What Is Logic?

Sara L. Uckelman
Department of Philosophy
Durham University
s.l.uckelman@durham.ac.uk

May 5, 2026

Contents

Contents **i**

1 What is logic? [last modified 11 Oct 21] **1**
1.1 Arguments . . . 3
1.2 Logic vs. a logic vs. logics . . . 4
1.3 Notation . . . 5
1.4 Acknowledgments . . . 9

I Formal Argumentation **11**

2 Aristotelian dialectics [last modified 22 Jan 19] **12**
2.1 The origins of logic . . . 12
2.2 Aristotle (384–322 BC) . . . 15
2.3 The inheritance of Aristotle . . . 16

3 Logic in medieval Western Europe [last modified 25 Oct 21] **18**
3.1 The 11th century . . . 18
3.2 The 12th century . . . 18
3.3 The 13th century . . . 21
3.4 Terminist logic . . . 24
3.5 *Obligationes* . . . 33
3.6 Fallacies, insolubles, and sophisms . . . 56
3.7 The 14th century . . . 60

4 Logic in ancient and medieval East and South-East Asia [last modified 27 Mar 26] **61**
4.1 Introduction . . . 61

4.2 Logic in Buddhist epistemology 62
4.3 Logic in ancient China 66
4.4 Analysis and discussion 70

II Classical Logics **73**

5 Introduction to Part II [last modified 22 Oct 20] **74**

6 Term logic [last modified 05 May 26] **76**
6.1 Language 77
6.2 Semantics 83
6.3 Proof theory: Natural deduction 89
6.4 Metalogical results 99
6.5 Summary 105

7 Propositional logic [last modified 17 Mar 26] **106**
7.1 Introduction 106
7.2 History 109
7.3 Language 112
7.4 Semantics 115
7.5 Proof theory 132
7.6 Soundness and completeness 150

8 Predicate logic [last modified 5 March 26] **156**
8.1 Introduction 156
8.2 Language 157
8.3 Semantics 171
8.4 Proof theory: General remarks 178
8.5 Proof theory: An axiomatic approach 179
8.6 Proof theory: Natural deduction for first-order languages 181

IIIExtensions of Classical Logics **207**

9 Introduction to Part III [last modified 4 Oct 18] **208**

10 Peano Arithmetic [last modified 5 Mar 26] **209**
10.1 Introduction 209

10.2 Language and semantics . . . 210
10.3 Recursive functions . . . 216
10.4 Proof theory . . . 222
10.5 Number theory in Peano Arithmetic . . . 228
10.6 Gödel numbers . . . 238
10.7 The incompleteness theorem . . . 247
10.8 Some background material . . . 248

11 Propositional modal logic: theory [last modified 02 Feb 26] 252
11.1 Introduction . . . 252
11.2 Language . . . 254
11.3 Semantics . . . 255
11.4 Proof theory . . . 271
11.5 Completeness . . . 312
11.6 Strange modal logics . . . 344

12 Modal logic: applications [last modified 19 Jan 21] 346
12.1 Epistemic logic . . . 346
12.2 Doxastic logic . . . 349
12.3 Agentive and deontic logic . . . 350
12.4 Modal term logic . . . 353
12.5 Provability logic and the logic of provability . . . 355

13 Temporal logic [last modified 01 Mar 21] 356
13.1 What is temporal logic? . . . 357
13.2 Minimal propositional tense logic . . . 361
13.3 Time and necessity . . . 368
13.4 Diodorus's master argument . . . 372
13.5 Metric propositional tense logic . . . 381
13.6 Future contingents . . . 385
13.7 Interval semantics . . . 386
13.8 *Since*, *Until*, and *While* . . . 396
13.9 Pure temporal logic . . . 399
13.10 Four grades of tense logical involvement . . . 400
13.11 The Event Calculus . . . 402
13.12 Further reading . . . 404

14 Quantified modal and temporal logic [last modified 01 Mar 21] 405

14.1 Introduction . . . 405
14.2 Philosophical issues . . . 406
14.3 Kripke-style quantified modal logic . . . 409
14.4 Prior's system Q . . . 416
14.5 Medieval quantified temporal logic . . . 417

15 Hybrid logic [last modified 26 Feb 19] 420

16 Dynamic logic [last modified 17 Feb 23] 427
16.1 Language . . . 428
16.2 Modeling epistemic uncertainty . . . 429

17 Counterfactuals [last modified 23 Nov 20] 432

IV Non-classical logics 433

18 Introduction to Part IV [last modified 4 Oct 18] 434

19 Intuitionistic and constructive logics [last modified 01 Mar 21] 435
19.1 Introduction to constructivity . . . 435
19.2 Propositional intuitionistic logic . . . 436
19.3 Other constructive logics . . . 440

20 Paraconsistent logics [last modified 13 Sep 18] 441
20.1 A return to conditionals . . . 441

21 Many-valued and fuzzy logics [last modified 12 Feb 26] 444
21.1 Many-valued modal logic . . . 444

22 Probabilistic logics [last modified 17 Jun 19] 447

23 Inductive and abductive logics [last modified 13 Sep 18] 448

V Back matter 449

A Exercises [last modified 12 May 25] 450
A.1 Exercises for Chapter 1 . . . 450
A.2 Exercises for Chapter 3 . . . 450

A.3 Exercises for Chapter 4 . . . 454
A.4 Exercises for Chapter 6 . . . 455
A.5 Exercises for Chapter 7 . . . 462
A.6 Exercises for Chapter 8 . . . 475
A.7 Exercises for Chapter 10 . . . 484
A.8 Exercises for Chapter 11 . . . 485
A.9 Exercises for Chapter 12 . . . 488
A.10 Exercises for Chapter 13 . . . 489

B Answers to Selected Exercises **491**

C Errata in [Goldstern and Judah, 1998] **496**
C.1 4.3 . . . 496
C.2 4.4 . . . 496
C.3 4.5 . . . 496
C.4 4.6 . . . 497
C.5 4.7 . . . 497
C.6 4.8 . . . 497
C.7 4.9 . . . 498
C.8 4.10 . . . 498

Index [last modified 27 Mar 26] **498**

Bibliography **501**

Chapter 1

What is logic? [last modified 11 Oct 21]

There are a plethora of introduction-to-logic textbooks out there: So why have I written another one? It is for exactly the same reason that there are so many of them already out there. The teaching of logic is a very personal matter, in that every person who does so has individual views on what are the important aspects to highlight, what every student taking a first-year logic course should come away knowing, what every student taking an upper-level logic course should come away knowing, what the scope of logic is, whether classical logic should be preferred to non-classical logic, whether there is One True Logic or many, in addition to more mundane concerns such as preferences about notation, often inherited from the teacher's own days as a student. A consequence of this is that every time I have taught logic, I have been dissatisfied with the available textbooks: None of them ever covered quite what I wanted them to, in the order and way I wanted them to, and with the notation I wanted. So rather than continuing to try to find the perfect textbook, I decided to write my own.

But I decided to do more than write yet another introduction-to-logic book. Instead, I wanted a book that I could use for both the intro course and the upper level course(s) at the same time. I wanted my first-year students to be given peeks into what could be built upon the foundations they were learning. I wanted my upper-level students to have all the basics available to them, in the same book, for reference when they were suddenly unsure whether they could, right at this very moment, perform universal generalization. Thus I do not expect to ever cover the entire book in one course, but hope that it can serve as a single book for a multi-year programme in logic.

I also wanted to write a book that could serve as something of a translation guide

between other textbooks. To that end, I have included extensive discussions of alternative notations, and a Glossary which serves as both a partial index to this book but also to other books. I hope that this will prevent me from rifling through my shelves in the middle of seminar to find a definition of Sahlqvist formulas because they have suddenly, tangentially, become relevant, and I do not have their form memorized off the top of my head.

The first question we must address is also the question that governs the book as a whole: What *is* logic? What is it that logicians do when they are "doing logic"? What is *a* logic? What are *logics*? Is there more than one logic? Or is there only one?

There are as many different answers to these questions as there are logicians (in fact, there are probably more answers than there are logicians, for some logicians are pluralists and accept more than one answer!), and unsurprisingly we will not provide a quick and easy answer here—after all, part of the point of this whole book is to address these questions.

Nevertheless, we have to start *somewhere*. That starting point will be the words of Roger Bacon, a medieval philosopher and theologian who studied and taught at Oxford and Paris, and later became a Franciscan monk. Writing around 1250, Bacon says in his *Art and Science of Logic* (an introductory textbook for undergraduates studying logic) that

> logic, as a science, is the habit of distinguishing what is true from what is false by means of rules or maxims or dignities by which we can comprehend the truth of a locution through our own efforts or with the help of others. And logic is so-called from '*logos*', which means discourse, and '*lexis*', which means reason or understanding — as it were, the science either of reason joined to discourse or of discourse joined to reason [Bacon, 2009, ¶3].

This definition highlights four important features of (the study of) logic:

1. It is aimed at distinguishing truth from falsehood.
2. It is rule-governed.
3. It can be a joint venture.
4. It involves discourse.

Contemporary logicians would probably stop Bacon after the first two, putting the emphasis on the truth-preserving, rule-governed nature of logic, and setting aside the discursive, cooperative aspects. One way in which this book differs from your standard 21st-century introduction to logic is that we will be emphasizing the importance of the discursive nature of logic: The purpose to which arguments are put has a bearing on the choice of appropriate logic.

1.1 Arguments

Keeping the preceding in mind, let us take as starting point the following definition of logic, which we will revise and precisify as we proceed:

Definition 1.1.1. *Logic* is the study of good arguments.

As a definition, this is an exemplary one; it is clear, simple, and elegant. But as an *explanation* of our subject matter, it is wanting, for we haven't yet said (1) what an argument is or (2) what a *good* argument is. It is easier to give a general, abstract account of what an argument is than it is to define which of these are good ones, and one of the things that we will see throughout the course of this book is that there is a plurality of ways in which "good" can be cashed out.

Let us begin, therefore, by giving a general definition of what we will count as an argument, and in doing so we will formally introduce our first symbolic notation:[1]

Definition 1.1.2. An *argument* is a pair consisting of a finite (and possibly empty) set Γ of sentences, $\varphi_1, \varphi_2, \varphi_3, \ldots$, called the premises, and a single sentence, ψ, called the conclusion.

Sometimes we will say that an argument is a set Γ containing a distinguished sentence, ψ, which is the conclusion, and the remainder of the sentences are the premises.

It should be noted that this is a very liberal account of argument, because it does not require any sort of connection between the premises and the conclusion. On this definition, the vast majority of arguments are going to be both (a) not recognizable as arguments on the ordinary usage of the term and (b) really, really bad.

This gives us a clue as to what makes an argument *good*: There must be some connection between the premises and the conclusion. In fact, all of the different ways of explicating "goodness" that we will explore in this book are rooted in this idea of a connection between the premises and the conclusion. The existence of such a connection is tied to the guaranteed or likely transmission of some special property from the premises of the argument to its conclusion; what this property is varies from explication to explication.

A good argument will then be one in which there is a special relationship between the members of the set Γ and the conclusion φ. Again as a starting point, let us take the following definition as a general account of good arguments:

Definition 1.1.3. A *good* argument is one in which the conclusion is a consequence of the premises.

[1] If you are not familiar with the Greek alphabet, you can refresh yourself in §1.3.

This definition too is only a starting point because it doesn't specify in what way the conclusion is a consequence of the premises, and it is by varying the type of consequence that we can come up with different accounts of "good argument". We will be covering a variety of types of consequence in this book, including:

- logical[2]
- probabilistic
- inductive
- dialogical
- epistemic

Because of the variety of ways in which 'good' is defined, we will give specific definitions for each context we investigate in this book.

1.2 Logic vs. a logic vs. logics

We can now distinguish *logic*, as a field of study—the study of good arguments—from *a logic*, which is a specific system of argumentation designed to make explicit certain features of the arguments of that system. One such feature that is often highlighted is the form or structure of a given argument, as there is often a close link between the goodness of an argument and the structure or form that it has. Individual logics will each be accompanied with mechanisms which allow us to make this form or structure explicit, and we will further use these mechanisms to prove general results about the logics (systems of argument) in question. These logics differ in the features that they highlight, and the contexts in which they are applied, but each logic can be characterized by three things: (1) its language, (2) its proof theory, and (3) its semantics. Thus, for every logic that we consider in this book, we will begin by identifying:

1. The *language*, that is, the set of logical and non-logical symbols and the ways in which they can be combined into syntactically (i.e., grammatically) correct strings.

[2]This is a poor choice of word, because it has the implication that the other types of consequence are *not* logical, but this goes against our view of logic as being the study of good arguments, and these good arguments can involve any of a variety of types of consequences. The reason for using this word here is primarily historical: If one thinks that the only good arguments are the ones where the strongest notion of consequence holds between the premises and the conclusion, then logic is the study of that notion of consequence alone, and hence that notion can be called 'logical'. In this book, we take a much more pluralistic approach to what counts as 'good', and hence what falls under the study of logic.

2. The *proof-theory*, that is, a system of formal proof, or the ways in which strings of symbols can be manipulated in that system.

3. The *semantics*, that is, what it is that gives meaning to the logical and non-logical symbols.

As soon as we have a language and a proof-theory, we are able to prove results at two levels: At the *object level*, we use a system's proof theory to prove results *within the system*. At the *meta level*, we reason in English, outside of the formal system, to prove results *about the system*, in particular, about the relationship between the proof-theory and the semantics.

An important goal of this book is to introduce even the novice logic student to the distinction between the object- and meta-level and the methods of proof which are used in both. Once an English-language argument has been translated into a particular logical language, we can characterise the goodness of the argument via either proof-theoretic or semantic means (cf. Table 1.1). In the best case scenario, we will give both types of characterizations, and prove that these two characterizations coincide, i.e., that every argument which is "good" according to the proof-theoretic characterization is also "good" according to the semantic characterization (when this is the case, we say that the proof theory is "sound"), and that every argument which is "good" according to the semantic characterization is also "good" according to the proof-theoretic characterization (when this is the case, we say that the proof theory is "complete".[3])

consequences	derivability
truth	representation
semantics	syntax
models	rules of inference

Table 1.1: The two faces of logic

1.3 Notation

In this section, we introduce and define notation that will be used throughout this book, as well as notation used in other logical textbooks to facilitate translation between texts.

[3]The notions of soundness and completeness are introduced in more detail in §7.6.

1.3.1 The Greek alphabet

Upper and lower case Greek letters are often used to stand for arbitrary logical sentences, or sets of sentences. The alphabet is reproduced in Table 1.2, and the names and forms of the letters should be mastered as soon as possible.

Name	**Upper case**	**Lower case**
Alpha	A	α
Beta	B	β
Gamma	Γ	γ
Delta	Δ	δ
Epsilon	E	ϵ
Zeta	Z	ζ
Eta	H	η
Theta	Θ	θ
Iota	I	ι
Kappa	K	κ
Lambda	Λ	λ
Mu	M	μ
Nu	N	ν
Xi	Ξ	ξ
Omicron	O	o
Pi	Π	π
Rho	P	ρ
Sigma	Σ	σ
Tau	T	τ
Upsilon	Υ	υ
Phi	Φ	φ
Chi	X	χ
Psi	Ψ	ψ
Omega	Ω	ω

Table 1.2: The Greek alphabet

1.3.2 Logical notation

Notation is not fixed across logical and philosophical literature. Here we give a brief summary of the different types of notation and the symbols used. Operators that take one argument are called *monadic* or *unary* operators. Operators that take two arguments are

called *dyadic* or *binary* operators. The standard logical operators and connectives are all either monadic or dyadic. Negation (§7.3), universal and existential quantification (§8.2), and necessity and possibility (§11.2) are all monadic operators. Conjunction, disjunction, implication, and equivalence (§7.3) are all dyadic operators.

Infix

The most common type of notation is *infix* notation, wherein the binary connectives are inserted between the formulas that they connect. In the list below, the symbol in red is the symbol used in this book.

- Negation: $\neg\varphi, {\sim}\varphi, \bar{\varphi}$.
- Conjunction: $\varphi\wedge\psi$, $\varphi\&\psi$, $\varphi \cap \psi$, $\varphi \bullet \psi$.
- Disjunction: $\varphi\vee\psi$, $\varphi \cup \psi$.
- Implication, material: $\varphi\rightarrow\psi$, $\varphi \supset \psi$.
- Implication, strict: $\varphi \to \psi$, $\varphi \strictif \psi$.
- Universal quantification: $\forall x\varphi$, $(\forall x)\varphi$, $(x)\varphi$.
- Existential quantification: $\exists x\varphi$, $(\exists x)\varphi$.
- Necessity: $\Box\varphi$ (generic necessity, logical necessity), $L\varphi$ (generic necessity, logical necessity), $K\varphi$ (epistemic necessity), $O\varphi$ (deontic necessity), $\delta\varphi$ (agentive necessity), $G\varphi$ (future necessity), $H\varphi$ (past necessity).
- Possibility: $\Diamond\varphi$ (generic possibility, logical possibility), $M\varphi$ (generic possibility, logical possibility), $F\varphi$ (future possibility), $P\varphi$ (past possibility).

Occasionally, other notation will be introduced in the book as needed.

Prefix/Polish

Less common is the so-called Polish notation, or prefix notation, where the connectives come before the propositions they connect. Arthur Prior (1914–1969), the father of modern temporal and tense logic, was heavily influenced by the Polish logicians, and wrote in Polish notation, so it is necessary to be able to read Polish notation if one wishes to read Prior's early work.

- Negation: $N\varphi$.
- Conjunction: $K\varphi\psi$.
- Disjunction: $A\varphi\psi$.
- Implication: $C\varphi\psi$.
- Universal quantification: $\Pi p\varphi(p)$.
- Existential quantification: $\Sigma p\varphi(p)$.
- Necessity: $L\varphi$.
- Possibility: $M\varphi$

Note that here there is not a clear distinction between propositions and predicates, as there is in ordinary infix notations for propositional and first order logic. We can think of propositions as 0-ary predicates.

1.3.3 Set theoretical notation

This book will not cover much set theory, explicitly, but we will make liberal use of set theoretic notation and concepts. The following provides a quick and dirty introduction to these concepts and notation:

Sets are defined by their members; if two sets contain exactly the same items, then they are the same set. Sets can be represented arbitrarily with capital Greek letters such as Γ, Δ, Θ, Λ, or explicitly, with their members indicated between { and }.

