# Magical Modality: The Logic of Possibility and Necessity

Mei Rose Connor

October 13, 2024

*A beautiful world we live in, when it IS possible, and when many other such things are possible, and not only possible, but done—done, see you!—under that sky there, every day. Long live the Devil. Let us go on.*

Ernest Defarge, *A Tale of Two Cities*, Charles Dickens

# Contents

# Chapter 0

# Our first meeting all together

## 0.0 Introduction of myself

My name is Mei Rose. My background in Formal Logic consists mostly of undergraduate courses in Philosophical Logic, but I studied Mathematical Logic at UW and at the Universiteit van Amsterdam, for two and one semester respectively. I worked at a software development group in my undergraduate university and wrote a proof-checker in JavaScript. Modal logic has been a passion of mine since taking a course on the Philosophy of Mathematics and seeing its history and uses since the times of Leibniz and earlier to puzzle about statements about possible worlds. My personal logic and mathematics heroes of history are David Hilbert and Amalie Emmy Noether. In recent history, I have looked up to Saul Aaron Kripke, who died in 2022 and left a wave of controversy in his wake that we will hopefully have time to discuss during this project. And alive today, inspiring me to be a Logician, is J. F. A. K. (Johan) van Benthem, the only person I know of who holds permanent research positions at three different universities at the same time.

## 0.1 Introductions of students

I am interested to know how your journey through mathematics and logic has been going so far. All I know about you is that the committee recommended you into the project, what courses you are enrolled in, and which courses you have taken that are relevant to the reading group.

- Please tell me about your favorite mathematical experiences. What made them special or unique?
- What sorts of things do you like to do outside of your coursework? Does mathematics or logic ever factor into those activities?
- Who inspires you in history, or the present day?

## 0.2 Overview and goals

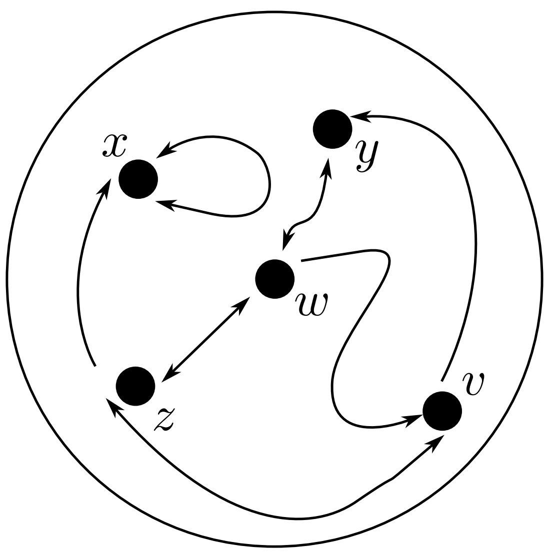
As stated in the project description, I aim for us to discuss many aspects of the so-called modal logic by the end of the semester. After establishing some groundwork in Kripke frames and models and a brief discussion of modal bisimulation, I plan to begin tackling the applications of modal logic. This is where your interests come in. I have the most experience working with modal logic applications in philosophy and linguistics, but there exist many more, such as information/preference theory, game theory, and more. Modal logic also touches more classical logic topics such as provability and model theory, and we can look into those if you are interested. (The "magical" [according to van Benthem] modal proof of Second Incompleteness is what gave this project its name.)

I would like to conclude the project with two connections between modal logic and first-order logic. These two connections are called van Benthem's Theorem and the Salqvist-van Benthem Algorithm. Each gives a slightly different viewpoint on how powerful modal logic is compared to first-order logic.

## 0.3 Propositional and First-order logic reminders

I will be assuming throughout this project that you are familiar with the following logical symbols from propositional logic: $\sim$ (logical negation, sometimes written as $\neg$), $\wedge$ (logical conjunction), $\vee$ (logical disjunction), $\rightarrow$ (material conditional), and $\leftrightarrow$ (biconditional).This means that you know their truth tables, and how to "read" them in a formula (e.g., $P \wedge Q$ gets read as "$P$ and $Q$".)

