A BRIEF INTRODUCTION TO MODAL LOGIC

JOEL MCCANCE

ABSTRACT. Modal logic extends classical logic with the ability to express not only 'P is true', but also statements like 'P is known' or 'P is necessarily true'. We will define several varieties of modal logic, providing both their semantics and their axiomatic proof systems, and prove their standard soundness and completeness theorems.

INTRODUCTION

Consider the statement 'It is autumn,' thinking in particular of the ways in which we might intend its truth or falsity. Is it necessarily autumn? Is it known that it is autumn? Is it believed that it is autumn? Is it autumn now, or will it be autumn in the future? If I fly to Bombay, will it still be autumn? All these modifications of our initial assertion are called by logicians 'modalities', indicating the mode in which the statement is said to be true. These are not easily handled by the truth tables and propositional variables of the propositional calculus (PC) taught in introductory logic and proof courses, so logicians have developed an augmented form called 'modal logic'. This provides mathematicians, computer scientists, and philosophers with the symbols and semantics needed to allow for rigorous proofs involving modalities, something that until recently was thought to be at best pointless, and at worst impossible.

This paper will provide an introductory discussion of the subject that should be clear to those with a basic grasp of PC. We will begin by discussing the history and motivations of modal logic, including a few examples of such logics. We will then introduce the definitions and concepts needed for a rigorous discussion of the subject. The third section will show that these systems behave themselves — that is, the set of provable statements and the set of true statements are identical. We will conclude with a discussion of some of the technical complexity of introducing quantification into modal logic.

1. HISTORY AND MOTIVATIONS

As with its classical cousin, the modern interest in modal logic begins with Aristotle.[1] In addition to his syllogisms dealing with categorical statements, the Greek thinker wished to formalize the logical relationships between what is, what is necessary, and what is possible. Unfortunately his treatment of modality suffered from a number of flaws and confusions, and while his categorical syllogisms became a staple of classical education, modal logic was dismissed as a failure. Kant and Frege both argued that modalities added no pertinent information to an argument,

The author would like to thank Professor Jen Brown for her helpful comments and Professor Bob Milnikel for his advice and support.

[1] This historical information in this section is drawn from Fitting [1].

merely hinting at why we might believe a given statement to be true. They believed that no more or less could be derived from the modal form a statement P that from P itself.

This claim has come to be seen as false. After all, if two statements are equivalent, they ought to imply each other. It seems reasonable to say that if P is the case then P must be a possible state of affairs, since what is true cannot be impossible. However, it is quite a bit less obvious to say that because P is possible, P is the case. It is possible, for example, that my bearded friend Charles is actually a very hirsute woman, yet there is no reason to believe that this is actually true. So while actuality implies possibility, possibility does not imply actuality. It seems, then, that there is more to modality than Kant and Frege suspected; modal statements are not quite equivalent to their nonmodal counterparts.

With the growing acceptance of such arguments in the past fifty years or so, there has been a revival of interest in modal logic, the product of which has been a number of interesting new modalities. Temporal logic, for instance, considers whether P is true now, will be true at some point in the future, or has been true in the past. This is actually a multimodal logic since it uses a number of modes of truth within a single language. Epistemic logic — the logic of knowledge and knowing — can also be multimodal. The modality in this case is that of whether agent A knows P or, alternatively, whether P could be true given what A knows. Related is the interpretation of $\Box$ which reads $\Box P$ as 'P is provable', a reading that has great use in the field of mathematical logic.

A final, interesting twist on the theme is action-based logics. As the name would imply, what we are interested in here is how actions alter the truth of a statement. The example in the introduction of flying to Bombay would be an action modality. Say 'It is autumn where I am' is true and 'I am in Ohio' is true. Then after flying to Bombay, both these statements will become false, as it cannot be autumn on both sides of the world simultaneously and, of course, I will no longer be in Ohio. While trivial in this example, a multimodal action-based logic provides a means of teasing out the finer implications of complex sequences of actions.

2. Propositional Modal Logic

Any complete system of logic needs at least three components: a rigorous language for writing out the statements in question, a means of interpreting the statements and determining their truth value, and a means of writing proofs.[2]

2.1. The Language of Modal Logic.

Any language, be it English, Chinese, or the language of logic, must have symbols and rules for combining them. In a spoken language the symbols are words and the rules are grammar. Our case is analogous, although rather than inventing a new system from whole cloth, modal logic begins with the familiar language of PC and just adds two new operators to handle modality. The result is the following set of symbols:

- A countably infinite set of letters A, X, P_1, P_2, &c., called 'propositional variables';
- The unary operators $\Box$, $\Diamond$, and $\neg$;
- The binary operators $\implies$, $\vee$, and $\wedge$; and
- Brackets (, and).

[2] The definitions in this section are adapted from Hughes [3] and Fitting [1].

P	$P \vee \neg P$
T	T
F	T

FIGURE 1. Truth table for '$P \vee \neg P$'.

We read $\Box$, $\Diamond$, $\neg$, $\vee$, $\wedge$, and $\Longrightarrow$ as 'box', 'diamond', 'not', 'or', 'and', and 'implies'. $\Box$ and $\Diamond$ are, of course, our new, modal operators. Their generic names allow us to create a single, rigorous system that can be adapted for the various modalities we may wish to implement. All these names are, of course, merely names thus far. We have not yet bestowed on them any formal interpretation. The next step, then, is to define which combinations of these symbols are 'legal' — which strings make sense given the interpretation we would like to eventually give. These rules reflect the intuition that tells us that '$(P \vee Q)$' is a good formula, yet '$) \vee PQ$' is symbolic gibberish. These 'good' formulae we will call 'well-formed formulae', or 'wffs' for short.

Definition 2.1. A 'well-formed formula' or 'wff' is any formula that meets one of the following three rules.

- Any propositional variable P standing alone is a wff.
- When α is a wff, so are $(\neg\alpha)$, $(\Box\alpha)$, and $(\Diamond\alpha)$.
- When α and β are both wffs, so are $(\alpha \vee \beta)$, $(\alpha \wedge \beta)$, and $(\alpha \Longrightarrow \beta)$.[3]

Note the recursion of this definition, first defining a 'base case' (called the 'atomic formulae') and going on to define more complex formula in terms of those wffs already known to us. This will give us a powerful tool for future proofs where we first prove something about isolated propositional variables and go on to show that it holds for the second and third formation rules listed above. This is called 'induction on the complexity of a formula', and will play a crucial role in our proofs for soundness and completeness below.