Definition 1.3.1. A set is *finite* if the number of members in it is one of the ordinary counting numbers, e.g., 1, 2, 3, etc.

Definition 1.3.2 (Empty set)**.** There is one unique set which contains no members, the empty set, represented with $\varnothing$.

Given the two ways to represent sets written above, the empty set can also be indicated by $\{\,\}$.

Definition 1.3.3 (Set membership)**.** If x is an element of a set Γ, we write $x \in \Gamma$; if x is not an element of a set Γ, we write $x \notin \Gamma$.

Definition 1.3.4 (Subset)**.** If every member of Γ is also a member of Δ, then Γ is a subset of Δ and we write $\Gamma \subseteq \Delta$. If $\Gamma \subseteq \Delta$ and there is at least one element in Δ which is not in Γ, then Γ is a proper subset of Δ, and we write $\Gamma \subset \Delta$.

Definition 1.3.5 (Union)**.** The union of two sets Γ and Δ, written $\Gamma \cup \Delta$, is the set containing anything that is a member of either Γ or Δ.

Definition 1.3.6 (Intersection)**.** The intersection of two sets Γ and Δ, written $\Gamma \cap \Delta$, is the set containing anything that is a member of both Γ and Δ.

Lemma 1.3.7. *If* $\Gamma \subseteq \Delta$*, then* $\Gamma \cap \Delta = \Gamma$.

Proof. If $\Gamma \subseteq \Delta$, then every element of Γ is also an element of Δ. It then follows that the set of all things that occur in both Γ and Δ is simply Γ itself. □

Definition 1.3.8 (Set difference)**.** The set difference between any two sets Γ and Δ, written $\Gamma \setminus \Delta$ or $\Gamma - \Delta$, is everything that is a member of Γ but not of Δ. If $\Gamma \subseteq \Delta$, then $\Delta \setminus \Gamma$ is the complement of Γ with respect to Δ.

Examples of these definitions can be found in Table 1.3. In this table, $\Delta \subseteq \Gamma$, $\Theta \subseteq \Gamma$, $a \in \Delta$, $a \in \Gamma$, $a \notin \Theta$.

Set	**Members**
Γ	$\{a, b, c, x, y, z\}$
Δ	$\{a, b, c\}$
Θ	$\{x, y, z\}$
$\Gamma \cup \Delta$	$\{a, b, c, x, y, z\}$
$\Delta \cup \Theta$	$\{a, b, c, x, y, z\}$
$\Gamma \setminus \Theta$	$\{a, b, c\}$
$\Gamma \cap \Delta$	$\{a, b, c\}$
$\Delta \cap \Theta$	$\varnothing$

Table 1.3: Some example sets

Definition 1.3.9 (Ordered pair)**.** The *ordered pair* $\langle x, y \rangle$ is the set $\{x, \{x, y\}\}$.

1.4 Acknowledgments

This book arose from lecture notes for the courses "Core Logic" (Universiteit van Amsterdam, Master of Logic, fall 2007, fall 2008, fall 2009), "Temporal Logic" (Universiteit

van Amsterdam, Master of Logic, spring 2010), "Philosophical Logic" (Durham University, Philosophy BA, 2014–15), "Formal & Philosophical Logic" (Durham University, Philosophy BA, 2015–16; 2016–17, 2017–18, 2018–19, 2019–20, 2020–21), "Introduction to Logic" (Durham University, Philosophy BA, 2016–17, 2017–18), "Introduction to Logic in the Middle Ages" (3rd Nordic Logic Summer School, 2017), "Logic II: Temporal Logic" (Università della Svizzera Italiana, Philosophy MA, 2019), and "Fundamentals of Logic" (Durham University, Philosophy BA, 2019–20, 2020–21, 2021–22, 2022-23, 2024–25, 2025–26[4]). As such, this entire book bears a great debt of gratitude to every single one of my students in these classes, who asked questions, caught typos, and made it all worthwhile. I'd also like to thank Nick Battle, who read the entirety of the book wholly on his own volition and sent many comments.

A special debt to Benedikt Löwe must also be acknowledged, due to its enormity. He designed the module "Core Logic" which I took as a first-year PhD student in Amsterdam, and which opened up my world to the history of logic and shaped the path of my dissertation research. I had the pleasure of being the teaching assistant for this class in later years, and eventually running it myself. The extensive notes and slides that he prepared for this class were invaluable when I was teaching, and much of what is currently Chapters 2 and 3 and the system for syllogistics in Section **??** is due to him.

[4]This class in particular caught many typos and formatting errors in the exercises at the end of the book, so especial thanks to them!

Part I

Formal Argumentation

Chapter 2

Aristotelian dialectics [last modified 22 Jan 19]

2.1 The origins of logic

Logic as a formal area of study did not develop out of a vacuum, but can be seen as a natural outgrowth of two distinct origins:

- Greek mathematics
- Greek rhetoric: "Eristic" and "Sophistry"

Pre-Greek mathematics, such as Egyptian and Babylonian mathematics, was not primarily concerned with proof, but more with computation. For them, geometry was *measurement of the earth*.[1] Greek mathematics broke with this tradition by developing the notion of/method of proof. It is often claimed that the first proof was given by Thales of Miletus (c. 625–c. 546 BC). The proof was either that the base angles of an isosceles triangle are equal or that the diameter of a circle halves the circle.[2] Whether or not this story is true, it is exemplary of a certain change in views towards mathematics. The Greeks recognized that there are certain things which cannot be taken as obviously or self-evidently true, but for which need you need to give a proof or argument.

The next important figure in the history of mathematics is Pythagoras (c. 569–c. 475 BC), still famous today for the theorem named after him. Pythagoras was born in Samos,

[1] Saying that the focus of pre-Greek mathematics was *computational* is not to say that it was not *theoretical*, e.g., that it was only applied or practical.

[2] Davis/Hirsch. Proclus comments on this, but it is not clear what it means, as we don't know what 'proof' means here. Panchenko discusses this in detail.

Ionia, and was influenced by Thales and Anaximander. He traveled widely (including to Egypt), and in 518 he founded a philosophical school in Croton, in the eastern part of the 'heel' of southern Italy. The school consisted of an inner circle of followers, called the mathematikoi, and an outer circle, called the akousmatics. The mathematikoi were permanent residents of the society, adhering to a life of vegetarianism and no personal possessions. The mathematikoi were taught directly by Pythagoras, and believed the following:

1. that at its deepest level, reality is mathematical in nature,
2. that philosophy can be used for spiritual purification,
3. that the soul can rise to union with the divine,
4. that certain symbols have a mystical significance, and
5. that all brothers of the order should observe strict loyalty and secrecy.

The akousmatics followed a less strict regime; they lived in their own houses, were allowed to have possessions, and did not have to be vegetarians. They came to the Society for instruction during the day.

We do have details of Pythagoras's life from early biographies which use important original sources yet are written by authors who attribute divine powers to him, and whose aim was to present him as a god-like figure. There is fairly good agreement on the main events of his life but most of the dates are disputed with different scholars giving dates which differ by 20 years.

Theaetetus (c417–c369 BCE was a student of Socrates.

Euclid (c. 325–c. 265 BC) wrote a compilation of mathematical knowledge which became the best-selling textbook of all time, his *Elements of Geometry*, and which was still used as a first course in geometry at the end of the 20th century, more than two thousand years after it was first written. We know very little about his life, despite the tremendous importance of his work.

Two aspects specifically characterize the use of proofs in Greek mathematics: The axiomatic method and proof by contradiction. In the axiomatic method, you take basic statements which are obviously true (these are called axioms), and then add rules of derivation. With these rules, you derive consequences from the axioms. Of course, the question is "what counts as obvious?" The 'Greek turn' in mathematics was recognizing that many statements which seem obvious in fact admit of proof. So it is at least superficially problematic to say that there are some obvious statements which be taken without proof.

Nowadays, proof by contradiction, or *reductio ad absurdum*, is considered to be a standard part of mathematical proof practice. However, it actually arose in a non-mathematical

context. The first examples of proof by contradiction are found in rhetoric, not math. However, by the time of Euclid, it had become standard mathematical practice.

We now give an example of a proof by contradiction, which is probably familiar to most readers:

Claim 2.1.1. $\sqrt{2}$ is not a fraction of integers.

Proof. Suppose it were. Then there are integers n and m without common divisor such that

$$\sqrt{2} = \frac{n}{m}.$$

But then

$$2m^2 = n^2.$$

In particular, n must be even. But then n^2 must be divisible by 4, and so m must be even. This is a contradiction (since we assumed that n and m have no common divisor), □

This proof was known to the Pythagoreans, and it caused them great consternation, given their belief that everything is number, or numerable.

We now turn to the other origin of logic, that is, rhetoric. Rhetoric itself developed out of two different areas, the Dialectic method, and argumentation in every day life.

The dialectic method is best exemplified by Socrates (469–399 BC), with his works of *elenchus* (*refutatio* in Latin) and *diaresis* (*divisio* in Latin). This is also exemplified by the system of proof by contradiction in mathematics, as discussed above, and in the works of Zeno of Elea[3] (c. 490–c. 425 BC), whose paradoxes are still well known today. These paradoxes are the paradoxes of the arrow, of Achilles and the turtle.

The Megarians were also dialecticians.

The other root of rhetoric, argumentation in everyday life, can be found in the following:

- Sophists
- Public disputations according to rules for questioner and answerer
- Plato, *Euthydemus*
- Aristotle, *Topics* and *Rhetoric*

In the Euthydemus, Plato discusses the sophists, and gives a very negative view of them. The *Topics* of Aristotle combine the logical systems found in the Prior and Posterior Analytics with what can be called argumentation theory. The *Rhetoric*, on the other hand, was almost purely pragmatic.

[3]Possibly adopted son of Parmenides (Diog. Laert.)

2.2 Aristotle (384–322 BC)

Aristotle was the son of the personal physician of Amyntas III, king of Macedonia. In 367 BCE, at the age of 17, he joined Plato's Academy in Athens. By this time the Academy had been in existence for 20 years, though when Aristotle arrived Plato was in fact not there, but on his first visit to Syracuse.

During Plato's absence, the Academy was administered by Eudoxus of Cnidos. Other teachers at the time were Plato's nephew Speusippus and Xenocrates of Chalcedon. Once Aristotle ceased being a student, he remained at the Academy for twenty years as a teacher. We have little information about what he taught while he was at the Academy.

It is likely that Aristotle left the Academy when Plato died in 347 BCE and Speusippus assumed the leadership, since he was opposed to Speusippus's views. Additionally, though he was put forward in the vote for the new head of the Academy, he did not succeed. After leaving the Academy, Aristotle worked at the court of Macedonia, which had signed a peace treaty with Athens in 346 BCE, from 343 to 336 BC.[4]

When Speusippus died in 340 BCE, Aristotle was again put forward was a potential new head of the Academy, and was again not elected. It was perhaps this that cause him to create his own Academy, the Lyceum, in 335 BCE under the patronage of Antipater, the Macedonian proconsul. The Lyceum was situated in the eastern part of Athens on the banks of the Iliuss. The Lyceum was a gymnasium with connected covered walks, each walk called a peripatos (giving rise to the term "peripatetic", originally just a word for people who walked around.[5])

When Alexander the Great died in 323 BCE, Aristotle retired to Chalcis because of the growing anti-Macedonian sentiment in Athens. He died the following year of a stomach complaint, aged 62. His works were given to the Dictator of Pergamon [Confirm this; my notes are unclear], who buried them. They were rediscovered under Sulla [get dating], and compiled by Andronicus. As a result, quite a few of the works are unclear, gappy, or confused. Andronicus also assigned the chapters and the titles, often based on contingent factors such as the original order of the writings. This resulted in the *Posterior Analytics* being named such not because they were written after the *Prior Analytics* but because they came after in the manuscripts, and also in the introduction of "metaphysics" as a separate discipline, where the term literally means "[material occurring] after the *Physics*".

[4] It is often reported that Aristotle was the tutor of the young Alexander the Great while he was at Philip's court. This is almost certainly a later invention.

[5] I would LOVE to be able to include this here: `http://www.transmogrifier.org/ch-img/ch930915.gif`!

Aristotle's logical works often go under the collective title of the *Organon* ("Works"). The *Organon* is composed of:

- **Categories**: Classification of types of predicates
- **On Interpretation** (Lat. *De interpretatione*, Gk. *peri hermeneis*): Basics of philosophy of language, subject-predicate distinction, Square of Oppositions
- **Prior Analytics**: Syllogistics
- **Posterior Analytics**: More on syllogistics; scientific reasoning.
- **Topics**: Logic except for syllogistics
- **On Sophistical Refutations** *Lat. (De Sophisticis Elenchis)*: Fallacies

The Square of Oppositions, as found in Aristotle's *De interpretatione*, can be seen in Figure 6.2, and it is discussed further in Chapter 6.

2.3 The inheritance of Aristotle

In this section, we trace the paths of the reception of Aristotle by later logicians.

To be added: A section on post-Aristotelian commentators and his reception in classical and post-classical Greece, in the Roman empire, and in late antiquity.

In western Europe, the most influential person in the transmission of Aristotle to future generations was Anicius Manlius Severinus Boëthius, c475–524. Martin Grabmann described him as the last of the Roman philosophers, and the first of the scholastic theologians [Grabmann, 1909].

Boëthius was a profligate writer, composing theological, philosophical, logical, and arithmetic treatises.

- theological treatises
- *Consolation of philosophy*
- commentaries on Aristotle
- *On hypothetical syllogisms*
- translations of Aristotle
  - *Categories*

– *On interpretation*
– *Prior Analytics*
– *Topics*
– *Sophistical refutations*

Chapter 3

Logic in medieval Western Europe [last modified 25 Oct 21]

3.1 The 11th century

Until 1100, only a few ancient sources in logic and grammar were available to the Latin west:

- Aristotle, *Categoriae* (trans. Boëthius).
- Aristotle, *De interpretatione* (trans. Boëthius).
- Porphyrios, *Isagoge*.

These make up what is called the *Logica Vetus* or the "old logic".

3.2 The 12th century

Over the course of the 12th century, most of the remaining texts of Aristotle became available in Latin, started with the rediscovery c. 1120 of Boëthius' translations of

- Aristotle, *Prior Analytics*.
- Aristotle, *Topica*.
- Aristotle, *Sophistici Elenchi*.

The *Topics* and the *Sophistical Refutations* deal with argumentation theory, interaction, and convincing people.

The rediscovery of the *Sophistical Refutations* sparked lively trade in the study of fallacies and fallacious reasoning in the 12th century:

- 12th century. *Parvipontani* (Adam of Balsham), Petit Pont in Paris. *Fallacie Parvipontane.*
- *Quaestiones Victorinae* (school of William of Champeaux, 1100–1150).

Around 1150: James of Venice translates

- Aristotle, *Posterior Analytics.*
- Aristotle, *De anima.*
- Aristotle, *Metaphysica.*

Both *De anima* and the *Metaphysics* caused problems because the theories in them were not in line with Christian dogma. The pink texts make up the *Logica Nova* or the "new logic" [Dod, 1982].

Peter Abelard was born in 1079 to a noble family. He decided in favour of a clerical career against the will of his noble father. In 1094, he left home to study logic and dialectics under Roscelin of Compiègne. Later, Abelard went to Paris, where he studied under William of Champeaux and participates Public debates during lectures on universals [Beonio-Brocchieri Fumagelli, 1969, p. 1].

- 1102–1111. Abelard teaches in Melun, Corbeil, Paris.
- 1111–1113. Abelard goes to Laon to study theology with Anselm of Laon.
- 1113–1118. Abelard becomes the mentor of Héloise (1100–1163), the daughter of a rich Parisian merchant. They have a child, Astrolabe, and marry.
- 1118. Héloise's uncle Fulbert hires thugs who castrate Abelard. Abelard becomes a monk at St. Denis, Héloise a nun at Argenteuil.
- 1118–1136. Abelard lives as a monk.
- 1121. First condemnation (Council of Soissons).
- 1132. Wrote his autobiography, *Historia Calamitatum Mearum* (A history of my calamities).

- 1136–1140. Abelard returns to teach in Paris.
- 1140. Second condemnation (Council of Sens).
- 1142. Abelard dies on the way to Rome.

His approach to theology was branded as too philosophical. There are questions about why he returned to teaching after having lived as a monk for so long; was it really because of monetary issues? By 1136 Abelard was [get rank] at [get place], and so would have had a fairly substantial income.

Abelard was a brilliant and innovative logician. There are four aspects of his logic that are of especial interest to the modern logician:

- **The Square of Oppositions.** Discussion of the existential content of universal statements: Does *"Omnis homo est albus"* imply that there is a man?
- **Modal Logic.** Distinction of modal statements into *de re* and *de dicto*.
- **Temporal Logic.** Isolation of the concept of "true at a time".
- **Propositional Logic.** Theory of conditionals.

For more information, see [Tweedale, 1982].

***quidam non* vs *non omnis*.** Abelard notices that the Aristotelian square of oppositions includes "existential import":

"Every B is A" implies "Some B is A",

so he reads "*omnis homo est albus*" as "there are men and all of them are white". Therefore, Abelard distinguishes between "*Non omnis homo est albus*" ("either there are no men or there are non-white men") and "*Quidam homo non est albus*" ("there is a non-white man").

***de re* vs *de dicto*.** Abelard notices the modal logic distinction *de re* versus *de dicto*. ("*expositio per divisionem*" and "*expositio per compositionem*"), thereby developing a way to understand the Aristotelian "two Barbaras" problem, where NAN Barbara is valid under one reading and ANN Barbara under the other. [See Catarina's article; *de re/de dicto* is not the same distinction as *in sensu compositu/in sensu diviso*.] However, he claims that *de dicto* modalities are not real modalities. (This changes in Pseudo-Aquinas, *De modalibus*.)

True at time t. One can isolate the problem in the Master Argument by noting that there is a confusion in it about tense and time. Not all statements in past tense are necessarily true, for example: "Socrates did not talk to Plato." One attempt of a solution

is to introduce a semantics of tense sentences that allows **truth at a time**. This was one step in the direction of modern temporal logics. Tense is accidental, but time is necessary.