As for the first-order logic I expect you to be familiar with, $\forall$ and $\exists$ are the starting point. Being able to read these symbols as "for all" and "there exists/there is at least one" is important, as is being able to write and interpret statements with relation or function symbols like $R(x, y)$ or $F(x, y, z)$. When I say that R is a two-place relation between a set and itself (as most of our relations will be in this project), I want you to be able to picture a diagram like the one below. The arrows show the relations. For example, we can say that $R(z, x)$ or that $(z, x) \in R$.




We all have preferences for what symbols and variables we use as Logicians. For example, I prefer to call propositions $P$, $Q$, etc. and use the $\sim$ symbol for the logical negation. I will try to accommodate your preferences, but know that I will fall back on mine if I forget. Van Benthem notably believes, or at least states, that you should be able to change notation for these sorts of things as one changes clothes.

Additionally, when we begin discussing things where the object language and the m etalanguage become important distinctions, my preferred metalanguage symbols for the above-referenced propositional logic connectives are: NOT for the logical negation, AND for the logical conjunction, OR for the logical disjunction, ONLY IF for the material conditional, and IFF for the biconditional. These last two are unlikely to be seen much, but I thought that I should establish them. For the metalanguage universal quantifier (equivalent to $\forall$) I like to use symbol $\bigwedge$, and for the metalanguage existential quantifier (equivalent to $\exists$) I like to use $\bigvee$. These should be distinguished from the object language $\wedge$ and $\vee$ that we will be using for the conjunction and disjunction, respectively.

## 0.4 A few new things

I mentioned above that there is a distinction between what is called the object language and the metalanguage. The object language is the (logical, though can be natural) language we are currently using. The majority of the time this will be the language of modal logic and its associated symbols. The metalanguage is the (same caveat) language we use to talk about the object language. A natural language example is the following, taken from A. Tarski: “Schnee ist weiss” if and only if snow is white. The object language, set off here in quotation marks, is German, and the metalanguage, left outside the quotation marks, is English. The philosophy people will notice that this is similar, but not quite the same, as the *de dicto/de re* distinction.

The distinction will come into play when we want to say things about modal logic, using the language of first-order logic. For example, when we define what it means for a formula of the form $\varphi \wedge \psi$ to be satisfied in a possible world $\alpha$, we will make a statement like: “” $\alpha \Vdash \varphi \wedge \psi$ IFF $\alpha \Vdash \varphi$ AND $\alpha \Vdash \psi$.” The logical conjunction in the part of this sentence before (and not including) IFF is in the object language of modal logic. But the logical conjunction in the part of the sentence after (and including) IFF is in the metalanguage of first-order logic. Essentially, this allows us to talk about logic using logic, and keep clear when we are doing which. We need to be really careful about this distinction. Please prompt me if I am not being clear about when we are using the object language and when we are using the metalanguage.

## 0.5 Before the next time we meet

Please obtain the book Modal Logic for Open Minds by van Benthem. If you are able to get it before the next meeting, please read the first chapter, “A whirlwind history, and changes in perspective” and the second chapter, “Basic language and semantics”. Section 2.4 will be optional, but do feel free to read it and let me know if you want the project to look more closely at “game semantics” like this. I recommend an attempt of Exercises 1(a), 1(b), 2(a), 3(a), and 3(b) to cement your understanding of the reading. We will discuss these during the next meeting. If you choose to read Section 2.4, Exercises 2(b) and 2(c) may help your understanding.

# Chapter 1

# Sing a song of Kripke: Kripke frames and models

Two of the most fundamental objects of study in Modal Logic are Kripke frames and Kripke models, both named for Saul Aaron Kripke, (1940–2022). My lecturer in Amsterdam, who taught me the mathematical foundations of the field of Modal Logic, always said that there were two kinds of Modal Logicians: those who prefer to think in terms of Kripke frames, and those who prefer to think in terms of Kripke models. I suppose that I am more of a Kripke model person myself, despite not really being a model theorist myself. You will see that the two objects roughly correspond to a syntactic (proof theoretic) and semantic (model theoretic) position, respectively. I will leave that matter of taste to you.