2.2. **Determination of Truth.** In PC we are primarily interested in the tautologies. These are the formulae that are true no matter what the truth of the propositions involved, things like '$(P \wedge Q) \Longrightarrow P$' and '$P \vee \neg P$'. These are called the 'valid formulae' of PC and are determined by simply checking the truth table for the formula in question. For example, we can tell that $P \vee \neg P$ is valid because every line of truth table in Figure 1 comes out to 'T'.

The situation for modal logic is somewhat more complex. After all, the whole point of our new symbols is to indicate a judgement that is independent of the truth of the formula in question. Nonetheless, we want an extension of the same intuitive notion: that the valid formulae are those that are 'True no matter what' or 'True in all situations'. The difference is that now it is possible that some propositions, while not tautological, will nonetheless always evaluate true. For example, it has been a popular move in theology to claim that it is necessary that there exist a greatest conceivable being. Such theologians are not generally claiming that God's existence is a tautology, but rather that in every conceivable world the proposition 'God exists' is true. Therefore, they argue, it is a necessary truth that God exists.

[3]The parentheses in this and the preceding rule, while crucial in some instances for rigor, will often be omitted when the meaning is clear from context and convention. This way we will not have to write $((\neg P) \vee Q)$ when $\neg P \vee Q$ is perfectly understandable.

It is this reading of 'necessarily true' as 'true in all possible worlds' that lead to the most popular interpretation of modal logic: Kripke's many-world semantics. Under this interpretation, the truth of a statement is relative to the world in question. For propositional formulae, this is determined simply by examining the state of affairs in that world. So if P and Q are both true in the current world, $P \wedge Q$ will be true in this world. The more interesting case comes with our new operators, $\Box$ and $\Diamond$. $\Box P$ is defined to be true in a world whenever P is true in all 'accessible' worlds. How we define accessibility depends on the modality, but 'conceivable' is a common one for the necessary/possible modality. So if P is true in all conceivable worlds, $\Box P$ is true — that is, P is necessarily true. $\Diamond P$ is similar, although in this case the modality is that of possibility. If P is true in at least one accessible world, $\Diamond P$ will be true as well — since it is true somewhere, it must not be impossible.

To consider another example, say we were using $\Box P$ to mean 'X believes P', where X is some person or ideology. The possible 'worlds' here are not really worlds at all, but people, ideologies, institutions — anything that can be said to believe a proposition. Accessibility in this case is interpreted as 'trusts'. So if Platonists trust physics, then physics is accessible to Platonism. More elaborately, say that Russell is a node in this network of people and ideologies. For this example, say Russell trusts physics, Richard Rorty, and atheism exclusively. Then $\Box P$ is only true for Russell if P is true for physics, Richard Rorty, and atheism. If P is not true in any one of these, than Russell does not believe P. Similarly, if only Richard Rorty believes P, then $\Diamond P$ is true for Russell. He can see how one would believe P, but is not fully convinced. Finally, note that the truth or falsehood of P in Scientology will have no effect on what Russell believes because he does not trust it — Scientology is not accessible to Russell by the relation 'trusts'.

With these examples in mind, we can now rigorously define the semantics of our symbols.

Definition 2.2. Let W be a non-empty set of what we will call 'possible worlds'. Let R be a binary relation from W to W, which we will call an 'accessibility relation'. Together, $\langle W, R\rangle$ form a 'frame'.

Definition 2.3. Let $\langle W, R\rangle$ be a frame and let $\Vdash$ (read 'forces') be a binary relation between W and the set of all well-formed formulae. Let $\Gamma \in W$. We will assume that $\Vdash$ obeys the following rules

- For all propositional variables P, either $\Gamma \Vdash P$ or $\Gamma \Vdash \neg P$.
- If α is a wff, then $\Gamma \Vdash \neg\alpha$ if and only if $\Gamma \nVdash \alpha$.
- If α and β are wffs, then $\Gamma \Vdash (\alpha \vee \beta)$ if and only if $\Gamma \Vdash \alpha$ or $\Gamma \Vdash \beta$.
- If α and β are wffs, then $\Gamma \Vdash (\alpha \wedge \beta)$ if and only if $\Gamma \Vdash \alpha$ and $\Gamma \Vdash \beta$.
- If α and β are wffs, then $\Gamma \Vdash (\alpha \Longrightarrow \beta)$ if and only if $\Gamma \nVdash \alpha$ or $\Gamma \Vdash \beta$.
- $\Gamma \Vdash \Box\alpha$ only if for every $\Delta \in W$, $\Gamma R \Delta$ implies that $\Delta \Vdash \alpha$.
- $\Gamma \Vdash \Diamond\alpha$ only if there exists a $\Delta \in W$ such that $\Gamma R \Delta$ and $\Delta \Vdash P$.

Together, $\langle W, R, \Vdash\rangle$ is a 'propositional modal model', which we will generally shorten to 'model'.

There are a few items of note in this definition. First, observe that all but the last two cases are identical to the truth-table semantics of PC. Only the final cases add anything new. It is also worth observing that just as we can rewrite $\vee$ and $\wedge$ using just negation and implication, we can also rewrite $\Diamond$ in terms of negation and $\Box$. Think of what it would mean for $\Diamond P$ to be false in a world. This would mean

Logic	Frame Conditions
K	no conditions
D	serial[4]
T	reflexive
B	reflexive, symmetric
K4	transitive
S4	reflexive, transitive
S5	reflexive, symmetric, transitive

FIGURE 2. Common systems of modal logic and the restrictions they place on the relations between frames.

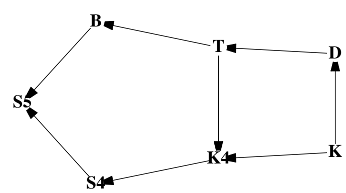

FIGURE 3. Relative strength of some standard frames in modal logic.

that in every accessible world P is false, so the negation of P will be true in all possible worlds — that is, $\Box\neg P$ is true. So $\Diamond P$ is equivalent to $\neg\Box\neg P$. A helpful analogy here is to think of why $(\exists x)(P(x))$ is equivalent to $\neg(\forall x(\neg P(x)))$. To say that there is an x such that $P(x)$ is true is to say that it is false that for every x, $P(x)$ is false. Similarly, $\Diamond X$ simply asserts that it is false that X is false in every accessible world.