Conditionals. A new (intensional) view of propositional logic: *Si non est A est B*" is equivalent to "*Aut est A aut est B.*"

$$\neg A \to B \leftrightarrow A \vee B?$$

Abelard reads "*Si est A est B*" as "necessarily, A implies B", and thus has a different reading of the disjunction as "necessarily, $\neg A$ implies B".

3.3 The 13th century

The rediscovery of these new texts combined with the development of the university system made the 13th century an exciting and vibrant period in philosophical, logical, and theological thought:

Resources Rediscovery of Aristotle leads to a lot of new material.

Institutions The centers of learning (Paris, Oxford, ...) institutionalize learning in the Universities.

Consolidation of the 13th century The great philosophers and theologians worked to embed Aristotelian teaching into the Christian philosophy.

Two of the most important people in this consolidation were Albert the Great and Thomas Aquinas:

(Saint) **Albert the Great**
Albertus Magnus (c. 1200–1280)
Doctor Universalis.
Founder of the *studium generale* in Cologne (1248), a predecessor to the current University of Cologne.
Predecessor of modern concept of sciences: *The aim of natural science is not simply to accept the statements of others, but to investigate the causes that are at work in nature.*
(Saint) **Thomas Aquinas** (1225-1274)
Student of Albert the Great.
Doctor Angelicus.

The 13th century was the Golden Age of Scholasticism. It is in this century that we see a switch in emphasis from "*the* correct form of reasoning" to "correct *forms* of reasoning", and

we see, in addition to reasoning and analysis (involving logic, metaphysics and semantics), based on authorities (philological and logical analysis of original texts), the development of specific forms of reasoning such as *quaestiones* and *disputationes*. These forms of reasoning play a similar rôle to that played by author instructions today: They give the formal requirements that a piece of reasoning must meet in order to be acceptable to the desired audience. In the early middle ages, there was no such thing as a publication culture; this only developed with scholasticism in the high middle ages.

The 13th century is also when we see the development of what is now called the *logica nova* 'new logic'. The *logica vetus* 'old logic' consisted of syllogisms and the Categories. The new logic comprised four main new branches of logic:

- *insolubilia*: fallacies and paradoxes.
- *syncategoremata*: and, or, not, if, every, some, only, except.
- *obligationes*: a game-theoretic approach to logic.
- "Terminist logic": *proprietates terminorum*. a predecessor of Leibniz's ideas.

We will discuss each of these in turn, but first we will give a brief overview of some of the important logicians in the 12th to 14th centuries:

- John of Salisbury (c. 1115–1180), *Metalogicon* (1159). The *Metalogicon* is very much in the spirit of the *logica nova*, being a mix of philosophy of language, logic, and metaphysics. It was written as a defense of logical teaching in grammatical disciplines. John was a student of Abelard, and later Thomas Beckett's secretary and eye-witness of his murder. Beckett had been made a priest the night before he was made Archbishop of Canterbury (he was previously the chancellor of King Henry of England).

- Robert Grosseteste (1168–1253). He was the teacher of Roger Bacon, and is often called the father of Philosophy of Science.

- William of Sherwood (1190–1249), teacher of Petrus Hispanus, *Introductiones in Logicam*. Discussions of propositional logic and the Square of Oppositions. His is also the earliest surviving text that has the syllogism mnemonics.

- Petrus Hispanus[1], *Summulae Logicales*. His *Summulae* was immensely popular, which fact is attested by the large number of surviving manuscripts and early printed editions. The text was still in use as a primary textbook in the 17th century.

- Robert Kilwardby (c. 1215–1279). Proofs of syllogistic conversion rules as syllogisms with two terms. He was Archbishop of Canterbury from 1273–1279. Cf. [Thom, 2007].

- Roger Bacon (1214–1292). A proponent of empiricism; inspired by Arabic science, and later called the *Doctor mirabilis*.

- Raimundus Lullus (Raymond Lull, Ramon Lull) (c. 1235–c. 1315). His *Ars magnus* is of interest because it contains discussions of mechanical computation of concepts. He was the first to attempt to automatize and mechanize reasoning, with words inscribed on metal circles. He was a pioneer in computation, and later on Leibniz refers to him, as his views also included the view of universal language.

- Johannes Duns Scotus (1266–1308). *Doctor Subtilis*. Duns Scotus was not so much a logician as a philosopher and theologian. He wrote extensively on epistemology and causality, and gave proofs for the existence of God via causal principles. He also introduced the idea of applying logic to reasoning about the trinity, via the concept of *distinctio formalis*.

- Pseudo-Scot: Wrote on modalities such as *dubium*, *scitum*, *opinatum*, *volitum*, *dilectum*. In 1639 Waddingham thought Pseudo-Scot was Scotus, and his works were included in Scotus compilations.

- William of Ockham (c. 1295–1349). *Entia non sunt multiplicanda praeter necessitatem.* He was a proponent of nominalism. His views were denounced. He was at the papal court in Avignon for 2–3 years. [See SEP article.] On 26 May 1328 Ockham flees to Pisa in northern Italy where he meets the German emperor. He later moves to Münich, and becomes increasingly involved in politics, leaving logic, philosophy, and theology behind.

- Golden Age of Terminist Logic: 1175-1250.

  – *Ars Meliduna* (1170–1180).

[1] Not to be confused with Pope John XXI, (c. 1205–1277).

  - *Tractatus Anagnini* (1200-1220).
  - William of Shyreswood (Shireswood/Sherwood, 1190–1249): *Introductiones in Logicam* (c. 1230–1240).
  - Petrus Hispanus (Pope John XXI.; c.1205-1277): *Summulae Logicales* (c. 1230–1240).

- **Oxford School.**
  Influenced by the *Parvipontani*.
  emphMain representative. William of Shyreswood.
- **Paris School.**
  Main representative. Petrus Hispanus.

- Geoffrey of Hapshall (c. 1270).

The modists opposed nominalism. Their different modes (*essendi, intelligendi, significandi*) are all different correspondence between words and reality. [Get list of modist names.] They were interested in the compositionality of semantics, and the fact that meaning cannot be wholly isolated from sentential context.

3.4 Terminist logic

Terminist logic developed out of the move from the analysis of meaning in words (what does *homo* mean?) to the analysis of meaning of terms in phrases (what part of the meaning of *homo* is responsible for the fact that "*omnis homo mortalis est*" is true?). Its origins were in the school of Chartres (c. 1030) and the 'contextual approach' to language and logic (cf. de Rijk, 1967). [Get full citation here.] Terminist logic was born out of the recognition that syllogistics, linguistic analysis, and grammar do not tell the complete story of language and truth:

- Syllogistics doesn't analyse the truth-status of categorical propositions any further.
- Linguistic analysis (predication vs non-predication) at the basis of the theory of categories.
- Grammar investigated the meaning of single words (outside of the context of propositions).

Some of the questions that the Terminists were interested in are very subtle:

- Compare *"homo est animal"*, *"homo est species"*, and *"homo est disyllabum"*. In each of the cases, the meaning of *homo* is slightly different.
- What do qualifiers do with meanings? If I go from *"omnis homo est philosophus"* to *"paene omnis homo est philosophus"*, how does the explanation for the meaning change? What is *paene*? Is it a sentential operator? An operator on *omnis*?

Two very important concepts that the Terminists made use of are *significatio* and *suppositio*. We will discuss both of these in turn.

3.4.1 Signification

The roots of the theory of signification (*significatio*) can be found in Aristotle and Boëthius. For these two, terms signify by establishing an understanding. Signification has a causal component. These lead to the triad of written / spoken / mental language:

- Written language signifies spoken language
- spoken language signifies mental language
- mental language signifies the things.

A term's *significatio* is determined by *impositio*, i.e., the word's original application (baptism). This notion is due to Priscian: *proprium est nominis significare substantiam cum qualitate*: *philosophus* signifies "human with the quality of being a philosopher".

Bacon says that "a signifying utterance is one by which every animal communicates something to all or some of its species" [Bacon, 2009, ¶118]. He distinguishes between natural signification and signification *ad placitum*, which is sometimes translated as 'at one's pleasure' or 'conventionally'. Of words that signify *ad placitum*, nouns are those which signify without the addition of time, while verbs are those which signify with time—that is, verbs are essentially tensed.

Lambert says that "the signification of a term is the concept of a thing, for which concept of a thing a vocal sound is imposed *ad placitum* of the one instituting it" cite. For Lambert, four things are required for signification:

- A thing.
- A concept of the thing.

- A vocal sound.
- A union of vocal sound with concept.

Peter of Spain says that "the signification of a term is the conventional representation of a thing by an utterance" cite, and that every signifying term signifies either universally or partially. There are two types of signification: Signification as a substance and signification as a modifier.

Finally, Sherwood says that "signification is the presentation of the form of something to the understanding" cite.

Signifying terms can be divided into two categories: categorematic and syncategorematic. Generally speaking, categorematic terms are terms that signify on their own (this definition goes back to Cassiodorus get reference), and terms that only *consignify*, that is, they only signify when combined with categorematic terms, are called *syncategoremata*. However, it is important to recognize that there is no single unified medieval tradition; different authors approached the distinction in different ways. These different ways can, for the most part, be classified into three categories: syntactic, semantic, and mixed. Most modern authors, e.g., [Braakhuis, 1981, pp. 141–142], identify only the first two. However, there are versions which do not clearly fit into either, but instead draw upon both syntactic and semantic elements, often relying on the medieval notion of supposition (see §3.4.2). These suppositional or mixed approaches in fact turn out to be the most interesting ones.

The syntactic approach hearkens back to the origin of the distinction as a grammatical one, introduced in Priscian's *Institutiones grammaticae*.[2] Kretzmann in the *Cambridge History of Later Medieval Philosophy* presents the syntactic distinction on the basis of how terms can be used in sentences:

> Any word that can be used alone as a subject term or as a predicate term is classifiable as a categorematic word; all other words are classifiable as syncategorematic words, those that can occur in a proposition, whether categorical or hypothetical, only along with at least one properly matched pair of categorematic words [Kretzmann, 1982, p. 211].

Braakhuis, in describing this tradition, describes syncategorematic terms as ones "that bear on the interrelation between subject and predicate of a proposition" [Braakhuis, 1981, p. 142]; the implication is then that the terms which can be subjects or predicates are categorematic.

[2] Though while Priscian was indeed strongly influential on the medieval development of the distinction in logical contexts, it would be a mistake to take his grammatical definition as representative of the logical definition [Kretzmann, 1982, pp. 212–213].

Paul of Venice, writing in the late 14th/early 15th century calls this 'the common definition', and summarizes as follows:

> A syncategorematic term is that which, taken as significant, cannot be the subject or the predicate, or a part of the distributed subject or predicate, of a categorical proposition [of Venice, 1979, p. 7].[3]

He criticizes the syntactic approach and rejects it on semantic grounds, and gives two counterexamples. He argues that sentences such as "Everything seeing every man is an animal" or "You are not seeing every man" are both significative and in them the syncategorematic term 'every' occurs as part of a distributed subject and as part of a distributed predicate, respectively. (The status of 'every' as a syncategorematic term is not in question.)

One consequence of the syntactic characterization is that it "produces mutually exclusive and jointly exhaustive classes that coincide almost perfectly with certain groupings of the parts of speech recognized by medieval grammarians: the categoremata are the names (both substantival and adjectival), the personal and demonstrative pronouns, and the verbs (excluding auxiliary verbs); the syncategoremata are all the others–e.g., the conjunctions, adverbs, and prepositions" [Kretzmann, 1982, pp. 211–212]. This quickly leads to an objection to defining the distinction in this fashion: Any mutual exclusive and jointly exhaustive division will be too neat. We have already seen an example of a term that is considered both categorematic and syncategorematic, so that categories cannot be exclusive. That they cannot be exhaustive is also argued by Paul of Venice, who offers three examples of terms which are neither categorematic nor syncategorematic: terms such as *nihil* 'nothing', the copula, and material terms 'A', 'B', 'C', which stand for terms but are not significative in themselves [of Venice, 1979, pp. 6–7].

The semantic tradition focuses on how words signify. Such an account is espoused by Henry of Ghent (c. 1260) who says:

> They are called syncategorematic as if to say 'consignificant'—i.e., significant together with others, namely, with categoremata—not because they signify *nothing* on their own, but because they have a signification that is not definite but indefinite, a signification whose definiteness they derive from those [words] that are adjoined to them [Kretzmann, 1982, p. 213].[4]

[3] *Terminus syncategorematicus dicitur qui, significative acceptus, non potest esse subiectum aut praedicatum, aut pars subiecti aut pars praedicati distributi, propositionis categoricae* [of Venice, 1979, p. 6].

[4] *Dicuntur sincathegoreumatice, quasi: consignificative, idest:* cum aliis significative *scilicet cum cathegoreumaticae; non quia de se nichil significant, sed quia habent significationem non finitam sed infinitam, cuius finitationem trahunt ab adjunctis* [Kretzmann, 1982, p. 214, fn. 12]. For the dating of Henry's treatise, see [Braakhuis, 1981, p. 135].

Thus, it is not so much that syncategorematic terms do not signify, but rather that they signify only in an indeterminate way, and that they modify the signification of the other terms in the proposition [Biard, 2011, §1]. This is similar to Paul of Pergola's definition. In his *Tractatus de sensu compositu et diviso*, written in the first half of the 15th century, Paul divides non-complex terms into those which signify *per se* and those which don't. A categorematic term is one which signifies *per se* and "which both by itself and when posited with something else has a proper or determine signification" (distinct) or "which both by itself and when posited with something else has a signification but [it is] indeterminate" (indistinct).[5] A syncategorematic term "is that which neither by itself nor when posited with something else signifies something, but has only a function (*officium*), such as: 'every', 'none'".[6]

The third category of approaches combine syntactic and semantic aspects, usually (but not always) appealing to suppositional properties of the terms. For example, Peter of Spain in his *Syncategoreumata* says:

> Syncategorematic words signify something or other. Now they do not signify things that are capable of functioning as subjects or as predicates. Therefore they signify things that are dispositions of things capable of functioning as subjects or predicates [of Spain, 1992, p. 39].[7]

This definition focuses on the grammatical possibility of a term to be a subject or predicate of a sentence, but it does so not on the basis of their classification of words as noun, verbs, or other, but rather on the basis of their signification. Albert of Saxony (c. 1356) takes up this point in the first argument against his fifth disputed question on logic, "Does every noun signify something other than itself, or what is similar to it in the mind, in speech, or in writing?":

> It is argued first that they do not, because syncategorematic terms, like 'all', 'no', etc. are nouns and, nevertheless, they do not signify anything other than themselves, or things similar to themselves in the mind, in speech, or in writing [of Saxony, 2010, p. 64, ¶84].

If categorematic terms are simply defined as "nouns and verbs" without any other attention to their semantic aspects, then *omnis*, *nulla*, and the like would be categorematic, not

[5] *Qui tam per se quam cum alio positus habet significatum proprium seu determinatum... qui tam per se quam cum alio positus habet significatum sed indeterminatum* [of Pergola, 1961, p. 7]. Translations not otherwise footnoted are my own.

[6] *Qui nec per se, nec cum alio positus aliquid significat* [of Pergola, 1961, p. 7].

[7] *Dictiones sincategorematice significant res aliquas. Sed non significant res subicibiles vel predicabiles. Ergo significant res que sunt dispositiones subicibilium vel predicabilium* [of Spain, 1992, p. 38].

syncategorematic. Other 'mixed' approaches focus on the ways in which the term supposit. For example, in his treatise on the divided and composite senses, William Heytesbury distinguishes between sentences in which *infinita* supposits categorematically and those where it supposits syncategorematically.[8]

The most articulate mixed approach that we consider is that of Paul of Venice. After rejecting a syntactic characterization of the distinction, Paul offers the following alternative definition of categorematic and syncategorematic terms:

Definition 3.4.1. A categorematic term is a sign, inwardly as well as outwardly simple, in accordance with a common law, without any sort of unifying effect on subject and predicate, but leading *per se* to a conception of something other than itself and what is equiform to it [of Venice, 1979, p. 3].[9]

Definition 3.4.2. A syncategorematic term is a sign that carries out a function and in the absence of a new imposition is significant *per se* of nothing other than itself and what is equiform to it [of Venice, 1979, p. 5].[10]

Paul notes that in both definitions, 'a sign' indicates the genus of 'term'; that is, both categorematic and syncategorematic terms must be significative. Thus, he also implicitly rejects a definition of the distinction which makes syncategorematic terms have signification only in a derivative sense. His further explication of the types of functions that a sign can carry out—which is a semantic rather than syntactic matter—is what causes him to revise the standard rules for conversion, which are generally stated in syntactic terms; we discuss this below.

Note 3.4.3. While we often speak of 'categorematic' and 'syncategorematic' *terms*, we should in fact speak of 'categorematic' and 'syncateorematic' *uses* of terms.

We should not be talking of 'syncategorematic terms' or 'categorematic terms'. If we do, then it doesn't make sense to speak of one and the same term being categorematic or syncategorematic. Instead, if we wish to talk about categorematic *infinita* and syncategorematic *infinita*, then we must also talk about equivocation: For it cannot be that one and

[8] "The consequence is invalid because in the one proposition the term 'infinite' supposits categorematically, in the other syncategorematically" [Heytesbury, 1988, p. 422], *Et sic de talibus non valet consequentia: Quia in una propositione iste terminus infinitum supponit cathegorematice, in alia sincathegorematice* [Heytesbury, 1494, fol. 3v].

[9] *Terminus categorematicus est signum, tam implicite quam explicite simplex, de communi lege, non extremorum aliqualiter unitivum, sed alterius a se et suo consimili per se in notitiam deductivum* [of Venice, 1979, p. 2].

[10] *Terminus syncategorematicus est signum officii executivum, nullius a se et suo consimili sine nova impositione per se significativum* [of Venice, 1979, p. 4].

the same word, *infinita*, is both a categoreme and a syncategoreme, but rather there are in fact two homophonic words. But it seems unnecessarily complex to postulate two distinct words, categorematic *infinita* and syncategorematic *infinita*. What is more natural—and indeed, what is more commonly done by the medieval authors—is to say that a term or a word can be *used* either categorematically or syncategorematically, and such usage depends on numerous factors, both syntactic and semantic.[11]

Definition 3.4.4 (Grammarians' definition)**.** A term is a *categorema* if it can be the subject or the predicate of a proposition. Other meaningful terms are *syncategoremata*.

