: $\forall u \forall v \forall w((Ruv \land Rvw) \to Ruw)$	$\mathsf{K}p \to \mathsf{KK}p$ (Positive Introspection)
euclidean: $\forall w \forall u \forall v((Rwu \land Rwv) \to Ruv)$	$\neg\mathsf{K}p \to \mathsf{K}\neg\mathsf{K}p$ (Negative Introspection)

Table 2.1: Four epistemic principles.

aml:el:acc: tab:four

2.5 Accessibility Relations and Epistemic Principles

aml:el:acc: sec

Given what we already know about frame correspondence in normal modal logics, we might want to see what the characteristic formulas look like given epistemic interpretations. We have already said that epistemic logics are typically interpreted in S5. So let's take a look at how various epistemic principles are represented, and consider how they correspond to various frame conditions.

Recall from normal modal logic, that different modal formulas characterized different properties of accessibility relations. This table picks out a few that correspond to particular epistemic principles.

Veridicality, corresponding to the T axiom, is often treated as the most uncontroversial of these principles, as it represents that claim that if a formula is known, then it must be true. Closure, as well as Positive and Negative Introspection are much more contested.

Closure, corresponding to the K axiom, represents the idea that an agent's knowledge is closed under implication. This might seem plausible to us in some cases. For instance, I might know that if I am in Victoria, then I am on Vancouver Island. Barring odd skeptical scenarios, I do know that I am in Victoria, and this should also suggest that I know I am on Vancouver Island. So in this case, the logical closure of my knowledge might seem relatively intuitive. On the other hand, we do not always think through the consequences of our knowledge, and so this might lead to less intuitive results in other cases.

Positive Introspection, sometimes known as the KK-principle, is sometimes articulated as the statement that if I know something, then I know that I know. It is the epistemic counterpart of the 4 axiom. Correspondingly, negative introspection is articulated as the statement that if I *don't* know something, then I know that I don't know it, which is the counterpart of the 5 axiom. Both of these seem to admit of relatively ordinary counterexamples, in which I am unsure whether or not I know something that I do in fact know.

2.6 Bisimulations

aml:el:bsd: sec

One remaining question that we might have about the expressive power of our epistemic language has to do with the relationship between models and the formulas that hold in them. We have seen from our frame correspondence results that when certain formulas are valid in a frame, they will also ensure that those frames satisfy certain properties. But does our modal language, for example, allow us to distinguish between a world at which there is a reflexive arrow, and an infinite chain of worlds, each of which leads to the next? That is, is there any formula A that might hold at only one of these two worlds?

Bisimulation is a relationship that we can define between relational models to say that they have effectively the same structure. And as we will see, it will capture a sense of equivalence between models that can be captured in our epistemic language.

Definition 2.7 (Bisimulation). Let $M_1 = \langle W_1, R_1, V_1\rangle$ and $M_2 = \langle W_2, R_2, V_2\rangle$ be two relational models. And let $\mathcal{R} \subseteq W_1 \times W_2$ be a binary relation. We say that $\mathcal{R}$ is a *bisimulation* when for every $\langle w_1, w_2\rangle \in \mathcal{R}$, we have:

1. $w_1 \in V_1(p)$ iff $w_2 \in V_2(p)$ for all propositional variables p.
2. For all agents $a \in A$ and worlds $v_1 \in W_1$, if $R_{1_a} w_1 v_1$ then there is some $v_2 \in W_2$ such that $R_{2_a} w_2 v_2$, and $\langle v_1, v_2\rangle \in \mathcal{R}$.
3. For all agents $a \in A$ and worlds $v_2 \in W_2$, if $R_{2_a} w_2 v_2$ then there is some $v_1 \in W_1$ such that $R_{1_a} w_1 v_1$, and $\langle v_1, v_2\rangle \in \mathcal{R}$.

When there is a bisimulation between M_1 and M_2 that links worlds w_1 and w_2, we can also write $\langle M_1, w_1\rangle \leftrightarrows \langle M_2, w_2\rangle$, and call $\langle M_1, w_1\rangle$ and $\langle M_2, w_2\rangle$ *bisimilar*.

The different clauses in the bisimulation relation ensure different things. Clause 1 ensures that bisimilar worlds will satisfy the same modal-free formulas, since it ensures agreement on all propositional variables. The other two clauses, sometimes referred to as "forth" and "back," respectively, ensure that the accessibility relations will have the same structure.

Theorem 2.8. *If $\langle M_1, w_1\rangle \leftrightarrows \langle M_2, w_2\rangle$, then for every formula φ, we have that $\mathfrak{M}_1, w_1 \Vdash \varphi$ iff $\mathfrak{M}_2, w_2 \Vdash \varphi$.*

Even though the two models pictured in Figure 2.2 aren't quite the same as each other, there is a bisimulation linking worlds w_1 and v_1. This bisimulation will also link both w_2 and w_3 to v_2, with the idea being that there is nothing expressible in our modal language that can really distinguish between them. The situation would be different if w_2 and w_3 satisfied different propositional variables, however.