Now that we have a clear interpretation of our symbols, we can finally define what it means for a formula to be valid in a given system.

Definition 2.4 (**L**-valid)**.** Let $\langle W, R, \Vdash\rangle$ be a model. We say that this model is 'based on the frame $\langle W, R\rangle$'. Let α be a well-formed formula. α is 'valid in $\langle W, R, \Vdash\rangle$' if $\Gamma \Vdash \alpha$ for every $\Gamma \in W$. α is 'valid in the frame $\langle W, R\rangle$' if it is valid in every model based on $\langle W, R\rangle$. Finally, if α is valid in a collection of frames **L**, then we say it is **L**-valid.

Since frames are just a set with a relation, it is natural to want to choose collections of frames based on properties of R. This is exactly what logicians have done, beginning with the system **K**, which places no restrictions on the frame whatsoever. Other common systems and frame conditions are listed in Figure 2.

Note the interconnections implied by this table. For example, any formula that is **K**-valid ought to be valid in all these systems, since its truth relies on no particular frame structure. Similarly, what is true in **B** will be true in **S5**, since the former is identical to the latter with the exception of not necessarily being transitive. The exact relationships are illustrated in Figure 3.

2.3. **An Axiomatic Treatment of Proofs.** We now have a means of writing statements and of talking about which are true in which circumstances. However,

[4]For every world $\Gamma \in W$ there exists at least one $\Delta \in W$ such that Δ is accessible to Γ.

the fact that the reader has probably taken an introductory *proof* course should suggest that more is required from a system of logic. We do not want to have to compute truth tables for every possible world whenever we want to establish the veracity of a statement. Instead we want a system that formalizes the usual reasoning process of mathematics: beginning from what we know and applying rules and axioms to generate new theorems.

In the discussion that follows we will consider the system **K**. Since **K** makes no assumptions about the structure of the frames, we will be able to augment **K** with more axioms to transform it into any of the systems discussed so far.

First, the definition of a proof:

Definition 2.5. An axiomatic 'proof' is a finite sequence of formulae, each of which is either an axiom or else follows from the earlier terms of the sequence by one of the rules of inference. An axiomatic 'theorem' is the last line of a proof.

The axioms of our system are as follows:

Definition 2.6. There are two classes of axioms for **K**

Classical Tautologies All valid formula of PC (that is, all the tautologies of traditional propositional logic) will be taken as axioms.

Schema **K** For any wffs α and β, we will assume that

$$\Box(\alpha \implies \beta) \implies (\Box\alpha \implies \Box\beta).$$

Our proof system will also have two rules of inference.

Definition 2.7. A 'rule of inference' is an ordered pair (Γ, α), where Γ is a set of wffs and α is a single wff. If the propositions of Γ are theorems of the system, so is α.

K has two rules of inference.

- Modus ponens: $(\{\alpha, \alpha \implies \beta\}, \beta)$.
- Necessitation: $(\{\alpha\}, \Box\alpha)$.

Since we are introducing axioms, it is worth taking a moment to ensure that they are reasonable. While the classical tautologies and modus ponens are the same as in PC, Schema **K** and Necessitation may take some argument. We will start with Schema **K**: $\Box(X \implies Y) \implies (\Box X \implies \Box Y)$. Say that we have a proof for $\Box(X \implies Y)$ and that we are interpreting $\Box$ in terms of necessity. Then what we are saying is that we have proven it necessarily true that $X \implies Y$. If we can prove that X is necessarily true then we will have shown that X is true in every possible world. Since $X \implies Y$ must be true in every possible world, it is reasonable to say that Y must be as well and that Y is therefore necessarily true. Schema **K** is also reasonable under epistemic logic. If know that $X \implies Y$ and we know X, we also know Y.

First, note that since we are in a proof system, X does not mean 'X is the case.' We are rather asserting that 'X is provable.' Consider the two unimodal logics we have discussed: the logic of necessity and epistemic logic. For the necessary/possible modality we are claiming that if we have a proof for X from our other axioms and rules of inference, X must be necessarily true. This is exactly the sort of result we want from a proof system. Similarly, in epistemic logic we are claiming that if we have a proof for X, we know X. This also makes sense. If we have proven X, we

Name	Scheme
D	$\Box P \Longrightarrow \Diamond P$
T	$\Box P \Longrightarrow P$
4	$\Box P \Longrightarrow \Box\Box P$
B	$P \Longrightarrow \Box\Diamond P$

FIGURE 4. Common axiom schemes for modal proof systems.

expect X to be both true and justified. So Necessitation seems a reasonable axiom to adopt.

Before giving an example of an axiomatic proof, we must first introduce the Derived Rule of Regularity. Derived rules are rules of inference that can be deduced from the axioms and rules of inference already available. We can treat them as rules of inference when convenient, since they can always be translated back into the official axioms and rules when need be.

Definition 2.8 (Derived Rule of Regularity)**.** $(\{X \Longrightarrow Y\}, \Box X \Longrightarrow \Box Y$.

This is a derived rule since for any X and Y, the following 'official' steps will produce the same result:

$X \Longrightarrow Y$	Occurs somewhere in the proof
$\Box(X \Longrightarrow Y)$	Necessitation on the previous
$\Box(X \Longrightarrow Y) \Longrightarrow (\Box X \Longrightarrow \Box Y)$	Axiom **K**
$\Box X \Longrightarrow \Box Y$	Modus Ponens on the previous two lines

Example 2.9 (An axiomatic proof)**.** $\Box(X \wedge Y) \Longrightarrow (\Box X \wedge \Box Y)$.