Definition 3.4.5 (Logicians' definition)**.** An incomplete list of about fifty words that are discussed as syncategorematic, usually prepositions, conjunctions, and quantifiers. Among them are words like *omnis*. Important syncategoremata: *et*, *ut*, *cum*, *vel*, *omnis*, *uterque*...

William of Sherwood's list of syncategorematic terms includes:

- all, every;
- whole;
- number words;
- infinitely many;
- both;
- of every sort;
- no;
- nothing;
- neither;
- but;
- alone;
- only;

[11] This is not a new suggestion; it occurs in, e.g., [Murdoch and Thijssen, 2001, p. 129] and [Braakhuis, 1981, p. 143], albeit not explicitly in the latter. However, we feel quite strongly that this suggestion has not been taken up as widely as it should be, and thus wish to stress this point, which comes up again in the next section.

- is;
- not;
- necessarily, contingently;
- begins, ceases;
- if;
- unless;
- but that;
- and;
- or;
- whether;
- '*ne*';
- whether... or.

3.4.2 Supposition

Suppositio is the analysis of the meaning of terms in propositions. In the view of many modern philosophers, (e.g., Moody in [Moody, 1953]), *suppositio* can be seen as a theory of reference.

Situation 1.[My notes: seems loosely and incorrectly defined.]

- Under what conditions is *omnis homo philosophus est* true?
- If *philosophus* supposits ('stands for') for every instance of *homo* (*suppositio mobilis*). It is called 'mobile' supposition because you can instantiate. From *Aristoteles homo est* you can conclude *Aristoteles philosophus est.*

Situation 2.

- Under what conditions is *omnis homo praeter Socratem philosophus est* true?
- If *philosophus* supposits for every instance of *homo* except for Socrates.

At this point we need to make two remarks about how Latin differs from English. First, Latin doesn't have an indefinite article. A sentence like *Homo est philosophus* is ambiguous because it could mean 'man is a philosopher', 'a man is a philosopher', or 'some man is a philosopher' (though the latter is usually written *Aliquis homo est philosophus*). Many medieval puzzles exploit this inherent ambiguity.

Second, the medievals didn't use quotation marks:

- *Homo est disyllabum.*
- 'Human' is disyllabic.

Situation 3.

- Under what conditions is *homo est disyllabum* true?
- If *disyllabum* supposits for every instance of *homo*. (But here, *homo* is a singular term standing for '*homo*').
- *Flawed instantiation: Aristoteles homo est. Aristoteles disyllabum est.*

A consequence for logic: Whether conversion rules can be applied depends on the type of supposition in the proposition.

homo est disyllabum.	
aliquis homo est disyllabum.	
aliquis disyllabum est homo.	(simple conversion)
disyllabum est homo.	

Disyllabic is a man. Is this a medieval example?

The final part of supposition theory addresses the question of how changes in the tense of the main verb affects the supposition of the terms in the sentence.

- What makes *Aristoteles academicus erat* true?
- *Attempt 1.* If *academicus* supposits for *Aristoteles*.
  But if *academicus* supposits for *Aristoteles*, then *Aristoteles academicus est* is true.
- *Attempt 2* (modern reading). If there was a point in the past when *academicus* supposited for *Aristoteles*.

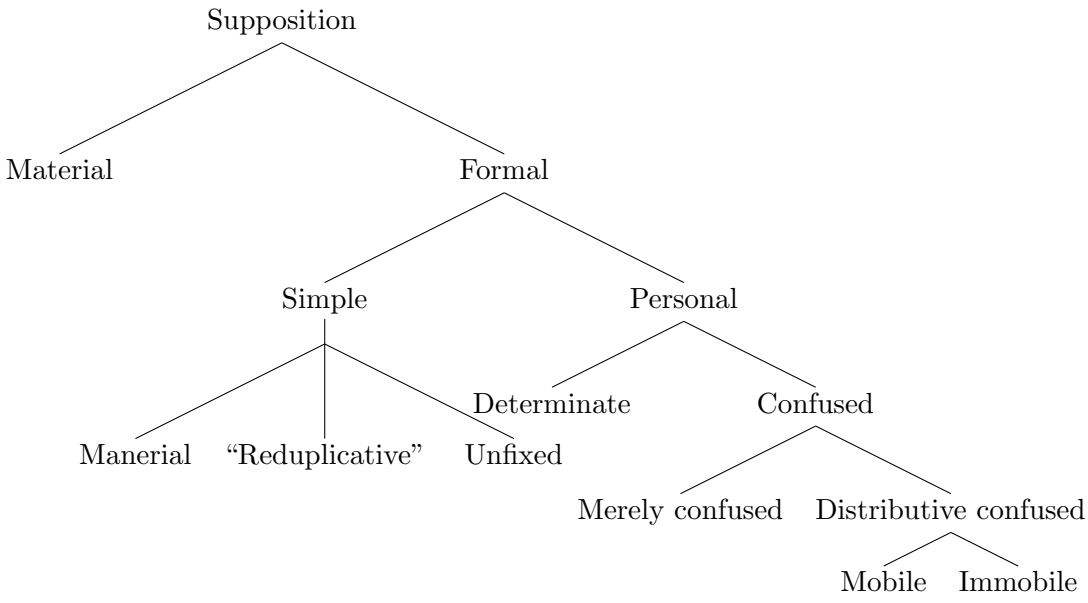


Figure 3.1: Sherwood's division of types of *suppositio*.

- Medieval theory: ampliation and restriction: *si terminus communis verbo de praeterito supponeret, posset supponere pro non-enti, ut hoc* homo cucurrit *verum est pro Caesare* (William of Sherwood, *Introductiones*).
- In general: the predicate determines the type of *suppositio* and whether *ampliatio* has to be used in order to determine the truth conditions.

Still: Even ampliation only allows to move back and forth in time, or along other modal accessibility relations (conceivability, possibility etc.). *Omnis chimaera est chimaera* is false regardless of ampliation since there was never and will never be an instantiation of "*chimaera*" that *chimaera* could supposit for.

St. Vincent Ferrer (1350–1419), *De suppositionibus dialecticis* (1372). "*rosa est odorifera*" is true even if there are no roses and never have been roses.

This type of supposition mongering helped give Scholasticism its bad name in later times.

3.5 *Obligationes*

An obligational disputation, or *obligatio*, is a dialogue between two agents, the Opponent and Res, where the Opponent puts forward a sequence of propositions, and Res is obligated

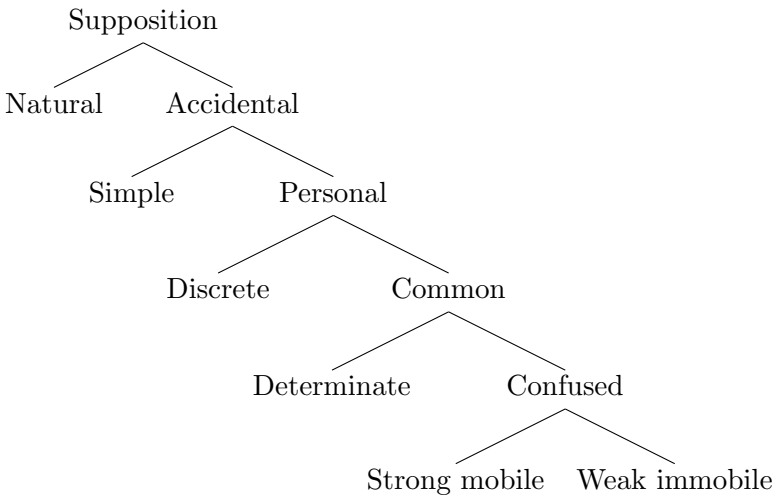


Figure 3.2: Lambert's division of types of *suppositio*.

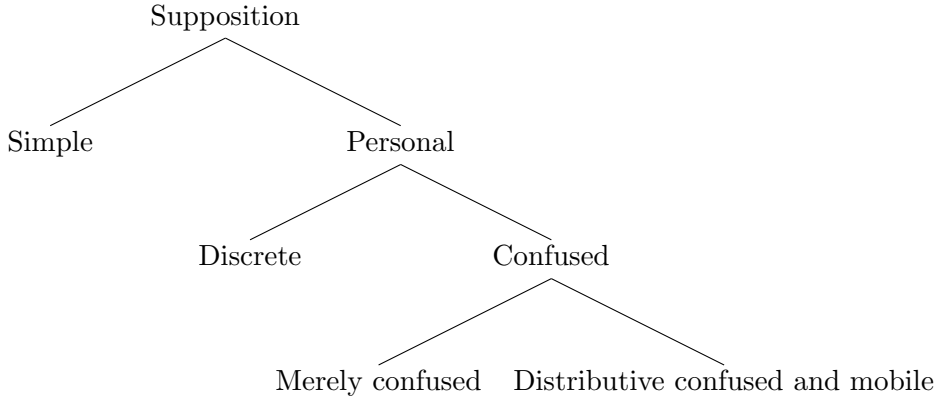


Figure 3.3: Bacon's division of types of *suppositio*.

(hence the name) to follow certain rules in his responses to the Opponent's propositions. More precisely, the Opponent puts forward an initial statement, called the *positum*, which Res can either accept or refuse to accept. If he accepts, the *obligatio* begins. If he does not, no *obligatio* begins. If the *obligatio* begins, the Opponent puts forward propositions and Res has three ways that he can respond: He can grant or concede (*concedere*) the proposition, he can deny (*negare*) the proposition, or he can doubt (*dubitare*) it, that is, remain agnostic. (Some authors, such as William of Ockham [**?**] and the anonymous author of the *Obligationes Parisienses* [de Rijk, 1975], mention a fourth option, which is to 'draw distinctions' (*distinguere*), that is, to clarify an ambiguity on the part of the Opponent, but

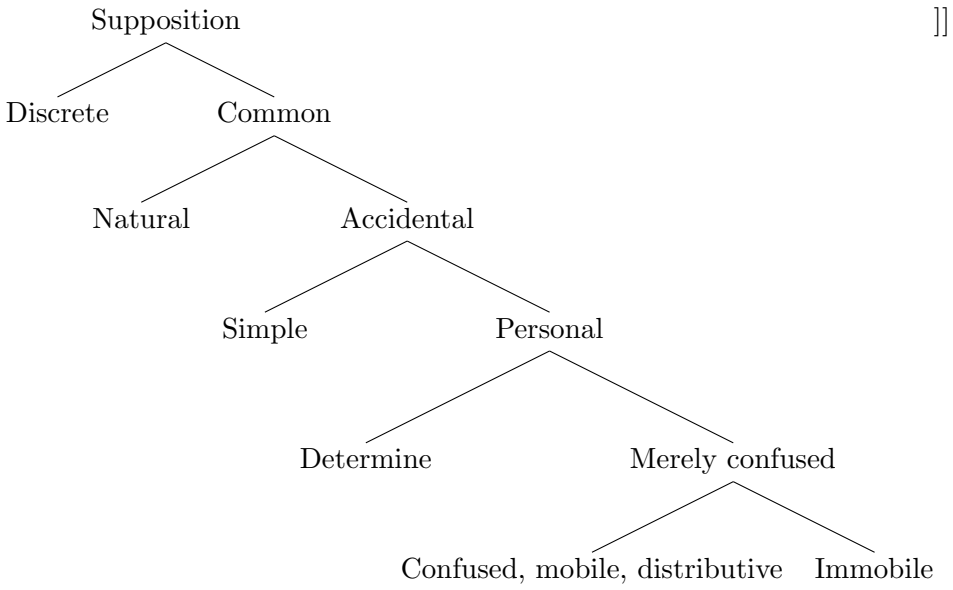


Figure 3.4: Peter's division of types of *suppositio*.

for ease of modeling we will ignore this action in the present paper.) The *obligatio* continues until the Opponent calls "*Cedat tempus*" ("Time's up"), whereupon the responses of Res are analysed with respect to the rules Res was supposed to follow, to determine whether Res has responded well or badly.

The earliest texts on *obligationes* date from the early 13th century [de Rijk, 1974, de Rijk, 1975, de Rijk, 1976], and while their roots are clearly grounded in Aristotle's discussion of dialectical exchanges in the *Topics* VIII, 4 (159a15–24) and in the *Prior Analytics* I, 13 (32a18–20) (cf. [Yrjönsuuri, 1994, §II.A]), the systematic development of the theory of *obligationes* over the course of the 13th and 14th centuries tends to show little adherence to the Aristotelian definitions. While the specific details vary from author to author, a number of distinct types of *obligationes* discussed by multiple authors can be identified. The six most common are: *positio*, *depositio*, *dubitatio*, *sit verum* or *rei veritatis*, *institutio*, and *petitio*. Of these six, *positio* is universally the most widely studied, both by medieval and modern authors.

A number of people who wrote on *obligationes*:

- Nicholas of Paris
- Walter Burley (Burleigh; c. 1275–1344)
- Roger Swyneshed (d. 1365)

responsio antiqua	*responsio nova*
Walter Burley	Roger Swyneshed
William of Sherwood	Robert Fland
Ralph Strode	Richard Lavenham
Peter of Candia	
Paul of Venice	

Figure 3.5: *Responsio antiqua* and *responsio nova*

- Richard Kilvington (d. 1361)
- William Ockham (c. 1285–1347)
- Robert Fland (c. 1350)
- Richard Lavenham (d. 1399)
- Ralph Strode (d. 1387)
- Peter of Candia
- Paul of Pergola
- Peter of Mantua
- Paul of Venice (c. 1369–1429)

3.5.1 *positio*

We discuss two theories typical of the *responsio antiqua* and the *responsio nova* (see Figure 3.5). The first is Walter Burley, in whose *De obligationibus* is a presentation of what can be called the "standard" set of rules. The other is Roger Swyneshed, in whose *Obligationes* (1330–1335) there is a radical change in one of the rules results in a distinctly different system.

Walter Burley's treatise *De obligationibus*, written around 1302, when he was a master of arts at the University of Oxford, gives a standard treatment of *positio*. The text of this treatise is edited in [Burley, 1963] and a partial translation of the text, including the section on *positio* in its entirety, is found in [Burley, 1988a]. Burley defines the general goal of an *obligatio* as follows:

> The opponent's job is to use language in a way that makes the respondent grant impossible things that he need not grant because of the *positum*. The respondent's job, on the other hand, is to maintain the *positum* in such a way that any impossibility seems to follow not because of him but rather because of the *positum* [Burley, 1988a, p. 370].[12]

Summarizing *positio* according to Burley:

- Two players, Opp and Res.
- Opp starts by positing a *positum* φ^*.
- Res can admit or deny the *positum*. If he denies it, the game does not begin.
- If he admits the *positum*, the game starts. We set $\Phi_0 := \{\varphi^*\}$.
- In each round n, Opp proposes a statement φ_n and Res either "concedes", "denies" or "doubts" this statement according to certain rules. If Res concedes, then $\Phi_{n+1} := \Phi_n \cup \{\varphi_n\}$, if he denies, then $\Phi_{n+1} := \Phi_n \cup \{\neg\varphi_n\}$, and if he doubts, then $\Phi_{n+1} := \Phi_n$.
- We call φ_n pertinent (relevant) if either $\Phi_n \vdash \varphi_n$ or $\Phi_n \vdash \neg\varphi_n$. In the first case, Res has to concede φ_n, in the second case, he has to deny φ_n.
- Otherwise, we call φ_n impertinent (irrelevant). In that case, Res has to concede it if he knows it is true, to deny it if he knows it is false, and to doubt it if he doesn't know. Real world knowledge is relevant here.
- Opp can end the game by saying *Tempus cedat.*

Example 3.5.1.

Opponent	Res	
I posit that Cicero was the teacher of Alexander the Great: φ^*.	I admit it.	$\Phi_0 = \{\varphi^*\}$.
Cicero was Roman: φ_0.	I concede it.	Impertinent and true; $\Phi_1 = \{\varphi^*, \varphi_0\}$.
The teacher of Alexander the Great was Roman: φ_1.	I concede it.	Pertinent, follows from Φ_1.

[12] *Opus opponentis est sic inducere orationem ut faciat respondentem concedere impossibilia quae propter positum non sunt necessaria concedere. Opus autem respondentis est sic sustinere positum ut propter ipsum non videatur aliquod impossibile sequi, sed magis propter positum* [Burley, 1963, p. 34].

Example 3.5.2.

Opponent	Res	
I posit that Cicero was the teacher of Alexander the Great: φ^*.	I admit it.	$\Phi_0 = \{\varphi^*\}$.
The teacher of Alexander the Great was Greek: φ_0.	I concede it.	Impertinent and true; $\Phi_1 = \{\varphi^*, \varphi_0\}$.
Cicero was Greek: φ_1.	I concede it.	Pertinent, follows from Φ_1.

Example 3.5.3. ("order matters!")

Opponent	Res	
I posit that Cicero was the teacher of Alexander the Great: φ^*.	I admit it.	$\Phi_0 = \{\varphi^*\}$.
The teacher of Alexander the Great was Roman: φ_0.	I deny it.	Impertinent and false; $\Phi_1 = \{\varphi^*, \neg\varphi_0\}$.
Cicero was Roman: φ_1.	I deny it.	Pertinent, contradicts Φ_1.

Here are some properties of Burley's *positio*.

- Provided that the *positum* is consistent, no disputation requires Res to concede φ at step n and $\neg\varphi$ at step m.
- Provided that the *positum* is consistent, Φ_i will always be a consistent set.
- It can be that Res has to give different answers to the same question (Example 3.5.4).
- Opp can force Res to concede everything consistent (Example 3.5.5).

Example 3.5.4. Suppose that Res is a student, and does not know whether the King of France is currently running.

Opponent	Res	
I posit that you are the Pope or the King of France is currently running: φ^*	I admit it.	$\Phi_0 = \{\varphi^*\}$.
The King of France is currently running: φ_0	I doubt it.	Impertinent and unknown; $\Phi_1 = \{\varphi^*\}$.
You are the Pope: φ_1.	I deny it.	Impertinent and false; $\Phi_2 = \{\varphi^*, \neg\varphi_1\}$.
The King of France is currently running: $\varphi_2 = \varphi_0$.	I concede it.	Pertinent, follows from Φ_2.