## 1.0 The bare definitions

A *Kripke frame* (also known as a relational frame, among other names) is an ordered pair $\mathfrak{F} = (W, R)$ that consists of a set of possible worlds $W$ and an accessibility relation between them denoted $R$. If we are discussing multiple different frames ($\mathfrak{F}$ and $\mathfrak{G}$ perhaps), we might choose to distinguish their sets of possible worlds as $W^{\mathfrak{F}}$ and $W^{\mathfrak{G}}$ respectively. And similarly, for their accessibility relations, we may denote the two relations as $R^{\mathfrak{F}}$ and $R^{\mathfrak{G}}$.

A *Kripke model* (also known as a relational model, possible–worlds model, among other names) is an ordered trirple $\mathfrak{M} = (W, R, V)$ that consists of a set of possible worlds $W$, an accessibility relation $R$, and a valuation function $V$. The valuation assigns propositions such as $P$ truth values (either true or false) in worlds such as $\alpha$. Another way of thinking about this is that $V$ tells which possible worlds have which propositions true in them, therefore assigning propositions to worlds or sets of worlds. Another way to think about Kripke models is that they are Kripke frames that come with a truth function $V$. In fact, the defintion that I was taught was precisely that: a Kripke model is an ordered pair $\mathfrak{M} = (\mathfrak{F}, V)$ where $\mathfrak{F}$ is a Kripke frame and $V$ is a valuation function.

### Some quick notation notes

The accessibility relation $R$ tells us how we can "get from" one possible world in the Kripke frame to another. If $(\alpha, \beta) \in R$, then we can say that "$\alpha$ sees $\beta$", or "we can reach $\beta$ from $\alpha$". The accessibility relation for a Kripke frame is usually written $R(\alpha, \beta)$ or as simply $R\alpha\beta$.

When a proposition $P$ is true in a certain possible world $\alpha$ in a Kripke model, we say that "$P$ is true at $\alpha$", "$\alpha$ satisfies $P$", or, in certain texts, "$\alpha$ forces $P$". These all mean the same thing. They are written with the so–called *modal turnstile* symbol $\Vdash$. The symbol is an infix, so its correct usage is something like $\alpha \Vdash P$ in the case that $\alpha$ satisfies $P$. Some people like to use the *semantic turnstile* symbol $\models$ for this, reflecting the fact that we are asserting something about the semantics of a proposition. I think that the distinction between satisfaction in a Kripke model and satisfaction in general is important, so I will stay away from overloading the $\models$ symbol. Similarly to the "factions" that have arisen over whether the more primitive object is the Kripke frame or model, there are "factions" that "argue" about whether to use $\Vdash$ or $\models$ for modal satisfaction. Ultimately, this is a matter of taste.

## 1.1 A fast overview of the semantics of the symbols

You've already seen this in Chapter 2 of *Modal Logic for Open Minds* as Definition 2.2.2, but I will give it here anyway so that you don't have to go hunting for it the next time you need it. These are the truth–definitions of the operators as they behave in Kripke models. So here we go! For a Kripke model $\mathfrak{M}$ and a possible world $\alpha$ of that Kripke model, we say that ...

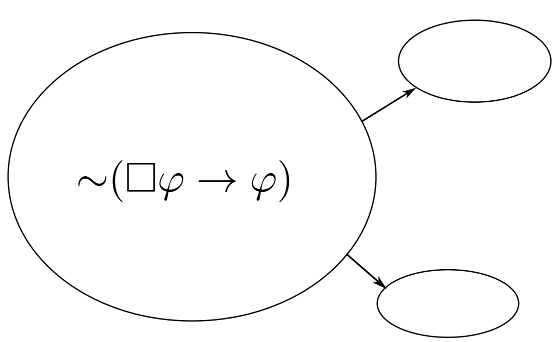
- $\alpha \Vdash P$ IFF the valuation operator $V$ assigns True to $P$.
- $\alpha \Vdash \sim\varphi$ IFF NOT ($\alpha \Vdash \varphi$).
- $\alpha \Vdash \varphi \wedge \psi$ IFF $\alpha \Vdash \varphi$ AND $\alpha \Vdash \psi$.
- $\alpha \Vdash \Diamond\varphi$ IFF $\bigvee \beta$ such that $R\alpha\beta$, $\beta \Vdash \varphi$.
- $\alpha \Vdash \Box\varphi$ IFF $\bigwedge \beta$ such that $R\alpha\beta$, $\beta \Vdash \varphi$.