Proof. The proof is as follows:

1	$(X \wedge Y) \Longrightarrow X$	Tautology
2	$\Box(X \wedge Y) \Longrightarrow \Box X$	Regularity on 1
3	$(X \wedge Y) \Longrightarrow Y$	Tautology
4	$\Box(X \wedge Y) \Longrightarrow \Box Y$	Regularity on 3
5	$[\Box(X \wedge Y) \Longrightarrow \Box X] \Longrightarrow$ $\{[\Box(X \wedge Y) \Longrightarrow \Box Y] \Longrightarrow$ $[\Box(X \wedge Y) \Longrightarrow (\Box X \wedge \Box Y)]\}$	Tautology
6	$[\Box(X \wedge Y) \Longrightarrow \Box Y] \Longrightarrow$ $[\Box(X \wedge Y) \Longrightarrow (\Box X \wedge \Box Y)]$	Modus Ponens, 2, 5
7	$\Box(X \wedge Y) \Longrightarrow (\Box X \wedge \Box Y)$	Modus Ponens, 4, 6

Q.E.D. ■

As stated before, the other systems we have introduced are identical to **K**, with only a few added restrictions on the frames. Axiomatically, these systems simply add more axiom schemes. The added axioms are show in Figure 4. The common systems of modal logic are formed by adding combinations of the schemes to the axiom system **K**, as shown in Figure 5.

3. SOUNDNESS AND COMPLETENESS

It is an interesting fact that our system of proof never introduced a formal link to our method for determining validity. How do we know, then, whether the theorems that we prove are valid in our collection of frames? How do we know if our proof system is strong enough to derive all the valid statements in our collection

Logic	Added Axioms
D	D
T	T
K4	4
B	T, B
S4	$T, 4$
S5	$T, 4, B$

FIGURE 5. Common logics and the axiom schemes necessary to properly axiomatize them.

of frames? These conditions, known as soundness and completeness respectively, are important if a proof system is to be of any use. This section will show that the axioms we have introduced are powerful enough to meet these conditions.

Since we know that we can write all our other operators in terms of $\Box$, $\neg$, $\Longrightarrow$, and $\wedge$, we will restrict our language to these operators in this section. While having the full library of operators is notationally efficient and important to know, the extra cases merely add clutter to the exposition that follows.

3.1. **Soundness.** Showing soundness is relatively simple. If our axioms are valid and our rules of inference never derive an invalid statement from a valid one, then clearly every provable formula will be valid.

We begin with the soundness of **K**.

Theorem 3.1 (**K** is Sound). *If X has a proof using the axiom system* **K***, then X is* **K***-valid.*

Proof. We begin with the rules of inference

Modus Ponens. Assume that the wffs X and $X \Longrightarrow Y$ are **K**-valid. So for all models based on all frames $\langle W, R\rangle$ and for all $\Gamma \in W$, $\Gamma \Vdash X$ and $\Gamma \Vdash X \Longrightarrow Y$. Fix a frame $\langle W, R\rangle$ and a world $\Gamma \in W$. Since $\Gamma \Vdash X \Longrightarrow Y$, we know that either $\Gamma \nVdash X$ or $\Gamma \Vdash Y$. Since we know that $\Gamma \Vdash X$, $\Gamma \Vdash Y$. Since our choice of world, model, and frame within **K** were all arbitrary, Y is **K**-valid. So modus ponens is a valid rule of inference.

Necessitation. Assume that X is **K**-valid. So for all models based on all frames $\langle W, R, \rangle$ and for all worlds $\Gamma \in W$, $\Gamma \Vdash X$. Fix a frame $\langle W, R\rangle$ and a world $\Gamma \in W$. Consider a $\Delta \in W$ such that $\Gamma R \Delta$. Since X is true in every world, clearly $\Delta \Vdash X$. So X is true in every world accessible to Γ. Ergo, $\Gamma \Vdash \Box X$. Since our choices of frame and world were all arbitrary, $\Box X$ is **K**-valid. So Necessitation is a valid rule of inference as well.

All that remains is to prove our axioms valid. This is trivial for the Classical Tautologies since our method of determining truth for non-modal statements is identical to the method used in PC. Since the Classical Tautologies are exactly the valid statements of PC, clearly they are valid. The only remaining task therefore is to show the validity of Schema **K**.

Schema **K***.* We wish to show that $\Box(X \Longrightarrow Y) \Longrightarrow (\Box X \Longrightarrow \Box Y)$ is true for all worlds in all frames. Let $\langle W, R\rangle$ be a frame and Γ a world in W. Assume that $\Box(X \Longrightarrow Y)$ is true in Γ. This means that if $\Gamma R \Delta$ then $\Delta \Vdash (X \Longrightarrow Y)$. We want to show that $\Box X \Longrightarrow \Box Y$ is true in Γ. We know this will be true if it is always the case that either $\Box X$ is not true in Γ or $\Box Y$ *is* true in Γ. Assume that $\Box X$ is true in

Γ. So for all $\Delta \in W$, $\Gamma R \Delta$ implies that $\Delta \Vdash X$. Fix such a Δ. Since $\Box(X \implies Y)$ is true in Γ and $\Gamma R \Delta$, we know that $\Delta \Vdash X \implies Y$. So Y must be true in Δ, making $\Box Y$ true in Γ. So in every world Γ in any frame, either $\Box(X \implies Y)$ is false or $\Box X \implies \Box Y$ is true. Hence the axiom is **K**-valid.

Since our axioms are sound and our rules of inference produce sound theorems from other sound theorems, every provable X in the axiom system **K** is **K**-valid. ■

The proofs for the soundness of the other systems are largely corollaries of the above proof. Since they have the same rules of inference and only add axioms, all that must be done is to show that the axiom schemes are **L**-valid, where **L** is whichever collection of frames in question.

Theorem 3.2 (**D**, **T**, **K4**, **B**, **S4**, and **S5** are Sound)**.** *Let* **L** *be* **D***,* **T***,* **K4***,* **B***,* **S4***, or* **S5***. If X has a proof in the axiom system* **L***, the X is* **L***-valid.*

Proof. Since **K** is valid and each system **L** is just **K** with some extra axiom schemes, all we must do is show that each axiom scheme is **L**-valid in the appropriately structured frame.

(D: $\Box P \implies \neg\Box\neg P$) We wish to show $\Box P \implies \neg\Box\neg P$ is true in all serial frames — that is, in all frames where every world is related to at least one other world. Let $\langle W, R, \Vdash\rangle$ be a model and Γ a world in W. Say that $\Box P$ is true in Γ. Then if $\Gamma R \Delta$, P is true in Δ. We want to show that $\neg\Box\neg P$ is true in Γ. In other words, that there exists a Δ accessible to Γ such that $\Delta \Vdash P$.