Example 3.5.5. Suppose that φ does not imply $\neg\psi$ and that φ is known to be factually false.

Opponent	Res	
I posit φ.	I admit it.	$\Phi_0 = \{\varphi\}$.
$\neg\varphi \vee \psi$.	I concede it.	Either φ implies ψ, then the sentence is pertinent and follows from Φ_0; or it doesn't, then it's impertinent and true (since φ is false); $\Phi_1 = \{\varphi, \neg\varphi \vee \psi\}$.
ψ	I concede it.	Pertinent, follows from Φ_1.

positio according to Swyneshed.

- All of the rules of the game stay as in Burley's system, except for the definition of *pertinence*.
- In Swyneshed's system, a proposition φ_n is pertinent if it either follows from φ^* (then Res has to concede) or its negation follows from φ^* (then Res has to deny). Otherwise it is impertinent.

Here are some properties of Swyneshed's *positio*.

- Provided that the *positum* is consistent, no disputation requires Res to concede φ at step n and $\neg\varphi$ at step m.

- Res never has to give different answers to the same question.
- Φ_i can be an inconsistent set (Example 3.5.6).

Example 3.5.6. Suppose that the respondent is a student in Paris, and not a bishop. Write ψ_0 for "You are in Rome" and ψ_1 for "You are a bishop".

Opponent	Res	
I posit that you are in Rome or you are a bishop: $\psi_0 \vee \psi_1$	I admit it.	$\Phi_0 = \{\psi_0 \vee \psi_1\}$.
You are in Rome or you are a bishop: $\psi_0 \vee \psi_1$.	I concede it.	Pertinent, follows from Φ_0; $\Phi_1 = \{\psi_0 \vee \psi_1\}$.
You are not in Rome: $\neg\psi_0$.	I concede it.	Impertinent, and true; $\Phi_2 = \{\psi_0 \vee \psi_1, \neg\psi_0\}$.
You are not a bishop: $\neg\psi_1$.	I concede it.	Impertinent, and true; $\Phi_3 = \{\psi_0 \vee \psi_1, \neg\psi_0, \neg\psi_1\}$.

Φ_2 is an inconsistent set of sentences.

positio according to Kilvington. Richard Kilvington (d. 1361).

- *Sophismata*, c. 1325.
- *obligationes* as a solution method for sophismata.
- He follows Burley's rules, but changes the handling of impertinent sentences.
- If φ_n is impertinent, then Res has to concede if it were true if the *positum* was the case, and has to deny if it were true if the *positum* was not the case.

3.5.2 *depositio*

Depositio is just like *positio*, except that Res is obliged to deny or reject the initial proposition (the *depositum*). A *depositio* with *depositum* φ will be completely symmetric to a *positio* with $\neg\varphi$ as the *positum*. Nevertheless, early treatises on *obligationes*, such as that by Nicholas of Paris which dates from ca. 1230–50, still treat *depositio* at some length. "In the section on *depositio* the question is raised as to whether, when a contingent true proposition is deposited, any (contingent) true proposition that is compossible with the *depositum* can be proved, in the same way that any (contingent) false proposition that is compossible with the *positum* can be proved or conceded when a contingent false proposition is posited"

[Braakhuis, 1998, pp. 156–57].[13] The use here of 'prove' is infelicitous; in a case like that which Braakhuis has just described, nothing at the propositional level is being *proved*. It is only meta-level properties, such as consistency and compossibility[14], that are proved or disproved in an obligational disputation.

3.5.3 *dubitatio*

In *dubitatio*, Res must *doubt* the statement that the Opponent puts forward (called the *dubitatum*). While *dubitatio* was discussed in 13th century texts, often at some length, later authors (both later medieval and modern authors) call *dubitatio* a trivial variant of *positio*, and thus spend little time discussing it. For example, Paul of Venice [of Venice and Ashworth (ed. and trans.), 1 reduces *dubitatio* to *positio* (in much the same way that he, and others, reduces *depositio* to *dubitatio*); Swyneshed, Lavenham, John of Wesel, Richard Brinkley [Brinkley et al., 1995], and John of Holland [of Holland and Bos, ed., 1985] do not mention *dubitatio* at all. However, as we have argued elsewhere [Uckelman et al., 2018, Uckelman, 2011a], such a trivializing view of *dubitatio* fails to recognize the higher-order aspects of the disputation, the mixing of both knowledge and truth, which result in a significantly more difficult type of game, which, contrary to Stump [Stump, 1985, p. 372, fn. 14], does not involve any type of three-valued logic; while there is a tripartite structure of the dispositions of Res (i.e., that certain propositions must be doubted, certain must be denied, and certain others must be conceded), these dispositions should not be taken as assigning (epistemic) truth values to the propositions. Just as *positio* is only interesting when the *positum* is false, *dubitatio* is only interesting when the truth value of the *dubitatum* is known (whether it is true or false). Thus, part of the complexity of the game arises from the interaction between knowledge, truth, and the obligations of Res, as Res in many cases is required to respond *dubio* 'I doubt it' to propositions that he actually knows.

One of the earliest texts that discusses *dubitatio* at any length is the anonymous *Obligationes Parisienses* edited by de Rijk in [de Rijk, 1975]. This text, tentatively datable to the first or second decade of the 13th C [Uckelman et al., 2018, §2.1], is also one of the few that discusses the status of *dubitatio* as a legitimate distinct species of *obligationes*. The author gives two general rules, for propositions which are not equivalent with the *dubitatum*, and eight specific rules and facts which give more explicit rules governing Res's actions. Because we have analysed these rules in detail elsewhere (along with the rules in William of Sherwood's treatise [of Sherwood, 1963]) [Uckelman et al., 2018], we do not discuss them

[13] "*Datur pro regula quod falso possibili posito de quolibet falso compossibili illi debet concedi ipsum esse*" [Braakhuis, 1998, p. 202].

[14] Two sentences are compossible if they are both possibly true together at the same time.

here, but instead look at one of the most extensive and comprehensive 13th-century discussions of *dubitatio*, found in Nicholas of Paris's treatise. Nicholas gives seven rules for *dubitatio* [Braakhuis, 1998, pp. 72–76]. Many of these rules are similar to those for *positio* and *depositio*:

Rule 3.5.7. Just as in false *positio* it is impossible to put forward "a falsehood is put forward" nor in *depositio* "a falsehood is to be deposed", by the same reason it is impossible to doubt "a falsehood is doubted".[15]

Rule 3.5.8. Just as in *positio* a *positum* put forward in the form of the *positum*, and everything convertible to it in the time of positing is to be conceded and its opposite and things convertible with it is to be denied and just as in *depositio* a *depositum* put forward in the form of the *depositum*, with its convertibles, must be denied and its opposite with things convertible with it must be conceded; so in *dubitatio* for a *dubitatum* put forward in the form of *dubitatum* and for its convertibles and moreover for the opposite of the *dubitatum* with its convertibles must be answered "prove!"[16]

Rule 3.5.9. For everything irrelevant to the *dubitatum* the response must be according to its quality.[17,18]

Rule 3.5.10. All the responses must be directed to the same instant.[19]

However, the rules that govern relevant propositions have interesting consequences. He says:

Rule 3.5.11. For everything antecedent to the *dubitatum* the response must be "false" or "prove!" and never "true".[20]

[15] "*Sicut in falsa positione non potest poni* falsum poni *nec in depositione* falsum deponi, *ita nec in dubitatione potest dubitari falsum dubitari*" [Braakhuis, 1998, p. 223].

[16] "*Sicut in positione positum sub forma positi propositum et omne convertibile cum illo in tempore positionis est concedendum et suum oppositum cum suo convertibili negandum, et sicut in depositione depositum sub forma depositi propositum cum suo convertibili negandum et suum oppositum cum suo convertibili concedendum, ita in dubitatione ad dubitatum sub forma dubitati propositum et ad suum convertibile et preterea ad oppositum dubitati cum suo convertibili respondendum est 'proba'*" [Braakhuis, 1998, p. 223].

[17] "*Ad omne impertinens dubitato respondendum est secundum sui qualitatem*" [Braakhuis, 1998, p. 225].

[18] Note that while Nicholas's definition of relevance [Braakhuis, 1998, p. 179] taken strictly mirrors Swyneshed's, it is clear from the succeeding rules that his notion of relevance is the same as Burley's and other 13th C authors.

[19] "*Omnes responsiones retorquende sunt ad idem instans*" [Braakhuis, 1998, p. 227].

[20] "*Ad omne antecedens ad dubitatum respondendum est 'falsum' vel 'proba' et nunquam 'verum'*" [Braakhuis, 1998, p. 224].

	Opponent	Res
1	'Socrates is white'	I doubt it.
2	'Socrates is white'	I doubt it.
3	'Socrates is pale/fair'	I doubt it.
4	'Socrates is black'	I concede it.
5	'It is false that Socrates is white'	I concede it.
6	*Cedat tempus*	

Figure 3.6: *Dubitatio* according to Nicholas of Paris.

Rule 3.5.12. For everything consequent to the *dubitatum* it is possible to reply "it is true" or "prove" and never "it is false".[21]

Here we see another cause of the complexity of *dubitatio*, and how it cannot simply be reduced to *positio* or *depositio*: This is because the rules, unlike the rules for the other types, are not deterministic. Whereas there is always a unique correct response for Res in *positio* (in both the *responsio antiqua* and *nova*), here, the rules give Res a range of choices. This non-determinacy means that there is a plurality of ways that Res may act, and still be disputing according to the rules, a feature which no other version of *obligatio* has. However, this feature of *dubitatio* seems not to have been noticed by later authors who insisted that *dubitatio* could be reduced to *positio*.

The second to the last rule he gives is the most curious:

Rule 3.5.13. The questioning exercise cannot be bounded/terminated (*terminari*).[22]

However, he gives no explanation of this rule, or why it is introduced, and while he says that this rule is "just as in the preceding [types of] *obligationes*"[23], no similar rule is mentioned in his discussions of other types of *obligatio*.

We give an example of *dubitatio* in Figure 3.6, adapted from [Braakhuis, 1998, pp. 223–224]. Res has responded correctly in rounds (2) and (3), since (2) is identical with the *dubitatum* and (3) is equivalent to the *dubitatum*, and he also responds correctly in round (5) since black and white are exclusive, so Socrates's being white and his being black are repugnant (inconsistent) but he responds incorrectly in round (4) when he concedes that Socrates is black; for from the fact that it is doubtful whether Socrates is white it does not

[21] "*Ad omne consequens ad dubitatum potest responderi 'verum est' vel 'proba' et nunquam 'falsum est'*" [Braakhuis, 1998, p. 224]

[22] "*Non possit terminari disciplinalis questio*" [Braakhuis, 1998, p. 226].

[23] "*Sicut in precedentibus obligationibus...*" [Braakhuis, 1998, p. 226].

follow that Socrates is black, and further, by conceding that Socrates is black, Res is later forced to deny the *dubitatum*, thus violating his primary obligation.

Nicholas's *dubitatio* has similar formal properties to *positio*. Provided that the *dubitatum* is neither a contradiction nor a tautology, it can be proved that Res can *win* the disputation playing by Nicholas's rules for *dubitatio*: that is, there is never any case where he will be forced either to concede or to deny the *dubitatum* [Uckelman, 2011a, Theorem 24].

3.5.4 *impositio / institutio / appellatio*

The obligation involved in *impositio*, also called *institutio* or *appellatio*, functions in a relevantly different manner from the obligation in *positio*, *depositio*, or *dubitatio*. Whereas in these latter three, Res's obligation involves how he is to respond to the *obligatum*, *impositio* involves an obligation to redefine certain terms or phrases. The anonymous author of *Obligationes Parisienses* notes that "*Institutio* is divided into certain *institutio* and uncertain or obscure *institutio*, for example if the name 'Marcus' is fixed that it might be a name of Socrates or Plato, but you would not know of which".[24] Lavenham defines *impositio* as an "obligation by means of which a term or proposition is assigned a [new] signification"[25], and gives the following examples:

> An example of the second: I impose that '*a*' signifies precisely that God exists. The whole "'*a*' signifies precisely that God exists" is what is obligated.[26]

and

> . . . I impose that this term 'man' may be converted with this word 'donkey', or I impose that this proposition 'God exists' signifies precisely that man is donkey.[27]

In the first example, '*a*' is being instituted as the name of a proposition that signifies that God exists; likewise in the second example of the second quote, the phrase 'God exists' is instituted as the name of a proposition signifying that man is donkey; thus any time that 'God exists' is asserted in a disputation, it must be understood as meaning 'Man is donkey'. In the first example of the second quote, the institution is not at the level of propositions

[24] "*Institutio dividitur in institutionem certam et incertam sive occultam, ut ⟨si⟩ instituatur hoc nomen 'Marcus' quod sit nomen Sortis vel Platonis sed nescias utrius*" [de Rijk, 1975, p. 28].

[25] "*. . . obligatio mediante qua terminus vel propositio imponitur ad significandum*" [Spade, 1978, ¶21]

[26] "*Exemplum secundi: Impono quod 'a' praecise significet quod deus est. Hoc totum "'a' praecise significet quod deus est" est obligatum*" [Spade, 1978, ¶2].

[27] "*. . . impono quod iste terminus 'homo' convertatur cum ista dictione 'asinus', vel impono quod haec propositio 'Deus est' significet praecise quod homo est asinus*" [Spade, 1978, ¶21].

but at the level of words; it changes the meaning of the term 'man' so that it no longer means 'man' but instead means 'donkey'.

Impositio can take place in conjunction with any of *positio*, *depositio*, and *dubitatio*; that is, once a new imposition is introduced, then Res may also further be obliged to concede, deny, or doubt the initial *obligatum* of the disputation. Lavenham gives two rules that govern *impositio* in the context of *positio* (the assumption being that adapting these rules to handle *impositio* in the context of *depositio* will be straightforward):

> And it should be known that in *positio* these rules must be observed: It is not because of the imposition of any proposition to what it has to signify that the response to them varied... The second rule: Everything following from the *obligatum* by the imposition must be conceded, and everything repugnant denied.[28]

Simple *impositiones* like the ones given above are relatively easy; the only skill they require beyond the skills needed for *positio* is the skill to remember the new imposition of the term or proposition. However, much more complicated examples can be provided, such as the following (also due to Lavenham):

> I impose that in every false proposition in which 'a' is put down that it signifies only 'man' and that in every true proposition in which 'a' is put down that it signifies only 'donkey', and that in every doubtful proposition in which 'a' is put down that it signifies alternately with this disjunction 'man or non man'.[29]

Here, a is being used as a name for a term. This imposition may seem innocuous, but consider what happens with the proposition "Man is a" is put forward. Suppose the proposition is true. Then, it means "Man is donkey", which is impossible; hence, the proposition is false. But if it is false, it means "Man is man", which is in fact true! Thus, if it cannot be true or false, then it must be doubtful. But if it is doubtful, then it means 'Man is man or not man', which is true! No matter which assumption we make about the truth of the proposition, we are lead into contradiction.

---

[28] "*Et sciendum est quod in positione sunt istae regulae observandae: Non propter impositionem alicujus propositionis ad significandum est responsio ad illam varianda... Secunda regula: Omne sequens ex obligato ex impositione est concedendum et omne repugnans negandum*" [Spade, 1978, ¶¶22–23].

[29] "*Impono quod in omni propositione falsa in qua ponitur 'a' quod significet hominem solum, et quod in omni propositione vera in qua ponitur 'a' quod significet asinum solum, et quod in omni propositione dubia in qua ponitur 'a' quod significet convertibiliter cum hoc disjuncto 'homo vel non homo'*" [Spade, 1978, ¶24].

3.5.5 *Petitio.*

In *petitio*, the Opponent asks (petitions) Res to respond in a certain way. *Petitio* is rarely treated at any length, because, as a number of authors (Nicholas of Paris, Marsilius of Inghen [of Inghen, 1489], Peter of Mantua [Strobino, 2009], Paul of Venice) argue, *petitio* can be reduced to *positio*. In the mid-13th century, Nicholas of Paris says, when discussing how the verb *ponatur* 'to be put forward' initiates the obligation of Res and the phrase *cedat tempus* 'the time is finished' disperses the obligation:

> Whence... the sense is, when the Opponent says '*ponatur*': I request that you are restricted to conceding the *positum* and everything convertible with it and following from it, etc.[30]

A century and a half later, Paul of Venice reduces *petitio* to *positio* in the following way:

> For when I say,
>
> > I ask you to reply to A,
>
> this *petitio* makes the same claim as this *positio*:
>
> > I posit that you reply to A.
>
> Thus, in short, one should reply to all these species just as one would to a *positio* [of Venice and Ashworth (ed. and trans.), 1988, p. 39].[31]

Thus, from the disputational point of view, there is little more than cosmetic differences between *positio* and *petitio*.

However, for Burley, it is not *positio* but *institutio* (*impositio*) that *petitio* is connected to—*petitio* is the *actus* to *institutio*'s *habitus*, both of these governing what is noncomplex. He argues that "*petitio* is distinct from other species [of obligation], because a *petitio* posits the performance of an act that is mentioned in the statable thing [at issue], but the other species do not require this" [Burley, 1988b, p. 373–74].[32] An additional interesting fact about *petitio*, for Burley, is that, unlike *positio*, *petitio* can only engender an obligation for Res with Res's consent:

---

[30] "*Unde... sensus est, cum dicit opponens 'ponatur': peto ut restringaris ad concedendum positum et omne convertibile cum illo et consequens ad ipsum, etc....*" [Braakhuis, 1998, p. 183].

[31] "*Cum enim dico 'Peto te respondere ad A', adserit haec petitio sicut ista positio: 'Pono quod tu respondeas ad A. Ita breviter quod ad omnes huiusmodi species respondendum est sicut ad positionem'*" [of Venice and Ashworth (ed. and trans.), 1988, p. 38].

[32] "*Dicendum quod petitio est distincta ab aliis speciebus, quia petitio ponit actum exerceri, qui ponitur in enuntiabili, sed hoc non petunt aliae species*"[Burley, 1963, p. 41].