The rest of the logical connectives' semantics can be deduced from these, as $\sim$ and $\wedge$ are a functionally complete set. I will leave it as an exercise for you to complete if you desire.

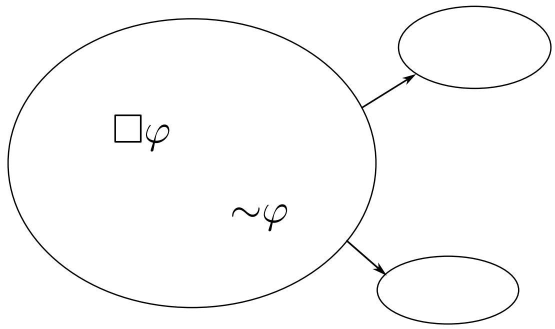
## 1.2 Modal axioms, what are they good for?

Modal axioms are formulae in the language of modal logic that are true for Kripke frames that have a particular structure to their accessibility relation. For example, in every Kripke frame that has a reflexive accessibility relation, it is the case that the Kripke model has the statement $\Box\varphi \to \varphi$ true. And the converse of this statement is true as well: Every Kripke frame for which the statement $\Box\varphi \to \varphi$ holds at every possible world has a reflexive accessibility relation. We can prove this using a picture.

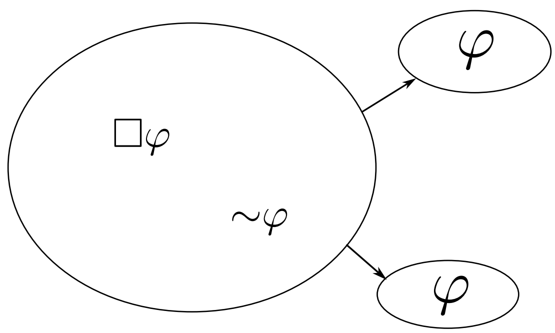
We will proceed by contradiction. Assume for the sake of contradiction that the axiom $\Box\varphi \to \varphi$ is false in some possible world $\alpha$ in our Kripke frame $\mathfrak{F}$. We will show this as the following (the large bubble is $\alpha$):



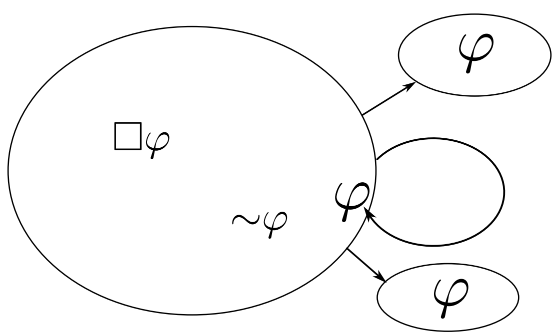
We will use the principle of classical logic that $\sim(\varphi \to \psi) \leftrightarrow \varphi \wedge \sim\psi$ to expand the assumption.



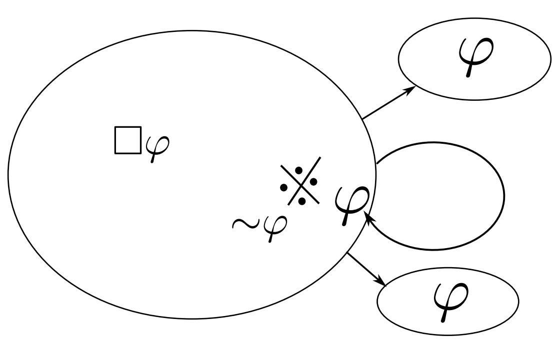

Separating the two conjuncts and applying the semantics of the box symbol means that every possible world that $\alpha$ can see satisfies $\varphi$.