Since $\langle W, R\rangle$ is serial, there exists at least one Δ accessible to Γ. Fix such a Δ. Since $\Box P$ is true in Γ by hypothesis and $\Gamma R \Delta$, we know that $\Delta \Vdash P$. It follows that $\neg P$ is false in Δ, so there exists a world accessible to Δ in which $\neg P$ is false. Therefore $\Box\neg P$ is not true in Γ, so $\neg\Box\neg P$ is. Since our frame and choice of world were arbitrary, $\Box P \implies \neg\Box\neg P$ is **D**-valid. Since the only extra axiom of **D** is D, **D** is sound.

(T: $\Box P \implies P$) Let $\langle W, R\rangle$ be a reflexive frame — that is, for all $\Gamma \in W$, Γ is accessible to itself. Fix such a Γ and assume that $\Box P$ is true in Γ. Then in every world accessible to Γ, P is true. Since Γ is accessible to itself, P must be true in Γ. So T is true in all reflexive models, including **T**. Since T was the only extra axiom of this system, **T** is sound.

(4: $\Box P \implies \Box\Box P$) Let $\langle W, R\rangle$ be a transitive frame. Let Γ be a world in W and assume $\Gamma \Vdash \Box P$. Then for every Δ accessible to Γ, P is true in Δ. Either Δ is related to another world Θ or it is not. If it is not, $\Box P$ is vacuously true in Δ. Say that $\Delta R \Theta$. Then since the model is transitive, $\Gamma R \Theta$. So since $\Box P$ is true in Γ, P is true in Θ. Since our choice of Θ was arbitrary, P is true in all worlds related to Δ. Therefore $\Box P$ is true in every Δ related to Γ, so $\Box\Box P$ is true in Γ. 4 is therefore true in all transitive frame, including **K4**.

Since 4 was the only added axiom to this system, **K4** is sound. Furthermore, since **S4** is both reflexive and transitive, T and 4 are both valid in **S4**. Since these were the only extra axioms, **S4** is sound.

(B: $P \implies \Box\neg\Box\neg P$) Let $\langle W, R\rangle$ be a symmetric, reflexive frame and let $\Gamma \in W$. Say that P is true in Γ. Since the frame is reflexive, we know there exists at least one world accessible to Γ. Let Δ be such a world. Since the frame is symmetric, Γ is accessible to Δ. We know that P is true in Γ, so $\neg\Box\neg P$ is true in Δ. As this holds for all worlds accessible to Γ, $\Box\neg\Box\neg P$ is true in Γ.

B is therefore valid in all symmetric, reflexive frames, including **B**. Since **B** is reflexive, T also holds in **B**. Since T and B were the only added axioms, **B** is sound.

Finally, **S5** simply requires that frames are reflexive, symmetric, and transitive. S, 4, and B are therefore all valid in **S5**. Since these were the only added axioms, **S5** is sound. ■

3.2. **Completeness.** The argument for completeness is significantly more complex that the one for soundness. This section begins with some preliminary definitions and results that work for any system **L**. However, the actual proof of completeness will be for **K**, with the remaining systems as corollaries as before.

Definition 3.3 (Consistent)**.** Let $\perp$ be an abbreviation for $(P \wedge \neg P)$. A finite set of formulae $\{X_1, X_2, ..., X_n\}$ is '**L**-consistent' if $(X_1 \wedge X_2 \wedge ... \wedge X_n) \implies \perp$ is not provable using the **L** axiom system. An infinite set is **L**-consistent if every finite subset is **L**-consistent.

Definition 3.4 (Maximal Consistency)**.** A set S of formulae is 'maximally **L**-consistent' if S is **L**-consistent and no proper extension of it is. That is, if $S \subseteq S'$ and S' is also **L**-consistent, then $S = S'$.

Theorem 3.5. *If S is* **L***-consistent, then it can be extended to a maximally* **L***-consistent set.*[5]

Proof. Our general strategy for this proof is to add either formulae or their negations one at a time to a consistent set until we 'run out' of formulae. The resultant set will be consistent because every finite subset of it was consistent. It will also be maximal, since every formula or its negation is already in the set.

Let S be an **L**-consistent set of formulae. Since every formula is finite combination of a finite set of symbols, we know from set theory that the set of all wffs is countable. We can therefore map them to the natural numbers like so: $X_1, X_2, X_3, ...$ We can use this fact to construct a sequence of sets $S_0, S_1, S_2, S_3, ...$ as follows:

$$\begin{aligned} S_0 &= S \\ S_{n+1} &= \begin{cases} S_n \cup X_{n+1} & \text{if } S_n \cup X_{n+1} \text{ is consistent} \\ S_n & \text{otherwise.} \end{cases} \end{aligned}$$

Note first of all that by construction, each S_i is consistent. Furthermore, $S_i \subseteq S_{i+1}$ for every $i \in \mathbb{N}$. We can therefore define the limit of this sequence S^* as $S_0 \cup S_1 \cup S_2 ....$ Our claim is that S^* is maximally consistent — that is, S^* is **L**-consistent and if $S^* \cup S'$ is consistent then $S^* = S'$.

S^* is **L***-consistent.* We will show this by contradiction. Assume that S^* was not **L**-consistent. Then there exists some finite subset $Z = \{Z_1, Z_2, Z_3, ..., Z_n\}$ of S^* such that $(Z_1 \wedge Z_2 \wedge ... \wedge Z_n) \implies \perp$. But we know that not every Z_i came from S since S was **L**-consistent. We also know that for every $Z_i \notin S$ there exists some $j \in \mathbb{N}$ so that $Z_i \notin S_j \cap S_{j-1}$. (That is, S_j is the first set in the sequence containing Z_i.) Since the set of Zs is finite we can fix k equal to the greatest such j. As indicated above, $S_0 \subseteq S_1 \subseteq ... \subseteq S_k$. So since each Z_i is in S_j for some $j \leq k$, it follows that each $Z_i \in S_k$. Therefore S_k is inconsistent with **L**. But we know

[5]The proof of this theorem and those that follow have been adapted from Fitting [1].

that S_i is **L**-consistent for every $i \in \mathbb{N}$, so this gives us a contradiction. Therefore S^* is **L**-consistent.

