Modal Logic for Open Minds

Johan van Benthem

February 2, 2010
Contents

Preface ix

1 A whirlwind history, and changes in perspective 1

I Core Concepts 7

2 Basic language and semantics 11

3 Expressive power and invariance 25

4 Validity and decidability 37

5 Axioms, proofs, and completeness 49

6 Computation and complexity 61

II Basic Theory 71

7 Translation and varieties of expressive power 75

8 Increasing deductive power: the landscape of modal logics 91

9 What axioms say: frame correspondence 101

10 Descriptive power: extended modal languages 109

11 Modal predicate logic 119
III  Selected Applications  127
 12  Epistemic logic  133
 13  Doxastic and conditional logic  147
 14  Dynamic logic of actions and events  155
 15  Logic and information dynamics  171
 16  Preference and deontic logic  189
 17  Modal logic and games  197
 18  The structure and flow of time  207
 19  Modal patterns in space  219
 20  Intuitionistic logic  233
 21  Provability logic  243

IV  Recent Theoretical Themes  251
 22  Fixed-points, computation, and equilibrium  255
 23  Issues in information dynamics  265
 24  System combination and undecidability  281
 25  Abstract model theory  293
 26  Deductive incompleteness  303

V  Coda  311
 27  Modal foundations for classical logic  315
  A  First-order predicate logic  323
  B  Modal algebra  333
Preface

Modal logic was born in the early part of the 20th century as a branch of logic applied to the analysis of philosophical notions and issues. While it still retains a bit of this grandeur, today, modal logic sits at a crossroads of many academic disciplines, and thus, it provides a unique vantage point for students with broad interdisciplinary interests. These notes are the accumulated material for a course taught for many years at Stanford to students in philosophy, symbolic systems, linguistics, computer science, and other fields. The purpose is to give them a modern introduction to modal logic, beyond lingering conceptions dating back to the distant past – and topics include both technical perspectives, and a wide range of applications showing the current range of the field. To check if the picture in these notes is representative, the reader may consult the 2006 *Handbook of Modal Logic*, Elsevier, Amsterdam, co-edited with my colleagues Patrick Blackburn and Frank Wolter, Elsevier Science Publications, Amsterdam. For philosophers, it may also be of interest to check with my 1988 lecture notes *Manual of Intensional Logic*, CSLI Publications, Stanford, which then represented my ideal of a modern introduction to the field. Some topics have panned out, but others have proved remarkably wide off the mark.

Part I is about basic techniques, Part II gives a first round of theory consolidating these. Part III then tells the story of a wide spectrum of modern applications, many of them about the study of agency, and Part IV is about theoretical issues again, arising out of these. Finally, there is a coda on modal perspectives in the heartland of classical logic itself. Working through this material will give you a modern view that enables you to understand many strands in current research, and maybe even participate in the enterprise, given the (Heaven knows) many open problems in the field today, far beyond the old "capitalism" of studying a zoo of modal logics like "$K$", "$T$", "$S4$", "$S5$" that once ruled.

Still, I felt there was room for something new: a less technical, but still substantial broad text at an earlier level, that initiates a larger student audience to the intellectual excitement of the field of modal logic as a whole, while training them in basic modern techniques that should allow them to see further than the generations before them.

**How to use these notes: theme selection**

This book is intended as an advanced undergraduate/beginning graduate course on Modal Logic. Each short chapter in these notes corresponds roughly to 1 1/2 hour class meeting, supported by a section. A typical course of mine would cover, after the introduction, most of the “mechanics” of the field: basic themes and theory (some 8 topics from Parts I and II), followed by a selection of some 6 current applications from Part III (knowledge and dynamics, but also time and space, were the usual favorites, for their concreteness). With some recap sessions, this came to 9 weeks in a typical Stanford spring quarter.

But as I kept on writing, more and more things crept in. Therefore the book can be used in other ways as well. For instance, after a brief recap of Part I, Part III could be excellent primary reading material for a graduate course on Philosophical Logic, especially, since many of its chapters are largely self-contained. Parts II and IV would also make good secondary material for a course on Metatheory of Logic. And finally, the whole text again could be used by researchers in areas where modal logic is applied these days (such as agent systems, artificial intelligence, or game theory) to learn more of what makes it tick.