But if our accessibility relation is reflexive, then every possible world can see itself. Then $\varphi$ must be true in $\alpha$ for this reason.




A contradiction arises from the fact that we already asserted that $\sim\varphi$ is true in this world when we separated the two conjuncts. Thus we cannot have a Kripke frame that contains this statement and is not reflexive.



I will fully accept this kind of "picture proof" from you as a demonstration of these types of theorems, most notably because proving these things formally (in a natural deduction system perhaps), is not the goal of this project. I have written formal proofs of fairly simple correspondences of this type that involved 5 layers of nested subderivations, which I do not believe would give you more enlightenment into the axioms than the "picture proofs".

You may have noticed that we used the reasoning principle of *modus ponens* in this proof. Without this essential piece of reasoning, logic has a hard time getting started. So we build this into every Kripke frame through the following axiom: $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi$. This axiom is so foundational, it was given the initial **K** in honor of Kripke. Nearly all modal axioms have initials. The one that we just discussed and proved a theorem about is called **T**. We will continue to explore modal axioms and their relation to Kripke frames and models as a theme in this project.

## 1.3 Looking into next week

For next week, I would like you to hear van Benthem's take on the "landscape of modal logics" by reading Chapter 8 in the book. Also, if you have gotten down to here, you should have read through my notes here thoroughly. I invite you to attempt his Exericse 2 in that chapter. And just for fun, becuase I think that you should do this to get acquainted with the way that modal axioms play with the accessibility relations on their Kripke frames, you should go to the Wikipedia page for the Kripke semantics and try to prove at least two of the correspondences shown in the table *Common modal axiom schemata*. Don't worry about reading the entire page and understanding it. The correspondences are what we will be building upon in the course of the project.

# Chapter 2

# Finding common ground: Modal bisimulation and modal equivalence

For the most part, I am going to hand this topic off to van Benthem, as he can explain it with far better examples than I can come up with on my own. I enjoyed this section of my modal logic course in Amsterdam more than nearly anything else, if only for the sense of humor that my lecturer brought to the topic. Looking back, I can't quite recall what was so amusing about this topic, but it still has that sort of nostalgic charm that one feels about subjects about which one has fond memories associated.

## 2.0 Definition of modal bisimulation

Modal bisimulation, in the words of van Benthem, requires "harmony" between two Kripke models in a way that will be clear from the definition. My lecturer called the two conditions for the interactions of the bisimulation 'back' and 'forth' which I think is appropriate. A bisimulation is a two—place relation between two Kripke models $\mathfrak{M}$ and $\mathfrak{N}$ such that two special conditions hold. The two conditions are the following:

1. For possible worlds $w$ and $w'$, there exists a bisimulation $E$ between $\mathfrak{M} = (W, R, V)$ and $\mathfrak{N} = (W', R', V')$ if when $Eww'$, then $w$ and $w'$ satisfy the same propositional letters.

2. If $Eww'$ and $Rwv$, then there is some $v' \in \mathfrak{N}$ such that $Evv'$ and $R'w'v'$. The converse of this must hold as well: if $Eww'$ and $R'w'v'$, then there exists sme $v' \in \mathfrak{M}$ so that $Evv'$ and $Rwv$.

The first of these condition is called "Harmony" by van Benthem, and the second is the "Back—and–Forth' condition. The first part, potentially confusingly, is the 'Forth' condition, while the converse of it is the 'Back' condition. The diagram the textbook gives makes these names a bit more clear, I think.