> So far it seems that *petitio* is not an obligation, for to require is not to obligate, because, unless the respondent consents, he is neither more nor less obligated in virtue of the opponent's requiring. We have to say [in reply] that *petitio* is an obligation—not just of any *petitio*, however, but only one that occurs with [the respondent's] consent [Burley, 1988b, p. 374].[33]

Further, because Opponent's initialization of the game is phrased as a request in *petitio*, there is a division into relative and absolute *petitiones* which is not present in *positio*. An example of an absolute *petitio* is the following: "I require you to grant that a man is a donkey"; an example of the second is "I require you to grant the first thing to be proposed by me". There is no way that this meta-level distinction could be made in *positio*.

In the *impositio*, Opp doesn't posit a *positum* but instead gives a definition or redefinition.

Example 3.5.14. "In this *impositio*, *asinus* will signify *homo*".

Example 3.5.15. "In this *impositio*, *deus* will signify *homo* in sentences that have to be denied or doubted and *deus* in sentences that have to be conceded."

Suppose Opp proposes "*deus est mortalis*".

- If Res has to deny or doubt the sentence, then the sentence means *homo est mortalis*, but this is a true sentence, so it has to be conceded. Contradiction.
- If Res has to concede the sentence, then the sentence means *deus est mortalis*, but this is a false sentence, so it has to be denied. Contradiction.

An *impositio* often takes the form of an insoluble.

3.5.6 *sit verum*

In early discussions of the genus of *obligationes*, six species are identified: *positio*, *depositio*, *dubitatio*, *petitio*, *institutio*, and *sit verum* or *rei veritas*. This sixth type is rarely discussed by the medieval authors. When it is, it is often not even explicitly defined, or it is reduced to one of the more common variants.[34] As a result, it is difficult to give a precise explanation or characterization of this type.

[33] "*Adhuc videtur quod petitio non sit obligatio, nam petere non est obligare, quia, nisi respondens consentiat propter petere opponentis, neque magis neque minus est obligatus. Dicendum quod petitio est obligatio, sed non quaelibet, sed solum petitio quae est cum consensum.*" [Burley, 1963, pp. 41–42].

[34] For example, John of Holland reduces *sit verum* to a form of *positio* (cf. John of Holland, *Four Tracts on Logic*, ed. by E.P. Bos (Ingenium Publishers, 1985)). Paul of Venice argues that the distinction between simple and complex *obligationes* is inconsistent, and as a corollary, *sit verum* cannot be distinguished from

In the early 13th-century anonymous treatise *Obligationes Parisienses*, the difference between *sit verum* and *positio* was illustrated by means of an example showing how disputations proceeding from the same *positum* and background assumptions differ depending on which set of rules is being used. We begin by looking in detail at the rules for *positio* in this treatise along with a rather lengthy example of it, since understanding this example is necessary for understanding how *sit verum* differs from *positio*.

Rule 3.5.16. Every proposition posited under the same form of speech under which it was posited, everything following from the *positum* and a thing conceded or things conceded, and everything true and not repugnant to these must be conceded.[35]

Rule 3.5.17. The opposite of the *positum*, and every false thing not following from the *positum* and a thing conceded or things conceded and the opposite or opposites of things correctly denied or a thing correctly denied, and every true thing repugnant to these must be denied.[36]

Comparing the two rules, it is clear that they are not symmetric, and that Rule 3.5.16 is incomplete since it omits mention of propositions which have previously been denied.[37] That these propositions are to be taken into account in determining whether a proposition should be conceded is clear from the method of evaluation of new *proposita* that the author goes on to give. This method also makes clear a more perspicuous and less convoluted way of stating the rules.

> However so that it may be properly judged of any proposition, the following consideration is made beyond the rules noted above: In receiving something in *positio*, if something is put forward, it should be considered of the proposition whether it is true or false. If true, either something has been previously conceded or previously denied, or nothing. If nothing, the opposite of the true proposition should be taken and the *positum* put down in a conditional in the antecedent, and the opposite of the true proposition in the consequent of that conditional. And so

petitio (Paul of Venice, *Logica Magna Part II Fascicule 8*, ed. E.J. Ashworth (Oxford University Press), 44–48). Cf. also Burley, 'Obligations (Selections)', 370.

[35] *Omne positum propositum sub eadem forma vocis sub qua sit positum, omne sequens ex posito et concesso vel concessis et omne verum non repugnans hiis, est concedendum.* de Rijk, 'Some Thirteenth Century Tracts', 29. All translations from this treatise are my own.

[36] *Oppositum positi et omne falsum non sequens ex posito et concesso vel concessis et opposito vel oppositis bene negatorum vel bene negati et omne verum repugnans hiis est negandum.* ibid., 29.

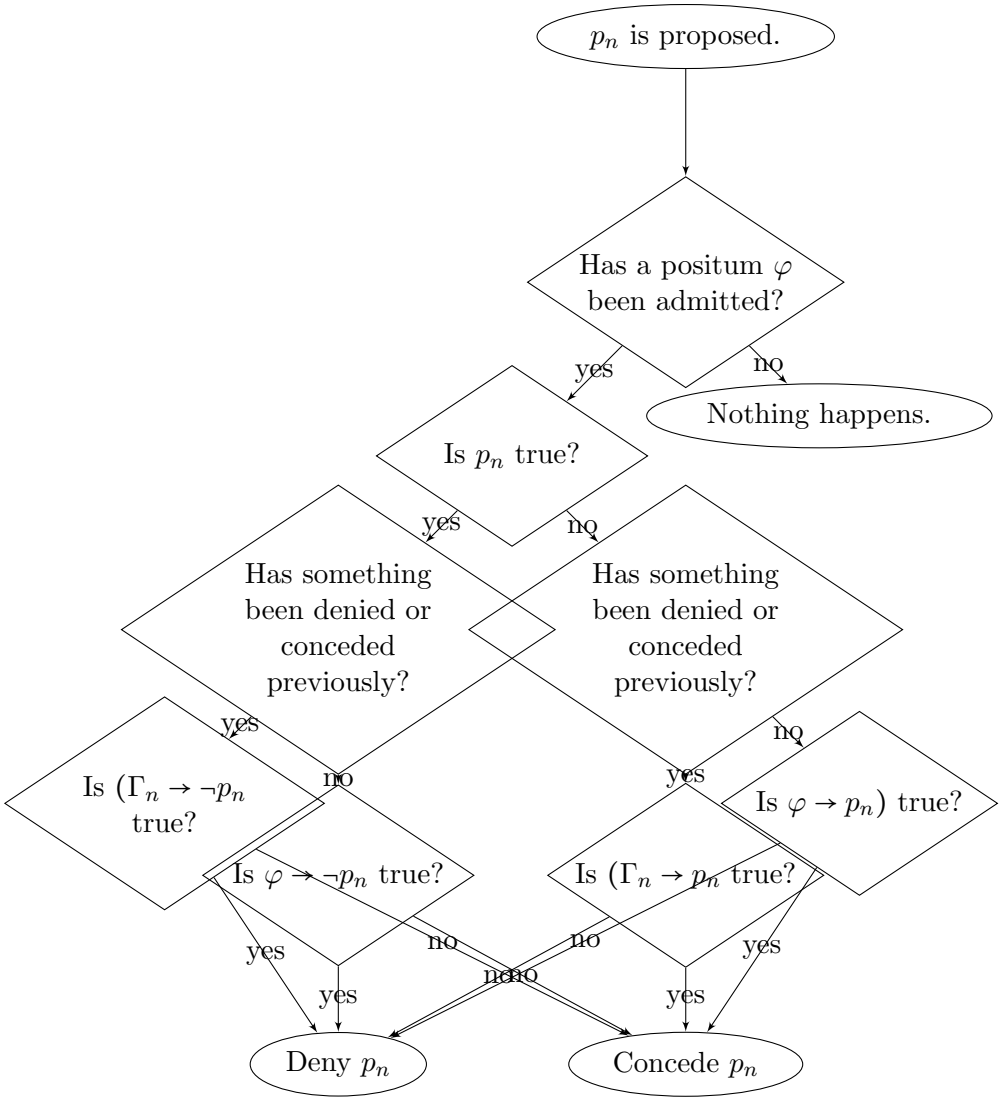
[37] This is not the only place in *Obligationes Parisienses* were rules are incompletely or incorrectly spelled out; see Sara L. Uckelman, Jaap Maat, and Katherina Rybalko, 'The Art of Doubting in *Obligationes Parisienses*, in *Modern Views of Medieval Logic*, eds. C. Kann, C. Rode, and S.L. Uckelman, (Peeters, forthcoming), §4.

arranged it is either true or false. If true, the true proposition is repugnant and must be denied; if false, it is not repugnant and must be conceded. If, however, something was previously conceded or previously denied, the *positum* should be taken with the conceded thing, or things if many have been conceded, and with the opposite of the denied, or opposites, and put down in the antecedent of a conditional, and the opposite of the true proposition in the consequent. That conditional will be either true or false. If true, the true proposition is repugnant and must be denied; if false, it is not repugnant and must be conceded. If however the proposition is false, either something was previously conceded or previously denied, or nothing. If nothing, the *positum* should be taken in the antecedent of a conditional and the false proposition in the consequent. That conditional will be either true or false. If true, the false proposition follows and must be conceded. If false, it does not follow and must be denied. If, however, something or things was previously conceded or previously denied, the *positum* should be taken with the concession, or concessions, and the opposite, or opposites, of the correct denial or denials, and put down in the antecedent of a conditional, and the false proposition in the consequent. That conditional will be either true or false. If true, the first proposition follows and must be conceded; if false, it does not follow, and must be denied.[38]

This algorithm is represented by the flowchart in Figure 3.5.6, and we can rewrite the rules as follows:

Rule 3.5.18. Every proposition identical to the *positum*, everything following from the

[38] *Ut autem bene iudicetur de quolibet proposito, super regulas predictas fit talis consideratio. Recepto aliquo in positionem si proponatur aliquid, consideretur de proposito an sit verum aut falsum. Si verum, aut aliquid sit preconcessum vel prenegatum aut nichil. Si nichil, sumatur oppositum veri propositi et propositum ponatur in conditionali in antecedente et oppositum veri propositi in consequente illius conditionalis. Et sic ordinata aut est vera aut falsa. Si vera, verum propositum repugnat et est negandum; si falsa, non repugnat et est concedendum. Si autem aliquid sit preconcessum vel prenegatum, sumatur positum cum concesso, vel concessis si plura fuerint concessa, et cum opposito negati vel oppositis, et ponatur in antecedente conditionalis et oppositum veri propositi in consequente. Illa conditionalis aut erit vera aut false. Si vera, verum propositum repugnat et est negandum; si falsa non repugnat et est concedendum. Si autem falsum sit propositum, aut aliquid est preconcessum aut prenegatum aut nichil. Si nichil, sumatur positum in antecedente conditionalis et falsum propositum in consequente. Illa conditionalis aut erit vera aut erit falsa. Si vera, falsum propositum est consequens et concedendum. Si falsa, non est sequens et negandum. Si autem aliquod fuerit preconcessum vel prenegatum vel aliqua, sumatur positum cum concesso, vel concessis, et opposito, vel oppositis, bene negati, vel bene negatorum, et ponatur in antecedente conditionalis et falsum propositum in consequente. Illa conditionalis aut erit vera aut falsa. Si vera, propositum primum est consequens et concedendum; si falsa, non est consequens et negandum.* de Rijk, 'Some Thirteenth Century Tracts', 29.



positum along with something(s) conceded and the opposite(s) of something(s) denied, and every true thing not repugnant to these must be conceded.

Rule 3.5.19. The opposite of the *positum*, every false thing not following from the *positum* along with something(s) conceded and the opposite(s) of something(s) denied, and every true thing repugnant to these must be denied.

The author illustrates this methodology with an example, which again we quote in full

despite its length:

> For example: Let it be posited that Antichrist exists. Thence [let] this [be posited]: 'Antichrist is', this is the *positum* and it is proposed under the same form of speech under which it was posited. Therefore it is conceded... Then it is put forward 'The Antichrist is colored'. This is something false... See therefore whether it may follow, putting it in the consequent of a conditional and the *positum* in the antecedent, in this way: 'If Antichrist exists, Antichrist is colored'. It is agreed that this is true. Therefore, Antichrist being colored is a consequence of the *positum*. Therefore, it should be conceded. You must put it in this way: 'Antichrist is white'. This is something false. You must consider whether something may be previously conceded or previously negated, or nothing. It is agreed that something is previously conceded. See therefore whether it follows, by fashioning a conditional, namely that 'if Antichrist [is] and Antichrist is colored, Antichrist is white'. This conditional is false. Therefore 'Antichrist being colored' is false and doesn't follow from the positum. Therefore it must be denied. Therefore it may be denied. Then so: 'Antichrist is not neutral'. This is something true. And it is agreed that something is previously conceded and something previously negated. See therefore whether Antichrist not being neutral may be repugnant, by putting down to the opposite of this in the consequent of a conditional, in this way 'if Antichrist [is], and Antichrist is colored, and Antichrist is not white, Antichrist is neutral'. This conditional is false. Therefore Antichrist being neutral does not follow. Therefore Antichrist being not neutral is not repugnant. And it is true. Therefore it is true and not repugnant. Therefore it must be conceded. With what is conceded it may be posited that 'Antichrist is black'. This is something false. It is agreed that some things are conceded and something is negated. See therefore whether Antichrist being black is a consequence of the positum and the conceded things and the opposite of the negated thing, in this way: 'If Antichrist is and Antichrist is colored and Antichrist is not white and Antichrist is not neutral, Antichrist is black'. This conditional is true. Therefore Antichrist being black is a false proposition which follows. Therefore it must be conceded.[39]

[39] *Verbi gratia. Ponatur Antichristum* ⟨*esse*⟩. *Inde sic* 'Antichristus est', *hoc est positum et propositum sub eadem forma vocis sub qua fuit positum. Ergo est concedenda... Deinde proponitur hec* 'Antichristus est coloratus.' *Hoc est quoddam falsum... Vide ergo an sit sequens, ponens ipsum in consequente conditionalis et positum in antecedente, hoc modo:* 'si Antichristuts est, Antichristus est coloratus'. *Constat quod hec est vera. Ergo Antichristum esse coloratum est sequens ex posito. Concedatur ergo. Inde sic:* 'Antichristus est albus.' *Hoc est quoddam falsum. Considera an aliquid sit preconcessum. Vide ergo an sit sequens, faciendo*

0.	Antichrist exists.	Admit	False, but consistent.
1.	Antichrist is.	Concede	Same as the *positum*.
2.	Antichrist is colored.	Concede	Follows from (1).
3.	Antichrist is white.	Deny	False, and does not follow from (1), (2).
4.	Antichrist is not neutral.	Concede	True, and not inconsistent with (1), (2), and the negation of (3).
5.	Antichrist is black.	Concede	False, but follows from (1), (2), the negation of (3), (4).

Figure 3.7: A schematic representation of the *positio*.

A summary of this example is given in Figure 3.7.

The keen-eyed reader will have noted that nothing is said, either in the rules or in the methodology, about the standard third response, beside concession and denial, of *Proba!* ("Prove it!"). In later *obligationes*, e.g., Burley, *Proba!* is the appropriate response when the proposition in question is irrelevant and the truth value is unknown, and it is explicitly mentioned in the rules. While it is omitted from the rules quoted above, following this example the author notes that it is possible for a proposition to be put forward of which it is doubtful whether it is true or false. In that case, "if it follows, you should concede it; if repugnant, you should deny it; if neither, you should respond 'prove it!'."[40] This comment, following directly on the heels of the first example of determinate *positio*, shows that Spade's conclusion that *Proba!* was not used in *positio* in *Obligationes Parisienses* [Spade, 1982, fnn. 17, 22] is simply mistaken.

The author uses this *positio* example to draw a contrast with *rei veritas*, as a way of introducing the latter species:

---

conditionalem, hanc videlicet: 'si Antichristus ⟨est⟩ et Antichristus est coloratus, Antichristus est albus.' *Hec conditionalis est falsa. Ergo*, Antichristum esse album *est falsum non sequens ex posito. Ergo est negandum. Negetur ergo. Deinde sic* 'Antichristus non est medius.' *Hoc est quoddam verum. Et constat quod aliquid est preconcessum et aliquid prenegatum. Vide ergo an* Antichristum non esse medium *sit repugnans, ponendo oppositum eius in consequente conditionalis, hoc modo* 'si Antichristus ⟨est⟩ et Antichristus est coloratus et Antichristus non est albus, Antichristus est medius.' *Hec conditionalis est falsa. Ergo Antichristum esse medium est non sequens. Ergo Antichristum non esse medium non ⟨est⟩ repugnans. Et est verum. Ergo est verum non repugnans. Ergo est concedendum. Qua concessa proponatur hec:* 'Antichristus est niger.' *Hoc est quoddam falsum. Et constat quod aliqua sunt concessa et aliquid est negatum. Vide ergo* Antichristum esse nigrum *an sit sequens ex posito et concessis et opposito negati, hoc modo:* 'si Antichristus est et Antichristus est coloratus et Antichristus non est albus et Antichristus non est medius, Antichristus est niger.' *Hec conditionalis est vera. Ergo* Antichristum esse nigrum *est falsum sequens. Ergo est concedendum.* ibid., 29.

[40] ibid., 30.

> And *rei veritas* differs from *positio* because when *rei veritas* has been done, then concerning any thing irrelevant or not following, then it is not to be denied, but on the other hand if *positio* has been done, then it is to be denied. When it is said '*rei veritas* is that Antichrist exists', then concerning this: 'Antichrist is white' the response is 'prove it!', but if *positio* has been done the response to the same is: 'It is false!'[41]

What the author is pointing out is that when a disputation is prefaced with a statement about "the truth of things" (*rei veritas*) or by saying "let it be true that..." (*sit verum*), this effectively shifts our focus during the disputation to only such situations where the *positum* is true, in a way that admitting a *positum*, or conceding or denying a proposition, does not. Suppose that, in *positio*, Res has conceded 'Antichrist exists' (even though this is false). If, then, Opponent puts forward 'Antichrist is white', this statement is irrelevant, and because it is false (since Antichrist does not exist, he cannot be colored), it should be denied. However, if "the truth of things" is that Antichrist exists, then when Opponent asserts that he is white, the correct answer is "prove it". If Antichrist exists, then he must be colored, but whether he is colored white or colored black is doubtful. In the next section, we make these differences in reasoning explicit.