Finally, it is a pleasure to acknowledge all the help that I have had. Eric Pacuit wrote extensive reader’s notes that transformed the text. Audrey Yap and Tomohiro Hoshi went through the text with their students, and provided valuable feedback and suggestions. I also received occasional comments from many others, including Fenrong Liu, Darko Sarenac, and Urszula Wybraniec-Skardowska. Then there was the proof-reading team of Viktoria Denisova, Nina Gierasimczuk, Lena Kurzen, Fenrong Liu, Minghui Ma, Ștefan Minică and Junhua Yu.
Throughout the stages of the production process, Fernando Velázquez-Quesada was responsible for making this book happen at all. With all this said, it remains to thank my Stanford students in this course for all the good times we have had over the years.

Johan van Benthem, Amsterdam & Stanford.
A whirlwind history, and changes in perspective

Some truths seem merely “contingent”, such as the fact what clothes you are wearing today: this could easily have been otherwise. But other truths seem “necessary”, such as the fact that, like it or not, you are not someone else. Modal notions of necessity, possibility, and contingency were standard fare in traditional logic up to the 19th century, and reasoning with them was considered a core part of the discipline. All these notions went out the door in the work of the founding fathers of modern logic, like Boole and Frege. In particular, in his famous little book Begriffsschrift from 1879, often taken to be the founding document of modern logic, Frege has a mysterious passage where he seems to be ticking off a list of things that are irrelevant to logic, and one of them is modality. According to that passage, saying that some proposition is necessarily true just means that it is true, plus some autobiographical information about how strongly you believe in it. That list was Kant’s Table of Categories¹, and what happened was that modern logic just kept “extensional” notions like negation and quantification, while dropping “intensional” ones like modality. The result are the familiar logical systems like propositional and predicate logic, which describe properties and relations of objects in fixed situations, represented by models. This historical restriction of the agenda and core tools has proved immensely beneficial, especially in the analysis of the foundations of mathematics, whose Golden Age was in the 1930s with classical results on provability, completeness, computability, and definability by Hilbert, Post, Gödel, Tarski, Turing, and many others. The millennium issue TIME 2000 placed Gödel, Turing, and Wittgenstein among the twenty most influ-

¹Frege does not bother to say this. The habit of citing sources and crediting other authors is much more recent than you might think!
ential intellectuals of the 20th century, an incredible harvest for a small discipline like logic. This book presupposes that readers know the attractions and power of this approach, including the notions of logical syntax, semantics, proof, and meta-theory of formal systems.

Even so, while extensional logics might be adequate for analyzing mathematical proof and truth in an eternal realm of abstraction, modality made a fast come-back. Soon philosophers started using modern logic to deal with patterns of reasoning as used by real agents, expressed in natural language: the noisy, diverse, and fascinating medium which is the trademark of mankind on this planet. And then, one finds that there is a host of notions of a “modal” character going far beyond mere truth: necessity, knowledge, belief, obligation, temporal change, action, and so on. Indeed, it is hard to think of any use of language which is purely informative: every sentence we utter resonates in a web of communication, expectations, goals, and emotions. Modal logic as we know it today tries to analyze this structure with techniques taken from the mathematical turn in modern logic. Incidentally, Frege had nothing against this move per se. In a famous analogy, he compared a formal language to a microscope: very precise, but limited in its realm of application, while natural language was more like the human eye: less precise, but universal in its perceptive sweep.\(^2\)

What follows here is a lightning history. For details and bibliographical references, we refer the reader to four sources. On the philosophy connection, see Roberta Ballarin’s entry on modal logic in the Stanford Encyclopedia of Philosophy (Ballarin, 2008) plus the chapter “Logic in Philosophy” by J. van Benthem in Jacquette (2007). Van Benthem’s Manual of Intensional Logic (van Benthem, 1988a) extends the canvas to linguistics and computer science, while the editorial introduction to the 2006 Handbook of Modal Logic (Blackburn et al., 2006) includes interfaces with all fields in play today.

For a start, soon after Frege and Russell, modal logic made its come-back through a study of the notion of strict implication \(\hspace{1mm} A \Rightarrow B \) (C. I. Lewis). This strengthens the usual propositional implication \( A \rightarrow B \), which amounts to a mere truth-functional link \(\neg(A \land \neg B) \) between the antecedent and the consequent, to the stronger modal connection \(\neg \Box(A \land \neg B) \): it is impossible for \( A \) to be true, and yet have \( B \) false. Modalities per se were then studied by Carnap, Kanger, Kripke, and