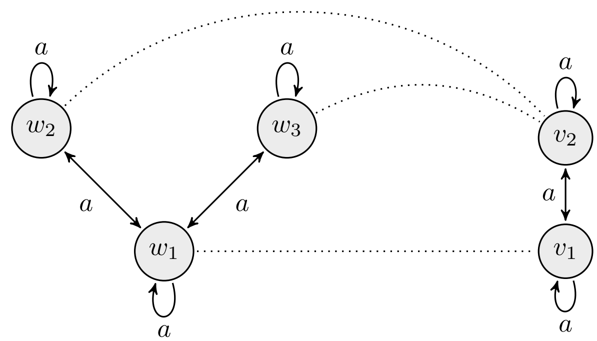

Figure 2.2: Two bisimilar models.

aml:el:bsd:fig:bisimilar

2.7 Public Announcement Logic

aml:el:pal:sec

Dynamic epistemic logics allow us to represent the ways in which agents' knowledge changes over time, or as they gain new information. Many of these represent changes in knowlege using informational *events* or *updates*. The most basic kind of update is a public announcement in which some formula is truthfully announced and all of the agents witness this taking place together. To do this, we expand the language as follows

Definition 2.9. Let G be a set of agent-symbols. The basic language of multi-agent epistemic logic with public announcements contains

1. The propositional constant for falsity $\bot$.
2. The propositional constant for truth $\top$.
3. A denumerable set of propositional variables: p_0, p_1, p_2, ...
4. The propositional connectives: $\neg$ (negation), $\wedge$ (conjunction), $\vee$ (disjunction), $\rightarrow$ (conditional)
5. The knowledge operator K_a where $a \in G$.
6. The public announcement operator $[\psi]$ where ψ is a formula.

The public announcement operator functions as a box operator, and our inductive definition of the language is given accordingly:

Definition 2.10. *Formulas* of the epistemic language are inductively defined as follows:

1. $\bot$ is an atomic formula.
2. $\top$ is an atomic formula.

3. Every propositional variable p_i is an (atomic) formula.
4. If φ is a formula, then $\neg\varphi$ is a formula.
5. If φ and ψ are formulas, then $(\varphi \wedge \psi)$ is a formula.
6. If φ and ψ are formulas, then $(\varphi \vee \psi)$ is a formula.
7. If φ and ψ are formulas, then $(\varphi \rightarrow \psi)$ is a formula.
8. If φ and ψ are formulas, then $(\varphi \leftrightarrow \psi)$ is a formula.
9. If φ is a formula and $a \in G$, then $\mathsf{K}_a\varphi$ is a formula.
10. If φ and ψ are formulas, then $[\varphi]\psi$ is a formula.
11. Nothing else is a formula.

The intended reading of the formula $[\varphi]\psi$ is "After φ is truthfully announced, ψ holds. It will sometimes also be useful to talk about common knowledge in the context of public announcements, so the language may also include the common knowledge operator $\mathsf{C}_G\varphi$.

2.8 Semantics of Public Announcement Logic

aml:el:psm:sec Relational models for public announcement logics are the same as they were in epistemic logics. However, the semantics for the public announcement operator are something new.

aml:el:psm:defn:mmodels **Definition 2.11.** *Truth of a formula φ at w* in a $\mathfrak{M} = \langle W, R, V\rangle$, in symbols: $\mathfrak{M}, w \Vdash \varphi$, is defined inductively as follows:

1. $\varphi \equiv \bot$: Never $\mathfrak{M}, w \Vdash \bot$.
2. $\varphi \equiv \top$: Always $\mathfrak{M}, w \Vdash \top$.
3. $\mathfrak{M}, w \Vdash p$ iff $w \in V(p)$
4. $\varphi \equiv \neg\psi$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w \nVdash \psi$.
5. $\varphi \equiv (\psi \wedge \chi)$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w \Vdash \psi$ and $\mathfrak{M}, w \Vdash \chi$.
6. $\varphi \equiv (\psi \vee \chi)$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w \Vdash \psi$ or $\mathfrak{M}, w \Vdash \chi$ (or both).
7. $\varphi \equiv (\psi \rightarrow \chi)$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w \nVdash \psi$ or $\mathfrak{M}, w \Vdash \chi$.
8. $\varphi \equiv (\psi \leftrightarrow \chi)$: $\mathfrak{M}, w \Vdash \varphi$ iff either both $\mathfrak{M}, w \Vdash \psi$ and $\mathfrak{M}, w \Vdash \chi$ or neither $\mathfrak{M}, w \Vdash \psi$ nor $\mathfrak{M}, w \Vdash \chi$.
9. aml:el:psm:defn:sub:mmodels-box $\varphi \equiv \mathsf{K}_a\psi$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w' \Vdash \psi$ for all $w' \in W$ with $R_a ww'$