### 2.0.0 Notation for bisimulations

The bisimulation itself, the relation between sets of possible worlds in $\mathfrak{M}$ and $\mathfrak{N}$, is usually given a capital letter name, like $E$. It can be written as either a prefix (as I have done to keep consistent with the way that I have been writing relations in these notes) or as an infix (as van Benthem does in our book). Essentially, it is relation $E \subseteq W \times W'$. In the case that we are saying that two Kripke models are bisimilar (the adjective form of the word *bisimulation*), then we write $\mathfrak{M} \leftrightarrows \mathfrak{N}$

## 2.1 Another notion of equivalence

Bisimulation is powerful because it gives us a notion of equivalence of Kripke models. We can define another notion of equivalence between Kripke models, which is called *modal equivalence*. The big idea behind this is the idea of $\tau$–theories. A $\tau$–theory of a possible world $w$ in a Kripke model $\mathfrak{M}$ the set of all $\tau$–formulae that are true in $w$. In our case, $\tau$ is the set of symbols in the modal logic we are working in, which is the system set out by van Benthem on the first page of Chapter 2. I use the general symbol $\tau$ here to represent the fact that construction works for any set of modal symbols, including those that fall outside of the $\Box$ and $\Diamond$.

Two worlds $w$ and $w'$ in two Kripke models $\mathfrak{M}$ and $\mathfrak{N}$ are modally equivalent if they have the same $\tau$–theories. There also an idea that two Kripke models can be modally equivalent, and this in the case that the $\tau$—theories of the *models* are identical. This requires the set of formulae satisfied in all states of $\mathfrak{M}$ to be the same as the set of all formulae satisfied in $\mathfrak{N}$. When there is a modal equivalence between two worlds, we write that $w \leftrightsquigarrow w'$. Analogously, we write a modal equivalence between Kripke models as $\mathfrak{M} \leftrightsquigarrow \mathfrak{N}$.

## 2.2 Analogies between analogies

We now come to the lead—up to one of the big theorems of the project. That theorem, known as van Benthem's theorem, will wait for a few more weeks. But I will give you a taste of it with the following lead–up theorem:

Let $\mathfrak{M} = (W, R, V)$ and $\mathfrak{N} = (W', R', V')$ be Kripke models under a set of modalities $\tau$. Then for every $w \in W$ and every $w' \in W'$, if $w \leftrightarroweq w'$, then $w \leftrightsquigarrow w'$. The proof is by symbolic induction on modal formulae, and relies *precisely* on each part of the definition of bisimulation.

This is powerful indeed! It gives us a characterisation of modal formulae, and also hints to us that bisimulation may be giving us more than we expected. What did we sell our souls for in exchange for this epic result? That question will be answered by the Sahlqvist–van Benthem Algorithm and van Benthem's theorem, both of which will reveal noth the expressivity and limitations of modal logic in the context of first–order logic.

## 2.3 For the week ahead

If you've made it this far in the notes, you have a basic understanding of what it means for possible worlds and Kripke models to be bisimilar and for them to be modally equivalent. I would like you to read van Benthem's section on bisimulation, which are contained within section 3.2 of Chapter 3. I think that with this knowledge, you should feel comfortable with Exercises 1(a) and 1(b) that come at the end of the chapter. Feel free to create your own Kripke models and find (or fail to find) bisimulations between them in order to get a good grasp of the concept. For fun, you may attempt to prove the above theorem if you feel comfortable with symbolic induction.

# Chapter 3

# Lost (and found) in translation: Modal logic's relationship with First-order logic

Modal logic, as you might have suspected, is a fragment of First–order logic. The immediate observation may be that they share all of the propositional connectives, and additonally, the accessibility relation adds a relation, requiring the power of First–order logic to express this. But there lies "something deeply hidden", as one of my mentor figures in undergrad would have expressed it. The last week on bisimulation should have given you a feeling for this kind of relationship between Kripke models. In the remarks following the final theorem of my notes for you, about modal equivalence and bisimulation, I suggested that we gave something up in order to have the power of this theorem. What was this exactly? We gained much, but we also lost some formulae along the way as we developed modal logic. Which formulae are we able to keep? They are precisely the ones that are invariant under bisimulation. In these notes, we will explain what this means.

## 3.0 Standard translation of modal formulae

We must first discuss what it means to translate a modal formula into a first–order one. This means that we are completely eliminating $\Box$ and $\Diamond$ from our formula, and turning propostional letters $P$, $Q$ and the like into single–place predicates that we will name things like $Px$ and $Qx$. There is a procedure for producing this kind of first–order formula known as "standard translation". I will define it inductively, as I did with the definitions of the connectives and modal operators a few weeks ago.