S^* *is maximal.* This will also be proven by contradiction. Say there was a proper extension of S^* that maintained **L**-consistency. Then there is some formula $Z \notin S^*$ such that $S^* \cup \{Z\}$ is **L**-consistent. But we know that our list of formulae $X_1, X_2, ...$ contained all formulae, so there exists $i \in \mathbb{N}$ so that $Z = X_i$. So either $X_i \in S_i$ or $X_i \notin S_i$. If $X_i \in S_i$, then $X_i \in S^*$ and $S^* \cup \{X_i\}$ is not a proper extension of S^*. So say that $X_i \notin S_i$. Then it must be that $S_{i-1} \cup X_i$ was not consistent. Since $S_{i-1} \subseteq S^*$, $S^* \cup \{X_i\}$ would not be consistent. Therefore there cannot exist a proper extension of S^* that maintains **L**-consistency, so S^* is maximally **L**-consistent. ■

Theorem 3.6. *If the set* $\{\neg\Box B, \Box A_1, \Box A_2, ...\}$ *is* **L***-consistent, so is* $\{\neg B, A_1, A_2, ...\}$.

Proof. We will proceed by contraposition. Assume that $\Gamma = \{\neg B, A_1, A_2, ...\}$ is inconsistent with **L**. This means that there is some finite $\Delta \subseteq \Gamma$ whose elements imply a contradiction. Since any extension of an inconsistent set is itself inconsistent, we can assume that Δ contains $\neg B$ and reorder any A_is in it so $\Delta = \{\neg B, A_1, A_2, ..., A_n\}$. From this we can generate the following proof:

$(\neg B \wedge A_1 \wedge A_2 ... \wedge A_n) \implies \bot$	Assumed
$(A_1 \wedge A_2 \wedge ... \wedge A_n) \implies (\neg B \implies \bot)$	$((X \wedge Y) \implies Z)$ is the same as $X \implies (Y \implies Z)$
$(A_1 \wedge A_2 \wedge ... \wedge A_n) \implies B$	$(\neg X \implies \bot)$ is equivalent to X
$\Box((A_1 \wedge A_2 \wedge ... \wedge A_n) \implies B)$	Necessitation
$\Box(A_1 \wedge A_2 \wedge ... \wedge A_n) \implies \Box B$	Schema **K**
$(\Box A_1 \wedge \Box A_2 \wedge ... \wedge \Box A_n) \implies \Box B$	Theorem 2.9
$(\Box A_1 \wedge \Box A_2 \wedge ... \wedge \Box A_n) \implies (\neg\Box B \implies \bot)$	$\neg X \implies \bot$ is equivalent to X
$(\neg\Box B \wedge \Box A_1 \wedge \Box A_2 \wedge ... \wedge \Box A_n) \implies \bot$	$((X \wedge Y) \implies Z)$ is the same as $X \implies (Y \implies Z)$

So $\{\neg\Box B, \Box A_1, \Box A_2, ...,\}$ has an inconsistent finite subset and is therefore inconsistent itself. The theorem is thus proven by contraposition. ■

Definition 3.7. (Canonical Model) We define the 'canonical model of **L**' $\langle W, R, \Vdash \rangle$ as follows. Let W be the set of all maximally **L**-consistent sets of wffs. We define the binary relation R on W as follows. Let $\Gamma, \Delta \in W$. We say that $\Gamma R \Delta$ if and only if for every proposition of the form $\Box X$ in Γ, X is true in Δ. Finally if P is a propositional variable, $\Gamma \Vdash P$ if and only if $P \in \Gamma$.

Theorem 3.8. *(Truth Lemma) Given a canonical model of* **L** $\langle W, R, \Vdash \rangle$*, a world* $\Gamma \in W$*, and a well-formed formula* α*, the following is true:*

$$\Gamma \Vdash \alpha \textit{ if and only if } \alpha \in \Gamma.$$

Proof. Let Γ be a world in the canonical model of **L** and let α be a well-formed formula. We will proceed by induction on the complexity of α.

Atomic formulae. Say α is a propositional variable P. By definition, $\Gamma \Vdash P$ iff $P \in \Gamma$.

Negations. Say that our formula is $\neg\gamma$, where γ is true in Γ if and only if $\gamma \in \Gamma$. Say that $\Gamma \Vdash \neg\gamma$. Then $\Gamma \nVdash \gamma$, so $\gamma \notin \Gamma$. Since Γ is maximally consistent, it follows that $\neg\gamma \in \Gamma$. Now say that $\neg\gamma \in \Gamma$. Since Γ is consistent, $\gamma \notin \Gamma$. By hypothesis it follows that γ is not true in Γ, so $\Gamma \Vdash \neg\gamma$.

Modalities. Say that our formula is $\Box\gamma$. First assume that $\Box\gamma \in \Gamma$. Say Δ is accessible to Γ. By the way we defined R we know that $\Delta \Vdash \gamma$. Since our choice of

Δ was arbitrary, this holds for all the worlds to which Γ is related. So $\Box\gamma$ is true in Γ.

Now say that $\Box\gamma \notin \Gamma$. Since Γ is maximally consistent, this means that $\neg\Box\gamma \in \Gamma$. Consider the set of all statements beginning with $\Box$, $\{\Box X_1, \Box X_2, ...\}$. We know from the previous theorem that if $\{\neg\Box\gamma, \Box X_1, \Box X_2, ...\}$ is consistent than so is $\{\neg\gamma, X_1, X_2, ...\}$. We can therefore extend this to form a maximally consistent set Δ. Now for every statement of the form $\Box P \in \Gamma$, P is in Δ. So by our inductive hypothesis, $\Delta \Vdash P$. Since for every $\Box P \in \Gamma$ $\Delta \Vdash P$, we know that $\Gamma R \Delta$ by definition. But since $\neg\gamma$ is true in Δ, it must be that $\Box\gamma$ is false in Γ. So by contraposition, if $\Box\gamma$ is true in Γ, then $\Box\gamma \in \Gamma$.