\(^2\)Steltzner (1996) explains how the major employer in Frege’s Jena was Zeiss Optics, with its visionary leader Carl Zeiss and Ernst Abbe, and how Frege was supported all of his life through anonymous donations from this source. Modern logic owes a lot to enlightened industrialists who wanted to give back to society.
many subsequent authors, explaining a necessity statement $\Box \varphi$ as saying that $\varphi$ holds throughout some relevant range of situations. This *multiple reference* view takes a modal necessity operator $\Box$ as a universal quantifier $\forall$, and the possibility operator $\Diamond \varphi$ as an existential quantifier $\exists$, both ranging over the relevant “worlds”, points in time, situations, or whatever relevant semantic entity, where $\varphi$ is true. But there are alternatives. As early as the 1930s, Gödel interpreted necessity $\Box \varphi$ as “mathematical provability” of $\varphi$ (an $\exists$-type account!), while Tarski interpreted modal formulas as describing subsets in topological spaces, with $\Box \varphi$ standing for the topological interior of the set defined by $\varphi$.\(^3\) These lecture notes will mainly take the now dominant universal range view, but we will briefly discuss alternatives in the appropriate places. Maintaining some bio-diversity of approaches is a good survival strategy for a field. But whatever view we take, it will be clear that modal logic thrives on co-existence with standard logical systems.

Another source of diversity are the many different technical approaches in the field. These lecture notes will cover both the traditional deductive (proof-theoretic) and semantic (model-theoretic) styles, with one excursion to algebraic methods, an important topic that we had to forego. But on the whole, we will take the viewpoint of “possible worlds semantics”, though resolutely cleansed from its outdated metaphysical interpretations. We will also introduce some new themes beyond the standard catechism, however; in particular, some awareness of expressive power and invariance, and of the computational complexity of modal languages. These further perspectives greatly enrich one’s view of what a modal logic – and indeed any logical system – actually is.

In terms of its natural habitat, modal logic was the main technical vehicle for philosophical logic since the 1950s, and its practitioners like Prior, Kripke, Hintikka, Lewis, or Stalnaker produced a series of beautiful systems, and associated notions and issues that became influential in philosophy, setting the agenda for debates in metaphysics, epistemology, and other fields. This is the period when labels like “modal logic”, “epistemic logic”, “doxastic logic”, “deontic logic”, “temporal logic”, etcetera, were coined, which still form a geography that is widely used, witness many chapters in the *Handbook of Philosophical Logic* (Gabbay and Günthner, 1983-1989). In the 1970s, this philosophical phase was consolidated into a beautiful mathematical theory by authors like Blok, Fine, Gabbay, Goldblatt, Segerberg, and Thomason. But simultaneously, modal logic crossed over to linguistics, when “Montague seman-

\(^3\)As we shall see later, this is a more complex semantic $\exists \forall$-type account.
tics” gave the study of intensional expressions in natural language pride
of place, using mixes of modal logic with type theory and other tools
from mathematical logic. In the same decade and especially through
the 1980s, modal notions found their way into computer science in the
study of programs and computation (Pratt, 1976), and into economics
in the study of knowledge of players in games (Aumann, 1976). And
this migration across the university is still continuing: in the 1990s,
modal languages have turned up in the study of grammars, data-base
languages, and more recently, in web design, and the structure of vec-
tor spaces used in mathematical image processing. The present lecture
notes reflect these realities, including ups and downs in specific fields –
and understanding modal logic today means seeing a total picture,
just like reading your worldwide investment portfolio.

In these twists and turns, something strange has happened, which
confuses many people. Many logicians still see modal logic as an enrich-
ment of classical logic. The modalities increase expressive power, and
may lead to intricate issues of the interplay between, say, quantifica-
tion over objects and modal reference to worlds. But there is another,
and perhaps by now the more widespread, perspective which views
things the other way around. Modal operators are themselves a sort
of quantifiers, but special “local” ones referring only to objects “acces-
sible” from the current one. Viewed in this way, modal languages are
not extensions, but rather fragments of classical ones, with restricted
forms of quantification – and this weakness is at the same time a clear
strength. Compared to classical systems, modal logics lower complex-
ity (they tend to be decidable; and their validities can be described in
transparent variable-free notations), and moreover, modal logics make
us aware of the expressive fine-structure of the richer languages they
are part of. One theme throughout these notes is the resulting “bal-
ance” between expressive power and computational complexity: gains
in one will be losses in the other. Such a balance is not peculiar to
modal logic: higher up, first-order logic itself is an elegant compromise
between expressive power and axiomatizability (note how second-order
logic gains in the first, and loses the second). Indeed, awareness of this
fundamental trade-off is essential to understanding the whole point of
using logical languages to formalize an area of reasoning.