1. Translate each propositional letter into its own single–place predicate. As mentioned above, this looks like turning things like $P$ and $Q$ into $Px$ and $Qx$.

2. Translate $\bot$ into $x \neq x$.

3. Keep in mind that statements of the form $\sim\varphi$ can be turned into $\varphi \rightarrow\bot$, and translate $\sim\varphi$ into the negation of the standard translation of $\varphi$.

4. Translate $\varphi \vee \psi$ into the disjunction of the standard translations of the disjuncts $\varphi$ and $\psi$

5. Translate $\Box\varphi$ into $\forall y(Rxy \rightarrow \Phi)$, where $\Phi$ is the standard translation of $\varphi$.

6. Translate $\Diamond\varphi$ into $\exists y(Rxy \wedge \Phi)$ where $\Phi$ is the standard translation of $\varphi$.

After perfoming this procedure, what has happened is you have turned a modal formula into a first–order one. What did this get us? It shows us which first–order formulae are secretly modal formulae. But one cannot always go backwards, turning all first–order formulae into modal ones. Which are the ones that work? We will now introduce the concept of *invariance under bisimulation*.

## 3.1 Invariance under bisimulation

Invariance under bisimulation for a formula $\varphi$ means that if two pointed Kripke models $(\mathfrak{M}, w)$ and $(\mathfrak{N}, v)$ are bisimilar, then $w \Vdash \varphi$ if and only if $v \Vdash \varphi$. It is an interesting procedure to prove that all modal formulae are invariant under bisimulation, and I will send a presentation that contains this result if anyone is interested. What is more interesting is the first–order formulae that are invariant under bisimulation.

## 3.2 The Big One: Van Benthem's Theorem

This is widely considered (at least amongst the modal logicians I have encountered) to be the most beautiful theorem in all of modal logic. My modal logic lecturer in Amsterdam told us that it was so ground–breaking when first proven and so influential that no one ever attempted to prove it again after van Benthem did in his Ph.D. thesis. This was the first time such a result had ever appeared in modal logic. (Additionally, this document contains the first recoreded reference to bisimuation between Kripke models.) What van Benthem proved set the stage for the next half century or more for modal logicians. Here it is:

**Theorem (van Benthem, 1976)** A first–order formula $\varphi(x)$ is invariant under bismulation if and only if it is equivalent to the standard translation of a modal formula.

I will not be proving this for you, as its proof is beautful (or so I have been told), but requires much study of things that are outside the scope of this project. The proof is given in Blackburn, de Rijke, and Venema's book *Modal Logic*. The arguments are clever and the method takes a while to get your head around. A note: these formulae are not necessarily the ones that define the correspondences between accessibility relations and the axioms. This can be seen in the chart on the Wikipedia page for the Kripke semantics, where some modal axioms have accessibility relations whose defining property is not a first–order property. For example, the McKinsey axiom $\Box\Diamond\varphi \to \Diamond\Box\varphi$ produces a Kripke model whose accessibility relation's defining property is a second–order formula.

What we do get from van Benthem's theorem is a new way to look at the Kripke semantics. We see that first–order formulae that express the relatonships between possible worlds that the Kripke semantics give us precisely the modal formualae that should express them.

## 3.3 Headed into next week

As usual, if you have gotten this far, you have nearly completed your reading for the week. In this case, you have completed it. I have not assigned you reading from *Modal Logic for Open Minds* this week because I do not think that we have covered the topics that he finds are prerequisite to expressing the above theorem. Translate your favorite 3 modal axioms into their standard translation. I will prepare some examples for next week. Also, please begin to think about which of the modal logic applications you would like to explore. The choices that I am most familiar with are the linguistic and the philosophical applications. Van Benthem covers many, many more in the book. Please feel free to look at the table of contents in the book and explore the first paragraph of chapters that contain applications you are interested in. I will give a "speed–dating" approach to the applications next week, and conclude by asking you which of the topics you would like to explore more deeply.