Implications. Say that our formula is $\gamma \implies \delta$ and that γ and δ are each true if and only if they are elements of Γ. Say that $\Gamma \Vdash \gamma \implies \delta$. Then either γ is false in Γ or δ is true. Say that γ is true, so δ must be true as well. Then γ and δ are both in Γ. Since they are both elements of the maximally a maximally consistent set, $\gamma \implies \delta$ must be in that set as well.

Now assume that $\gamma \implies \delta \in \Gamma$. Then either $\neg\gamma$ or δ is in Γ as well, since Γ is maximally consistent. Then either $\Gamma \Vdash \neg\gamma$ or $\Gamma \Vdash \delta$, so $\gamma \implies \delta$ is true in Γ.

Conjunctions. Say that our formula is $\gamma \wedge \delta$ and that γ and δ are true if and only if they are elements of Γ. Say that $\Gamma \Vdash \gamma \wedge \delta$. Then both γ and δ must be true, so $\gamma, \delta \in \Gamma$.

Now say that $\gamma, \delta \in \Gamma$. Then both $\Gamma \Vdash \gamma$ and $\Gamma \Vdash \delta$. So we know by definition that $\Gamma \Vdash \gamma \wedge \delta$. ■

With this machinery in place, we can finally show that **K** is complete.

Theorem 3.9. *Say that X is* **K***-valid. Then there is a proof of X in the axiom system* **K***.*

Proof. With the machinery above this proof becomes relatively straightforward. We will proceed by the contrapositive. Assume that X has no proof in **K**. Since X has no proof, the set $\{\neg X\}$ is consistent. (Since we cannot derive X, neither can we derive $X \wedge \neg X$.)

Extend this set to a maximally consistent set X^* and let $\langle W, R, \Vdash \rangle$ be the canonical model of **K**. Clearly X^*, being maximally consistent, is in W. Since $\neg X \in X^*$, $\neg X$ is true in X^*. Since X^* is maximally consistent, $X \notin X^*$. Therefore X is not true in X^*, meaning that X is not valid in the canonical model.

Now clearly the canonical model of **K** is in the collection of frames **K**, since **K** places no restrictions on its frames. We have therefore shown that there exists a model based on a frame in **K** in which X is not valid. Therefore X is not **K**-valid. ■

To show that the remaining systems are true, we merely need to show that the canonical model of a given system **L** meets the requirements of that system. For example, the proof for **T** consists in showing that the canonical model of **T** is based on a reflexive frame. However, it is helpful to begin with a quick lemma.

Lemma 3.10. *Let $\langle W, R \rangle$ be the canonical model for a system* **L** *and let $\Gamma \in W$. If Γ contains no statements beginning with $\Box$, then $\Gamma R \Delta$ for all $\Delta \in W$.*

Proof. Fix $\Delta \in W$ and let Σ be the set of all statements beginning with $\Box$ in Γ. Define $\Phi = \{X | \Box X \in \Sigma\}$. We know from the definition of the canonical model that $\Gamma R \Delta$ only if $\Phi \subseteq \Delta$. But since by hypothesis Σ is empty, so must be Φ. So $\Phi = \varnothing \subseteq \Delta$. Ergo, Δ is accessible to Γ. ■

Theorem 3.11. *The systems* **D**, **T**, **K4**, **B**, **S4**, *and* **S5** *are all complete.*

Proof. As stated above, the only thing to do here is show that for any axiom system **L** (where **L** is one of the above systems), the frame on which the canonical model is based is in collection of frames **L**. In each of the following proofs let $\langle W, R, \Vdash \rangle$ be the canonical model of **L**, where **L** is the system under consideration.

(**D**) Fix $\Gamma \in W$. Since we have taken D ($\Box P \implies \neg\Box\neg P$) as an axiom of **D**, we know that $(\Box P \implies \neg\Box\neg P) \in \Gamma$. Therefore $\Box P \implies \neg\Box\neg P$ is true in Γ according to the Truth Lemma. Say that there was no $\Delta \in W$ accessible to Γ and let X be a propositional variable. Since there are no worlds accessible to Γ, $\Box X$ is vacuously true. However, since D is true in Γ, we know that $\neg\Box\neg X$ is in Γ (since Γ is maximally consistent). Therefore $\Box\neg X$ must be false in Γ. But this can only be the case if there exists Δ accessible to Γ such that $\Delta \nVdash \neg X$. This contradicts our earlier assumption, so it must be that there exists a Δ accessible to Γ. Since this holds for all $\Gamma \in W$, the canonical frame of **D** is serial.

(**T**) Fix $\Gamma \in W$. Since we have taken T ($\Box P \implies P$) as an axiom of **T**, we know that T is in Γ. Therefore $\Box P \implies P$ is true in Γ. If there are no statements beginning with $\Box$ in Γ then $\Gamma R \Gamma$ is true automatically. So say there are such statements in Γ and let $\Sigma = \{\Box X_1, \Box X_2, ...\}$ be the set of all statements beginning with $\Box$ in Γ. Since axiom T is true in Γ, each X_i is true in Γ as well. So by the definition of canonical models, $\Gamma R \Gamma$. Since our choice of Γ was arbitrary, the frames of **T** are reflexive.

(**K4**) Let $\Gamma, \Delta, \Theta \in W$. Since we have taken axiom 4 ($\Box P \implies \Box\Box P$) as an axiom of **K4**, we know that it must be an element of Γ. So by the Truth Lemma, $\Gamma \Vdash \Box P \implies \Box\Box P$. Assume that $\Gamma R \Delta$ and $\Delta R \Theta$ are true. If Γ has no statements beginning with $\Box$ then we know from the lemma that $\Gamma R \Theta$. So say there are such statements in Γ and let $\Sigma = \{\Box X_1, \Box X_2, ...\}$ be the set of all statements beginning with $\Box$ in Γ. Since axiom 4 is true, we know that $\Box\Box X_i$ is true in Γ for each $i \in \mathbb{N}$. Fix such an i. As Δ is accessible to Γ, it follows that $\Box X_i \in \Delta$. Since it must therefore be true in Δ and $\Delta R \Theta$, X_i must be an element of Θ. So by the definition of canonical models, Θ must be accessible to Γ. Therefore the canonical model of **K4** is transitive.