How *sit verum* is counterfactual when *positio* isn't. In other work, we have shown how possible worlds models, suitably augmented with dynamic and epistemic operators, can provide a formal framework for the evaluation of *obligationes* [Uckelman, 2011a, Uckelman, 2011b, Uckelman, 2013].[42] This framework allows us to make explicit the knowledge of the Opponent and Res (both propositional and higher-order) and truths about the actual world; the inferential relations between propositions; and the actions (and their consequences) of the Res, continually reduce the number of "possible" worlds in the model by removing those which are inconsistent with his actions, e.g., to deny φ is to remove all worlds where φ is true. (For the complete technical details, we direct the reader to the sources in footnote 42, contenting ourselves here with an informal sketch.) We begin with a set of possible worlds, one of which represents the actual world, and the others of which are different combinations of truth values of the propositions involved in the disputation. (Propositions which are not proposed in the disputation are truly irrelevant and need not

[41] *Et* rei veritas *differt a positione quia rei veritate facta, circa aliquod impertinens non sequens, non est negandum, facta autem positione est negandum. Unde dicto* "rei veritas est quod Antichristus est"*, ad hanc* "Antichristus est albus" *respondendum est* "Proba!"*, sed facta positione ad eandem respondendum est:* "falsum est". de Rijk, 'Some Thirteenth Century Tracts", 28.

[42] Yrjönsuuri also motivates the use of possible worlds semantics in the analysis of *obligationes* in Mikko Yrjönsuuri, 'The Role of the Casus in Some Fourteenth Century Treatises on Sophismata and Obligationes', *Argumentationstheorie. Scholastische Forschungen zu den logischen und semantischen Regeln korrekten Folgerns*, ed. K. Jacobi, (Leiden, New York, Köln: E.J. Brill, 1993), 301–321.

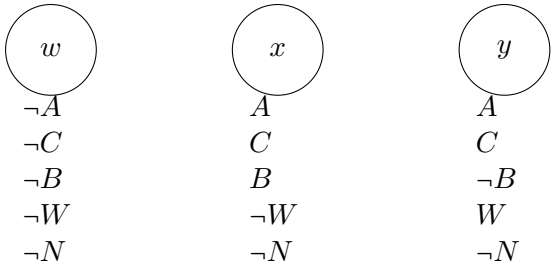


Figure 3.8: A minimal model.

A:	'Antichrist exists'
B:	'Antichrist is black'
C:	'Antichrist is colored'
N:	'Antichrist is neutral'
W:	'Antichrist is white'

Figure 3.9: Key to propositions

be taken into account.) For example, in Figure 3.8, there are three possible worlds, one of which, the actual world w where "Antichrist" exists is false and hence he is neither colored, nor black, nor white, nor neutral, since what doesn't exist doesn't have any color. In the other two worlds, x and y, Antichrist does exist, and hence is colored, but in one he is black and in the other he is white. All worlds are considered possible with respect to each other, so we have omitted the usual arrows.

In such a setting, a *positio* disputation is one (as Spade rightly points out) "incompletely specifying—but more and more completely as the disputation proceeds—a class of possible worlds in which the positum is true but that otherwise differ as little as possible from what, for all the respondent knows, the actual world is like."[43] Each action of the Res successfully removes from the set of possible worlds those worlds which are no longer 'possible' given what has happened in the disputation so far, thus constraining his future actions. This way of viewing *obligationes* shares many similarities to approaches for epistemic modals in dynamics semantics; as Asher and McCready note, "the available epistemic possibilities partly depend on what information has already been introduced in the discourse" [Asher and McCready, 2007]. In both *positio* and *sit verum*, the available responses of the Res depend on his actions in the disputation so far.

[43] [Spade, 1982, p. 11]. King repeats this point, cf. Peter King, 'Mediaeval Thought-Experiments', 52.

Figure 3.10: After Res has conceded A.

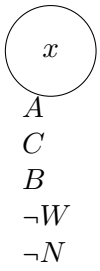


Figure 3.11: After Res has denied W.

For example, we can illustrate how the *positio* schematized in Figure 3.7 works by using the model in Figure 3.8. The *positum* A is false, but not impossible—since it is true in worlds x and y—and is admitted. In the next round A is put forward again, and since it is identical with the *positum*, it is conceded. The effect of this concession is that all worlds where A is false are removed from the model, i.e., Figure 3.10. Next, C is put forward. This is true in all remaining worlds, and hence follows from A; it is conceded, and the model does not change. Next, W is put forward. This does *not* follow from A, since in some worlds where A is true, W is false, namely, x. Per the rules for *positio*, it is then asked whether W is actually true or false; and the answer is 'false', as is clear by inspection of the actual world w in Figure 3.8. Since it is irrelevant and false, it must be denied; such denial results in Figure 3.11. The Opponent then puts forward $\neg N$; the analysis of this is the same as for C. Lastly, he puts forward B, which necessarily follows from $A \wedge C \wedge \neg W \wedge \neg N$, as is clear from Figure 3.11. Res concedes.

If, however, *rei veritas est quod Antichristus est*, then the Res should respond to W differently. He concedes A and C, since these both follow from A, but then when W is put forward and noted as irrelevant, the point of evaluation is no longer the actual world, but

instead the entire set of currently possible worlds, i.e., all the worlds in Figure (3.10). Here, there is one world, x, where W is false and one world, y, where it is true. Since neither counterfactual "if $A \wedge C$ were true, then W would be true" nor "if $A \wedge C$ were true, then $\neg W$ would be true" is true, the Res should reply "*Proba!*"[44] In fact, one of the interesting consequences of the *sit verum* rules is that irrelevant propositions will always be doubted; for if W were true in all worlds in Figure (3.10), or if it were false in all, then it would not be irrelevant, but relevant, and conceded or denied accordingly. This may seem like a minor point, but in fact it means that the counterfactual analysis shifts the disputation from a dynamic setting to a static one. Part of what makes *positio* difficult is that relevance must be recalculated any time that an irrelevant proposition has been conceded or denied. With the change introduced here, *sit verum* loses this dynamic aspect by doubting all irrelevant sentences. Thus, there is no longer any need to recalculate relevance to take into account sentences that might have become relevant on the basis of conceding or denying something irrelevant.

The author of *Obligationes Parisienses* gives no indication of recognizing this shift, nor indeed did other later authors ever comment on this. When Swyneshed introduced his *responsio nova* whereby the relevance of sentences was calculated at the beginning of the disputation and did not change thereafter, providing a static framework within which Res responded, his approach was taken as a radical change, sparking a back-lash from those in the Burleyan tradition who did not like the change (cf. [Spade and Yrjönsuuri, 2013, §3.2]).

3.6 Fallacies, insolubles, and sophisms

Around 1120, Boëthius' translation of the *Sophistici Elenchi* is rediscovered. Aristotelian discussions of fallacies. The medieval solution to these paradoxes followed the Aristotelian fallacy of *secundum quid et simpliciter*. The most famous insoluble: **the Liar**.

This sentence is false.

φ : φ is false.

[44] Formally, the rule for irrelevant propositions changes from:

If $\mathfrak{M}, w^* \vDash K_{\mathsf{Res}}\varphi$	then	$R(\varphi)$: concede
If $\mathfrak{M}, w^* \vDash K_{\mathsf{Res}}\neg\varphi$	then	$R(\varphi)$: deny
If $\mathfrak{M}, w^* \vDash \neg(K_{\mathsf{Res}}\varphi \vee K_{\mathsf{Res}}\neg\varphi)$	then	$R(\varphi)$: doubt

to

If $\mathfrak{M} \restriction \Gamma \vDash \varphi$	then	$R(\varphi)$: concede
If $\mathfrak{M} \restriction \Gamma \vDash \neg\varphi$	then	$R(\varphi)$: deny
If $\mathfrak{M} \restriction \Gamma \nvDash \varphi$ and $\mathfrak{M} \restriction \Gamma \nvDash \neg\varphi$	then	$R(\varphi)$: doubt

For notation, see [Uckelman, 2011b, §6].

In the early literature on insolubles, there are five solution types for this paradox:

- *secundum quid et simpliciter.*
- *transcasus.*
- Distinction between the exercised act and the signified act.[45]
- *restrictio.*
- *cassatio.*

Solutions

- *secundum quid et simpliciter.*
  Solution. Unclear.
  - Mentioned by Aristotle (*Sophistici Elenchi*, 180b2-3).
- *transcasus.*
  Solution. The Liar sentence is false.
  - Derives from the Stoic *metaptosis*: differing truth-values over time.
  - When I say "I am speaking a falsehood" I am referring to what I said immediately preceding to that sentence.
  - If I didn't say anything before that, then the sentence is just false.
- Distinction between the exercised act and the signified act.
  Solution. The Liar sentence is false.
  - Johannes Duns Scotus, *Questiones.*
  - The exercised act of the liar is "speaking the truth".
  - The signified act of the liar is "speaking a falsehood".
  - The liar expresses something which is not the truth, so it is false.
- *restrictio.*
  Solution. The Liar sentence does not have a truth value.
  - The *restringentes* do not allow assignment of truth-values to sentences with self-reference.

[45] Paul of Venice, following Thomas Bradwardine, connects this to the fallacy of equivocation.

  - Not only the Liar, but also the following insoluble: $\varphi : \psi$ is false. $\psi : \varphi$ is true (linked liars)
    ... and ... "This sentence has five words."

- *cassatio.*
  Solution. The Liar sentence does not have a truth value.

  - If you are uttering an insoluble, you are saying nothing.
  - Therefore an insoluble has the same truth value as the empty utterance: none.

Self-reference isn't the problem: it's reference to truth-values.

The most productive era in the theory of insolubles was from 1320 to 1350. The following people all contributed to the material that we have on theories of insolubles:

- Thomas Bradwardine (c. 1295–1349).
- Richard Kilvington (c. 1302–1361).
- Roger Swyneshed (mid 14th century).
- William Heytesbury (c. 1310–1372).
- Gregory of Rimini (mid 14th century).
- John Wyclif (c. 1330–1384).
- Peter of Ailly (*Petrus de Alliaco*; 1350–1420).

We will discus some of these people and their views in detail.

Thomas Bradwardine (c. 1295–1349).

> [The old logicians] understood little or nothing about insolubles. After them then arose the prince of modern philosophers of nature, namely Master Thomas Bradwardine. He was the first one to discover something worthwhile about insolubles (Ralph Strode, between 1350 and 1375).

- *Insolubilia*: 1321–1324; a new critical edition has recently been published, [Bradwardine,]. Spade describes the theory presented in this text as an "adverbial theory of propositional signification" [Spade, 1996, p. 178–85].
- Every sentence signifies that it is true.

- A sentence is *true* if and only if everything that it signifies is true (*sicut est*). A sentence is *false* if and only if there is something that it signifies which is false (*aliter quam est*).
- The Liar sentence signifies that it is false.

Roger Swyneshed (mid 14th century).

- A sentence is true if and only if it signifies *sicut est* and if it not *self-falsifying*. Self-falsifying sentences are always false. The definition of self-falsification is independent from any issue of self-reference.
- The Liar is self-falsifying, so it is false.
- *Consequences of Swyneshed's definition of truth.*
  - φ : φ is false.
  - ψ : φ is not false.
  - φ is false as it is self-falsifying. But then ψ is false, too. But φ and ψ are contradictories.

Another consequence for Swyneshed is that some pairs of contradictory sentences are both false.

William Heytesbury (c. 1310–1372).

- *1335. Regulae solvendi sophismata*, influenced by Kilvington's *Sophismata*.
- *The source of the paradox according to Heytesbury:* The Liar "φ : φ is false" is only paradoxical since we want to retain the usual theory of signification for it. If we give that up, there is no paradox. For example, φ could signify "*Socrates currit*" which is free of paradoxes.
- But φ cannot be evaluated according to the usual theory of signification. Therefore, anyone who utters φ must have some other hidden signification in mind. There is no way to analyze φ further before we know which one this is.

Some of the problems concerning the semantics of syncategoremata are part of the theory of sophismata:

- **Scenario 1.** Socrates enters the room and sees everyone. He leaves. Plato leaves the room. Socrates returns and sees everyone except for Plato.
  Socrates videt Platonem.

- **Scenario 2.** Plato is not in the room at all. Socrates enters the room twice and sees everyone in there.
  Socrates non videt Platonem.

3.7 The 14th century

14th and 15th century. Philosophy sharply divided into *via antiqua* and *via moderna*.

- "speculative grammar" based on *modi*.
- Boëthius of Dacia (d. 1290)
- Pierre d'Auvergne (d. 1303)
- Martin of Dacia (d. 1304)
- Thomas of Erfurt (c. 1330)
- Johannes Aurifaber (c. 1330)

Chapter 4

Logic in ancient and medieval East and South-East Asia [last modified 27 Mar 26]

4.1 Introduction

In Chapter 1, we offered a definition of logic as "the study of good arguments" (Def. 1.1.1). Much of this book is devoted to an understanding of "good" as "deductively valid". But while this is often how logic is treated, it is a narrow view of what counts as logic. We might think that while this makes up a "gold standard" for goodness of an argument, it is too strong, and instead there are many arguments that we make and use in daily life which we consider good, even if they fall short of that gold standard.

So let us revamp our perspective on logic into something broader. If we think back to our pre-theoretic ideas about arguments, one thing that is present in them which is *not* in the standard philosophical treatment is the idea of argumentation being a multi-agent activity: There are at least two people, each of which have different knowledge and beliefs, and either one is trying to convince the other of something, or both are working together to reveal a previously unknown truth. If we look at arguments and argumentation in this context, then there are two new standards by which we can judge arguments: (a) An argument for φ is good if it provides one with reasonable grounds for acting as if he knows φ; (b) an argument for φ is good if it convinces the other of φ in the course of the interaction.

Thus, we can distinguish a three-fold view of what arguments are about:

1. Ontic: This view of logic treats facts and the actual relations between facts as basic.

2. Epistemic: This view focuses on what we know.

3. Dialectic: This view focuses on what we accept for the sake of argument, even if we don't know it.

Standard Western approaches to logic generally take place at level (1), the relations between facts, rather than at the level of knowledge or dialectical acceptance—even when we look at epistemic logic (Chapter 12), analysis takes place at the factive level of what we *actually* know. An exception to this general tendency are the logical *obligationes* games (cf. Chapter 3) developed in the Middle Ages, which operate at the third level.

In this chapter, we look at logical or argumentative traditions that focus on levels (2) and (3).

4.2 Logic in Buddhist epistemology

Epistemology in the Indian philosophical tradition focuses on *pramāṇa* or the "means of knowledge": By what means/methods/mechanisms/instruments can we come to know things, or can we justify the knowledge that we have?

There are a number of different *pramāṇa*, that are discussed in the various traditions (including the Hindu traditions, the Buddhist traditions, and the Jain traditions), with the most common including the following six:

1. *anumāṇa*: inference.

2. *pratyakṣa*: perception.

3. *śabda*: testimony.

4. *upamāṇa*: comparison/analogy.

5. *arthāppati*: derivation from circumstance.

6. *anupalabdhi*: non-perception or negative proof.

What is important to note about all these *pramāṇa* is that they are *defeasible*: As tools, they give one reasonable evidence for knowledge, that is, conclusions you are justified in acting on, but they do not have the same strength as valid arguments, and, more importantly, they were not designed to be so.

Different traditions accepted all or only some of these *pramāṇa* as legitimate means of knowledge; however, all of them accepted *anumāṇa* or inference, which is what our focus will be on.

4.2.1 Buddhist accounts of *anumāṇa*

In the Buddhist tradition, what counts as a "good" *anumāṇa*? There is a long and varied history covering more than a millennium, but we will focus, here, on three people who contributed to this development, Vasubandhu, Dignāga, and Dharmakīrti.

Vasubandhu

Vasubandhu was a Buddhist philosopher who lived in the 4th–5th C. For information on his life and biography, see [Gold, 2018]. Amongst his numerous contributions to the development of Buddhist philosophy, Vasubandhu was also influential on the development of formal logic, via his treatise *Vādavidhi* (*Rules for Debate* or *Method for Argumentation*). Only fragments of this treatise remains, but it is translated into English in [Anacker, 1984].

Vasubandhu was the first to give a *formal* characterization of *anumāṇa*, or good inference. For Vasubandhu, an *anumāṇa* consists in three parts:

- the *pakṣa*, or the subject under investigation.[1] Call this p.
- the *hetu*, or ground property. Call this H.
- the *sādhya*, or target property. Call this S.

These three parts are arranged into a three-step argument form:

1. thesis (what is to be proven): "p has S".
2. ground (premise which grounds the thesis): "p has H".
3. indispensability (the warrant which gets you from the ground to the thesis): "H pervades S" (or "Whatever has H has S").

The third statement is also called a statement of "positive concomitance" or "invariable concomitance" [Anacker, 1984, p. 38]. The warrant is often not stated explicitly, but is instead indicated by given an example of something where both H and S co-occur. An example of such an argument that Vasubandhu gives is the following:

> Sounds of speech are non-eternal [thesis], because of their state-of-arising-due-to-an-effort [ground], *like a pot* [example of positive concomitance] [and] ... Whatever has come about through effort is not eternal [warrant] [Anacker, 1984, p. 39, annotations added].

[1] "The *pakṣa* is the object of sense or understanding one wishes to investigate" [Anacker, 1984, p. 38].

Here are two further examples of arguments which follow this form.

Example 4.2.1.

1. thesis: This chair has a color.
2. ground: This chair has the color red.
3. warrant: Whatever has the color red has a color.

Example 4.2.2.

1. thesis: This chair has a brain.
2. ground: This chair was made in outer space.
3. warrant: Whatever was made in outer space has a brain.

Clearly, Example 4.2.1 is a good argument, but Example 4.2.2 is not. What this shows us is that while the three-step form of argument is necessary for an inference to be a good one, it is not sufficient. Further constraints must be met if the justification or warrant is not to be a pseudo-justification or fallacious warrant. There are three types of pseudo-justifications [Anacker, 1984, p. 39]:

1. those which are not demonstrated, that is, "those where the characteristics stated in the exemplifcation do not exist".
2. those which are not sufficiently certain.
3. those which incur a self-contradiction.