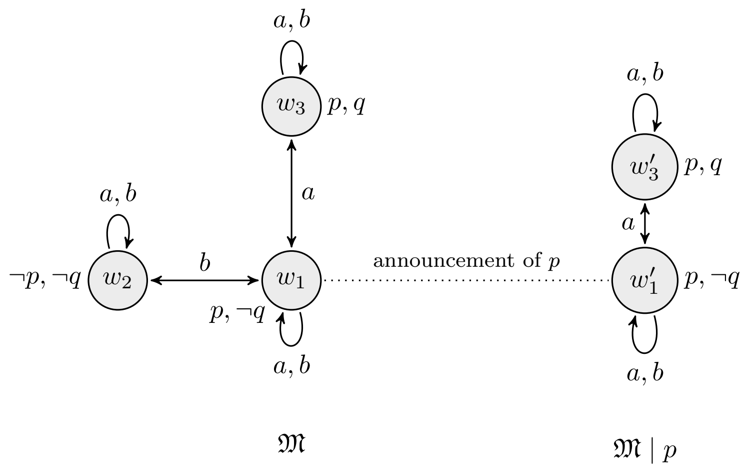

Figure 2.3: Before and after the public announcement of p.

aml:el:psm: fig:announcement-example

aml:el:psm: defn:sub:mmodels-pal

10. $\varphi \equiv [\psi]\chi$: $\mathfrak{M}, w \Vdash \varphi$ iff $\mathfrak{M}, w \Vdash \psi$ implies $\mathfrak{M} \mid \psi, w \Vdash \chi$

    Where $\mathfrak{M} \mid \psi = \langle W', R', V' \rangle$ is defined as follows:

    a) $W' = \{u \in W : \mathfrak{M}, u \Vdash \psi\}$. So the worlds of $\mathfrak{M} \mid \psi$ are the worlds in $\mathfrak{M}$ at which ψ holds.

    b) $R'_a = R_a \cap (W' \times W')$. Each agent's accessibility relation is simply restricted to the worlds that remain in W'.

    c) $V'(p) = \{u \in W' : u \in V(p)\}$. Similarly, the propositional valuations at worlds remain the same, representing the idea that informational events will not change the truth value of propositional variables.

What is distinctive, then, about public announcement logics, is that the truth of a formula at $\mathfrak{M}$ can sometimes only be decided by referring to a model other than $\mathfrak{M}$ itself.

Notice also that our semantics treats the announcement operator as a $\Box$ operator, and so if a formula φ cannot be truthfully announced at a world, then $[\varphi]B$ will hold there trivially, just as all $\Box$ formulas hold at endpoints.

We can see the public announcement of a formula as shrinking a model, or restricting it to the worlds at which the formula was true. Figure 2.3 gives an example of the effects of publicly announcing p. One notable thing about that model is that agent b learns that p as a result of the announcement, while agent a does not (since a already knew that p was true).

More formally, we have $\mathfrak{M}, w_1 \Vdash \neg \mathsf{K}_b p$ but $\mathfrak{M} \mid p, w'_1 \Vdash \mathsf{K}_b p$. This implies that $\mathfrak{M}, w_1 \Vdash [p]\mathsf{K}_b p$. But we have some even stronger claims that we can make about the result of the announcement. In fact, it is the case that $\mathfrak{M}, w_1 \Vdash [p]\mathsf{C}_{\{a,b\}}p$. In other words, after p is announced, it becomes *common knowledge*.

We might wonder, though, whether this holds in the general case, and whether a truthful announcement of φ will *always* result in φ becoming common knowledge. It may be surprising that the answer is no. And in fact, it is possible to truthfully announce formulas that will no longer be true once they are announced. For example, consider the effects of announcing $p \wedge \neg\mathsf{K}_b p$ at w_1 in Figure 2.3. In fact, $\mathfrak{M} \mid p$ and $\mathfrak{M} \mid (p \wedge \neg\mathsf{K}_b p)$ are the same model. However, as we have already noted, $\mathfrak{M} \mid p, w_1' \Vdash \mathsf{K}_b p$. Therefore, $\mathfrak{M} \mid (p \wedge \neg\mathsf{K}_b p), w_1' \Vdash \neg(p \wedge \neg\mathsf{K}_b p)$, so this is a formula that becomes false once it has been announced.

Photo Credits

Bibliography