(**B**) The axiom system **B** includes axioms T and B ($P \implies \Box\neg\Box\neg P$). We know that T implies that $\langle W, R \rangle$ is reflexive from above, so we simply need to show that B guarantees symmetry. We will show this by contraposition. Let $\Gamma, \Delta \in W$ and assume $\Delta R \Gamma$. Say that $X \notin \Delta$. Since Δ is maximally consistent, it follows that $\neg X$ is in Δ. Since we have assumed axiom B, we know then that $\Box\neg\Box\neg\neg X$ is true in Δ. Since Γ is accessible to Δ, this means that $\neg\Box X$ is true in Γ. But then we know that $\Box X$ is not in Γ. So by contraposition, if $\Box X$ is in Γ then X is in Δ. So Δ is accessible to Γ by definition, making the canonical model of **B** symmetric.

(**S4**) The axiom system **S4** includes axioms T and 4. Since we have already shown that T and 4 guarantee reflexivity and transitivity, we know that the canonical model of **S4** is both reflexive and transitive.

(**S5**) The axiom system **S5** includes axioms T, B, and 4. The above proofs are therefore enough to show that **S5** must therefore be reflexive, symmetric, and transitive. ■

4. Quantified Modal Logic

While there are uses for propositional logic — modal or otherwise — we are as mathematicians generally more interested in quantified or 'first-order' logic. Without the quantification provided by $\forall$ and $\exists$ it is extremely difficult (perhaps impossible) to show general statements about such important mathematical structures as the integers or the real numbers. However, adding quantifiers to modal logic turns out to be a somewhat complex matter, even more so than adding them to PC. In this last section we will illustrate the difficulties with an informal examination of just what is required to determine the truth of a quantified modal statement.

The language for quantified modal logic is, as with the propositional case, an expansion of the usual first-order logic. It consists of the following:

- The same operators and punctuation as propositional modal logic;
- The quantification symbols $\forall$ and $\exists$;
- As many variables as we need, represented by lowercase Roman letters.
- As many n-place relation symbols as we need, represented by capitalized Roman letters.
- A special relation $=$, the usual identity relation.

An atomic formula in this language will be any expression of the form $R(x_1, x_2, .., x_n)$, where R is an n-ary relation symbol, and build the rest of the formulas as you would expect.

How might we construct a model for such a language? Since we are interested in adding quantification, it would probably be useful to have some non-empty set over which to quantify. We call this set the 'domain'. We then determine the truth of a formula like $(\forall x)(P(x))$ in a world Γ by checking to see if $P(x)$ is true in Γ for every x in the domain. The evaluation for modalities also follows a similar route. $\Box(\forall x)(P(x))$ can be determined true or false by checking to see if $P(x)$ is true for every value of x in every accessible world. Similarly, $(\exists x)(P(x))$ says that there is some element in the domain for which $P(x)$ is true.

There is a danger on the horizon, however. Consider the veracity of $(\exists x)(\Box(x = x))$ in a world Γ: there exists an $x \in \Gamma$ such that in every accessible world Δ, x is identical to itself. On the surface this seems quite readily true; how could it be that something is not identical to itself? But there is a catch: how do we know whether or not x actually exists in Δ? Does it even make sense to talk of objects that do not exist? Yet we do not want to assert $x \neq x$, since that is a contradiction.

One possible solution is to simply require that if x is in one world's domain, it must be in them all. This is called a 'constant domain model', the frame of which is a collection of worlds W, and accessibility relation R, and a single, non-empty set D. However, it is not inherently unreasonable to want the domain to vary from world to world. In temporal logic, for example, it would be ridiculous to say that every object that exists now also existed 1,000 years ago. Temporal logic therefore is better suited to a 'varying domain model'. The frame of a varying domain model looks identical to that of a constant domain model: a triple $\langle W, R, D\rangle$. In this case, however, D is a function assigning each world in W to a domain. The actual domain for a world Γ is not D, but $D(\Gamma)$.

Unfortunately, this reintroduces the problem of deciding whether or not $x = x$ is true in Δ if $x \notin D(\Delta)$. One, somewhat unsatisfying option is to simply allow that perhaps some statements are neither true nor false. This approach results in what is called a 'partial model'. Another solution is to define a 'frame domain' F, where F is just the union of all the domains. This allows us to talk about a fixed x irrespective of whether or not it exists in the world in question. This way we can define =So x is still identical to x whether or not it exists in a particular world, since the relation = can just be defined as the set of all ordered pairs (y, y) such that $y \in F$. This also solves the problem of whether $P(x)$ is true or not in Δ when $x \notin D(\Delta)$. Without the frame domain, $P(x)$ would be vacuously true. With the frame domain we can reasonably talk about (x) not being in the relation P, so $P(x)$ can be sensibly valued false.

Before concluding, let us consider how what is involved in showing the truth of the statement $(\forall x)(P(x)) \implies (\exists y)\Box(P(y))$ in a varying domain model. We have a world Γ, and its domain $D(\Gamma)$. We must examine all the elements of $D(\Gamma)$ in order to show the hypothesis of the statement. We must then fix our y and check $P(y)$

Conclusion

Modal logic provides logicians with a way to take into account not just the facts themselves, but also how we come by these facts. While perhaps not useful in the study of *a priori* mathematics, a rigorous treatment of modalities provides logicians with the means to tease apart the implications of possibility, belief, and action on the truth of judgments. However, a rigorous, symbolic treatment is required to open up modal logic to the tools of mathematical logicians. By adding a single operator to classical propositional logic, $\Box$, we were able to develop such a treatment that behaves as we would like a logical system to behave: only the true statements are provable and all the provable statements are true. While the predicate case for modal logic requires further development, the propositional case is both sound and complete.

References

[1] Fitting, Melvin and Richard L. Mendelsohn, *First-Order Modal Logic*, The Netherlands: Kluwer Academic Publishers, 1998.

[2] Hughes, G. E. and M. J. Cresswell, *An Introduction to Modal Logic*, London, Great Britain: Spottiswoode, Ballantyne and Co Ltd: 1968.

[3] Hughes, G. E. and M. J. Cresswell, *A New Introduction to Modal Logic*, New York, NY: Routledge, 1996.