Of course, this does not mean that the “extension” view of modal
logics has become invalid. But consider an extended system like “modal
predicate logic”, which many people consider “obvious”. From a mod-
ern point of view, such a system is a potentially explosive combination
of diverse ideas: standard quantifiers over the object domain, restricted
local quantifiers over worlds, and also some (insidiously) hidden assumptions about how these two realms of objects and worlds are related. No wonder that the semantics of modal predicate logic has been under debate with both philosophers and mathematicians right up until today. We will bring the reader up-to-date in one of our chapters, showing how old debates between Kripke and Lewis on “trans-world identity” have returned in mathematics in the 1980s.

A related feature of research today is that tribal labels like “philosophical”, “mathematical” or “computational” logic mean less than they used to. Many topics in these lectures on modal logic cannot be classified as just one or the other, and this reflects intellectual realities. For instance, the modern study of rational agency and games combines fundamental insights from all these sources, without any particular pecking order. In our view, this is typical of logic: its themes migrate between academic fields, and in doing so, modify their initial agenda. But there is no reason to be pessimistic, the way some philosophers have the gloomy view that, once logic becomes technical, it leaves for good. Prodigal sons tend to return from their travels – though on the whole, a bit wiser than when they left. Some signs of such homecomings may be seen in current areas like formal epistemology and philosophy of action, and a number of illustrations will be found in these notes, when discussing logics of knowledge, interaction, and games.

The same is true for the popular division between “pure” versus “applied” logic, often misused as a label. Like any healthy discipline, logic generates theory in a process of reflection on applications, and this can go through many cycles. Indeed, in modal logic, fundamental theory has always a unifying force counteracting expansion. This shows at two places in these notes. Part II describes theory that arose out of reflection on basic developments in Part I. In Part III, we go on to describe a wide range of applications from the last decades, with a new round of modern theory in Part IV. And so it goes on and on.
Part I

Core Concepts
In this first Part, we discuss the major technical notions in modal logic, all stated for the basic language, but with a broader thrust for logic in general that will become clear as we proceed. Our emphasis is on propositional modal logic, and that for two reasons: (a) this is by now the dominant practice in the area, and (b) essential features of the modalities come out best on a weaker base. We may think of propositional modal logic as a system in between propositional logic and first-order predicate logic, core topics that the reader has probably studied in a first introductory course. But taking the modal view also throws new light on first-order quantification in the end – as will be explained in detail in the next part of these lectures.

Most chapters in this first part represent major logical themes by themselves – and they require at least one extended classroom session: expressive power, axiomatic deduction, completeness, and computational complexity. The latter topic is rather new in introductions to modal logic, and it might be skipped – though I personally feel that this material belongs to “what every educated student should know” these days. I have hesitated about also including correspondence theory as a core topic, but placed it in the next part eventually.
Basic language and semantics

2.1 Syntax of modal propositional logic

A logical formalism starts with a language, a system of patterns behind some practice of communication and reasoning. These patterns are formal and austere, but that is precisely why they highlight basic features of the phenomenon described, while also suggesting analogies across different situations. Our basic language has the following syntax:

Definition 2.1.1 (Basic modal language). Formulas are defined as follows. We first chose a basic set of unanalyzed propositions:

\[ AT := p, q, r \ldots \text{ plus } \top \text{ (“always true”) and } \bot \text{ (“always false”) } \]

Next, we define inductively how to construct further expressions, using the format:

\[ \varphi ::= AT \mid \neg \varphi \mid (\varphi \land \psi) \mid (\varphi \lor \psi) \mid (\varphi \rightarrow \psi) \mid \lozenge \varphi \mid \Box \varphi \]

Here is how one reads items on the last line: “all atoms are formulas”, “if \( \varphi \) is a formula, then so is \( \neg \varphi \)”, “if \( \varphi, \psi \) are formulas, then so is \( (\varphi \land \psi) \)” etcetera. The understanding is that formulas are all and only the syntactic strings arising from this recursive process in a finite number of steps. While this format, originally invented for defining programming languages, is more terse than the usual formulations in most logic textbooks, it is very perspicuous – and its brevity in ink and paper also helps save the tropical rainforest.

Remark (Notation). In many passages in these lectures, I will denote arbitrary propositions by proposition letters \( p, q, \ldots \) but sometimes also by capital letters \( A, B, \ldots \) or Greek symbols \( \varphi, \psi, \ldots \). This practice is not very consistent, but most readers should agree that it is nice to have different clothes to wear, depending on one’s mood.

There are many possible readings for the modality \( \Box \) (pronounced “box”), as we have hinted at in our Introduction: necessary truth,