An example of an undemonstrated *hetu* is "sound is non-eternal, because it is perceived by the eye"; this is an unwarranted inference because sound is not in fact perceived by the eye. An example of an insufficiently certain *hetu* is "it is eternal, because it is without a body"; this in uncertain because the referent of "it" is not known. Vasubandhu gives two examples of *hetu*s that incur self-contradiction, one which is contradictory for the Vaiśeṣika ("It is not eternal, because it is perceived through the senses") and one which is contradictory for the Sāṇkhya ("The effect is contained in the cause, because it comes to be (through the cause)").

Thus, according to Vasubandhu, an inference must satisfy the following three requirements in order to be a good inference (necessary condition):

1. The *hetu* must occur in the *pakṣa*.

2. The *hetu* must occur in similar examples, i.e., in examples which have the *sādhya*.

3. The *hetu* must not occur in the dissimilar examples, i.e., ones without the *sādhya*.

A *hetu* that meets all three of these conditions is called a *trairūpya-hetu*. The first requirement of the *trairūpya-hetu* is already expressed in the second step of the argument, the statement of the ground; thus, if the ground is false then the inference will not be a good one. The second and third requirements must be checked explicitly. Example 4.2.2 fails both the first requirement (the *hetu* does not in fact occur in the *pakṣa*) as well as the third requirement, as there are things which were made in outer space that do not have brains, i.e., the *hetu* does occur in dissimilar cases.

Dignāga and Śaṇkarasvāmin

Dignāga (480–540) built on and substantially revised and expanded Vasubandhu's foundation. Not all of his works survive, and not all of them are available in English translation.

Instead, in this section we focus on a work by Śaṇkarasvāmin, said to be a student of Dignāga, the *Nyāyapraveśa* [Tachikawa, 1971]. Dignāga takes the three-step argument form found in Vasubandhu as his basis, but adds a further requirement to the third step, namely, the statement of a *dṛṣṭānta* or 'example' of the positive concomitance; while some of Vasubandhu's arguments included such an example, it was not explicitly a part of the formal structure of the argument, without which the argument would not be formally good. In doing so, he makes the second requirement of the *trairūpya-hetu* occur explicitly in the argument.

Here are two inferences following this form:

Example 4.2.3. 1. thesis: sound is non-eternal (sound = *pakṣa*; non-eternal = *sādhya*).

2. ground: because of resulting from effort (= *hetu*).

3. concomitance + example: Whatever results from effort is observed to be non-eternal, like a pot. (pot = *dṛṣṭānta*).

(We saw this inference above, in Vasubandhu.)

Example 4.2.4. 1. thesis: "there is fire on the mountain" (mountain = *pakṣa*; fire = *sādhya*).

2. ground: because there is smoke (= *hetu*).

3. concomitance + example: Wherever there is smoke, there is fire, like in a kitchen. (kitchen = *dṛṣṭānta*).

Sometimes, an example of negative concomitance is also given: "unlike in a lake" (where there is neither smoke nor fire). When a negative *dṛṣṭānta* is provided, then the third requirement of the *trairūpya-hetu* is made explicit.

Thus, giving examples of positive and negative concomitance directly incorporates the requirements of the *trairūpya-hetu* of Vasubandhu. Given a property H and a property S, there are nine different ways that they can be combined; only two of these ways result in a valid inference. These nine ways can be illustrated with what is called the *hetucakra* or the "Wheel of Reasons" (see Figure 4.1).

H occurs in:	(1)	all S, all $\neg S$	(2)	all S, no $\neg S$	(3)	all S, some $\neg S$
H occurs in:	(4)	no S, all $\neg S$	(5)	no S, no $\neg S$	(6)	no S, some $\neg S$
H occurs in	(7)	some S, all $\neg S$	(8)	some S, no $\neg S$	(9)	some S, some $\neg S$

Figure 4.1: Wheel of Reasons

The only cases which lead to good inferences are (2) and (8): It must be the case that the *hetu* never appears without the *sādhya*, so that excludes all possibilities but the middle column, and it must be the case that the *hetu* appears with the *sādhya* at least once, which excludes the middle cell from the middle column.

Dharmakīrti

With the works of Dharmakīrti (7th C), the formal development of logic reached a pinacle; Gillon describes "his extensive writing on epistemology in general and on reason and argument in particular" as "a watershed in classical India philosophy" [Gillon, 2016, §4].

4.3 Logic in ancient China

Note: This section is being drafted and currently exists in the form of minimal lecture notes. It will be substantially expanded in the future.

4.3.1 The "White horse is not horse" argument

In a text attributed to Gōngsūn Lóng (c320–250 BCE), a series of arguments for the curious conclusion "white horse is not horse" is found. The text, the *Gōngsūn Lóngzǐ*, comprises a preface, three incomplete and/or corrupted chapters, and two short but very complex chapters. While this text was attributed to Gōngsūn Lóng, and the content certainly reflects views of the Warring States period (475–221 BCE), there is no clear evidence that Gōngsūn Lóng was an actual historical figure, and there is evidence that the text was not a static piece, but was continually developed and adapted, up through the early medieval period (6th C CE). The earliest extant version of the text is from the Daoist Canon (*Dàozàng*), which was completed in 1445 [Indraccolo, 2017, fn. 3].

Context of the argument:

- Argumentation vs. persuasion
- Involving the "Logicians" or the "Sophists" (*míngjiā*, lit. "experts on names")
- Conventionality of language and the rectification of names (cf. [Ren and Liu, 2019, p. 470]). "The white-horse paradox arose together with the need for more precise use of language, the rectification of names that many pre-Qin philosophers considered a prerequisite for rectifying the world" [Ren and Liu, 2019, p. 472].
- Relationship between language and the world.
- Gōngsūng Lóng's status as a "sophist" and the charge against him of "arguing for arguing's sake" [Indraccolo, 2017, p. 2].

What has been described as "the most famous case of argumentation in the history of Classical Chinese philosophy" [Indraccolo, 2017, p. 1] is found in the first chapter of the *Gōngsūn Lóngzǐ*. Though this is often called the "White horse is not horse *argument*," the text in fact is a series of seven back-and-forth exchanges, which can be grouped into four distinct arguments, each of when is aimed at establishing the admissability of asserting the conclusion "White horse is not horse" (*bái mǎ fēi mǎ*). How to translate?

In our presentation of the four arguments, we use the labels of [Ren and Liu, 2019].

Shape and Color Argument

The first argument, in §§1–4 of the text, begins with the assertion that "It is admissible to say 'white horse is not horse'." When the unnamed opponent questions the admissibility of this strange statement, Gōngsūn Lóng says it is admissible because 'horse' names body or shape, while 'white' names color—and body and color are not the same thing.

Underlying Gōngsūn Lóng's argument is the idea that two different names cannot refer to the same thing—this is the 'one name, one thing' principle that Ren and Liu forces "an isomorphism between language and the world" [Ren and Liu, 2019, p. 476]: If we call something a 'white horse' instead of a 'horse,' then we have to be saying something different by each of these, or we would not have two different names.

This argument highlights different ways "is" is used:

- is of identity
- is of existence
- is of predication

Intension vs. extension

Question 4.3.1. Is this "mere equivocation"?

Seeking Horse Argument

The second argument, in §§5–6, demonstrates how adding qualification changes meaning. The second argument can be read as expressing a *reductio* argument:

(1) Suppose "white horse is horse." (Assumption)

(2) If I am looking for a horse, then a horse of any color will do. (Fact)

(3) If I am looking for a white horse, then a horse of any color will do. (Substituting (1) into (2))

(4) If I am looking for a white horse, only a white horse will do. (Fact)

(5) A black horse is not a white horse. (Fact)

(6) A black horse will do. (From 3)

(7) A black horse will not do. (From 4 and 5)

(8) That's absurd.

(9) So it can't be that white horse is horse.

Combination Argument

Argument 3: "White horse" includes "horse"! This is an argument about names, not about their referents. So then how can we ever say that a white horse *is* a horse? How can we talk about the referents, rather than the names?

Against the conventionality of language: "Gongsung Long discovered that some names are not arbitrary; they are combinations of other names" [Ren and Liu, 2019, p. 473].

Separating White Argument and Yellow Horse Argument

4: But all horses have color! White horses vs. dun horses.

Highlights of the argument

"We argue that '(a) white horse is not (a) horse' is not a denial of member-class relationship, but that the two names 'white horse' and 'horse' can refer to the same thing, though they may correspond to the same individual" [Ren and Liu, 2019, p. 466].

- Dialogical in nature (functioning at level (3), discussed above):
  - opponent (unnamed) and proponent (Gōngsūng Lóng).
  - connections to medieval European disputations (cf. Section 3.5).
  - connections to the "thesis-first" approach (cf. Section 4.4).
  - *NOT* like Western logic.
- Focus on admissability rather than truth (cf. [Indraccolo, 2017, p. 3]).
- The role of logic in society. **Names matter**—what we choose to call things matter.

  "He [Hansen] confuses how language divides the world with how people conceive of the world," [Ren and Liu, 2019, p. 467].
- When discussing the fourth argument, Indraccolo says that Gōngsūng Lóng uses an "impure syllogism" [Indraccolo, 2017, p. 5]. To what extent should we torture language into syllogisms?
  - torture English-language arguments into syllogistic form(s)
  - torture *anumāṇa* into syllogistic form(s)
  - torture the white horse exchange into syllogistic form(s)

  How can we apply the theories and concepts of Western philosophy without misreading them into ancient Chinese thought? Specifically, what light may Western theories of mass nouns, mereological epistemology, and the discussion about realism and nominalism bring to our reading of the "White Horse Dialogue"? [Ren and Liu, 2019, p. 465].

4.4 Analysis and discussion

Is this logic? That is, does it make sense for what has been discussed in the previous two sections to have been included in a textbook on logic? The answer to this question will depend on what gets to count as logic, that is, on the titular question asked by this book.

Furthermore, when we make decisions about what gets to count as logic, thes eare also decisions about *who* gets to count as a (a) logicians and (b) logical.

A narrow definition of logic as deductive validity in the western tradition requires us to ask whether the systems discussed in the previous two sections are mappable onto one of the (usually classical) systems of deductive validity. With this in mind, we now turn to the question of how this approach to argumentation compares to Western approaches — and from that we will look at the question of whether we should be comparing them in the first place.

First, note that the building blocks of *anumāṇa* are objects (that is, locations, properties), and relations (that is, pervasion and occurrence). The statements in an inference either assert the occurrence of a property in a location, or the pervasion of two properties.

In 1824, H.T. Colebrooke reported to the Royal Asiatic Society of his discovery of the 'Hindu syllogism', a "schema for correct reasoning as described in the early Indian text" [Ganeri, 2001, p. vii]. A century later, H.N. Randle published an article on the 'Indian Syllogism', in *Mind*. Both of these events were landmarks in the reception of the Indian (both Buddhist and Hindu) traditions of reasoning in the west, and in both, the use of the term 'syllogism' is both problematic and indicative.

Colebrooke's lecture challenged the belief "that Europe, as a consequence of its classical cultural inheritance, had a privileged position of intellectual authority" and by using the term 'syllogism' to describe the inference forms outlined in the previous section, he "implied that the ancient Indians were as aware of the syllogism as the ancient Greeks", and thus "the problem was whether Colebrooke's discovery could be made consistent with the European self-understanding of its intellectual superiority over its colonies" [Ganeri, 2001, p.5–6]. If the Indian 'syllogism' was really a 'syllogism', then this posed a threat to European intellectual superiority: For if non-European cultures had the same concept of 'syllogism' (at that point, the 'Gold Standard' for argumentation), then it would be hard to say that the European tradition was superior.

The question then became to what extent the Indian 'syllogism' was a syllogism. How would one of Vasubandhu or Dignāga's arguments look like in syllogistic form? Perhaps:

> Some mountain is a smoky thing.
>
> Every smoky thing is a fiery thing.

Some mountain is a fiery thing.

It could be objected that this is not a good rendition of the original argument: The premises come before the conclusion (instead of the conclusion (the thesis) before the premises (the ground and the warrant); the examples are lost; the original argument talked about a specific individual mountain, rather than the 'some' mountain discussed here. All of these might be taken as objections to formalizing the original argument in syllogistic fashion like this. But they might also be taken as defects in the original argument: If that argument was supposed to be a syllogism, it's a "bad" syllogism, because look at how much we had to change it to get it into syllogistic form. As Ganeri says in reference to Colebrooke, "accepting that Indian logic is indeed syllogistic, the Indian version is found wanting in two respects: Of the five steps, two are 'manifestly superfluous'" and the examples of positive and negative concommittance are redundant [Ganeri, 2001, p. 8]. Hamilton called the 5-step argument a "clumsy agglutination" [Ganeri, 2001, p. 12]. Thus, British colonizers could retain their European superiority by showing that even if the colonized had a concept of syllogism, it was in some sense defective.

Similar claims about the "defectiveness" of the White Horse dialogue can also be found—and amongst much more modern commentators. Indraccolo says that Gōngsūn Lóng "draws a parallel comparison establishing a direct relationship betwween horses of different colors in a sort of *impure syllogism*, deliberately distorting and twisting the opponent's argument" [Indraccolo, 2017, p. 5, emphasis added].

There are two responses that can be made to these charges of bad syllogizing. The first is to say that judging these arguments against the standard of the Aristotelian syllogism is not appropriate, given that we now know that that isn't exactly the gold standard. We have predicate logic, which is just more expressive (in that it can represent and prove as valid arguments that cannot be represented in a pure syllogistic logic). Perhaps the correct way to represent *anumādna* is using predicate logic, giving us something like:

$$\forall x(Sx \to Fx)$$

$$S\mathsf{m}$$

$$F\mathsf{m}$$

This has an advantage over the previous that it least we can name and specify which mountain we're talking about, namely m rather than, say, n. But it isn't much better: the statements are still in the wrong order, the examples are missing, and the first premise is talking about objects, rather than properties. What next?

Maybe... we shouldn't have been asking this question. Maybe instead:

1. Why judge Indian and Chinese argumentation on the basis of Western logical standards? Why not judge Western logic against the standards of Indian or Chinese argumentation? (That is, why *by default* assume that the western standard is the better standard?)

2. Why judge Indian and Chinese argumentation on the basis of an ontic/ontological standard rather than on the basis of an epistemic standard, as discussed above?

But you might wish to reject even these questions. Must there be some standard against which all of these various systems can be compared/evaluated? If you say yes, you're probably a logical monist. If you say no, you're probably a pluralist.

Part II

Classical Logics

Chapter 5

Introduction to Part II [last modified 22 Oct 20]

In this book, we divide our study of logics into those which are classical and those which are non-classical. We call a logic *classical* if it is built upon the two following principles:

LEM The Law of Excluded Middle: Every sentence is, at any given time, either true or false.

PNC The Principle of Non-Contradiction: No sentence is both true and false at the same time.

Note that as stated, these are *meta*-principles of logic, articulated in the meta-language: How (and if) these principles are stated in an object language varies from logic to logic.

A consequence of these two principles is that every classical logic admits exactly two exhaustive and exclusive truth values, True and False, with every sentence being exactly one of these, no other option.

Proof. Let φ be a sentence in some classical logic. By LEM, it is either true or false.

Suppose it is true. Then by PNC, it is not false.

Suppose it is false. Then by PNC, it is not true.

Therefore, if φ is true, then it is *only* true and if φ is false, then it is *only* false; *and* it has to be at least one of these. □

Non-classical logics, which are the topic of Part IV, drop one or both of these assumptions. For example, intuitionistic logic (Chapter 19) drops the assumption of LEM, while paraconsistent logics (Chapter 20) reject PNC.

This definition of what counts as a "classical logic" is broader than is used in many other textbooks (cf., e.g., [Priest, 2001]), wherein any extension of propositional logic (Chapter 7) or predicate logic (Chapter 8) by the addition of new logical operators, including modal and tense operators, is considered non-classical. We prefer to distinguish classical logics from non-classical ones on the basis of the presence or absence of the two principles LEM and PNC, because it allows us to talk about, e.g., both classical and non-classical modal logics. The distinction is then one about the basic principles accepted, or not, by a system, rather than a question of the number or type of symbols included in the system.

Classical logic(s) are logics in which good arguments are those which are deductively valid.

Definition 5.0.1. A *deductively valid argument* is an argument that is necessarily truth-preserving. That is, there is no way in which the premises can be true without the conclusion being true.

'Truth-preserving' can be cashed out in two equivalent ways:

Semantic Every model[1] which makes all the premises true makes the conclusion true. (Every structure which is a model of the premises is a model of the conclusion.)

Proof-theoretic There is a derivation from the premises to the conclusion every step of which is truth-preserving.

If every proof-theoretic derivation of a system is truth-preserving, then that system is called *sound*. If every argument which is valid under the semantic definition can be derived in a particular system, then that system is called *complete*. If a system is both sound and complete, then these two ways of characterizing 'valid argument' are equivalent. Showing this equivalence is one of the most important meta-level results that can be proven about a system.

[1] 'Model' is a technical concept which must be defined specifically for any given logic.

Chapter 6

Term logic [last modified 05 May 26]

The earliest systematically developed system of logic is the syllogistic of Aristotle, which can be found in the *Prior Analytics* [Aristotle,]. Many readers will be familiar with what is probably one of the most famous arguments in philosophy:

Every human is mortal.
Socrates is human.
Therefore, Socrates is mortal.

This type of argument is often called a *syllogism*; however, Aristotle himself seems to treat the pair of premises alone, rather than the premises along with their conclusion, as a syllogism. For example, he notes that

> Some syllogisms... give more than one conclusion [Aristotle, , 53a4–6].

If a syllogism is defined to be a pair of premises and a conclusion, then it is impossible for a syllogism to have more than one conclusion.

The syllogistic is a term logic, not propositional or predicate logic, which means that its primary building blocks are terms (and copulae) rather than propositions or predicates and variables. While later extensions of the syllogistic, such as some medieval developments, allow terms that were proper names (like Socrates), Aristotle's own development, with which we begin, did not and admitted only general terms as subjects and predicates. Thus, a more typical presentation of a syllogism + conclusion is the following:

Every human is mortal.

Every Greek is a human.
Therefore, every Greek is mortal.

In this chapter, we abstract from the particular details of Greeks, men, and mortality, and look at the structure of syllogistic arguments. We define a formal language for expressing categorical statements and categorical syllogisms, define a notion of 'goodness' for syllogistic arguments, introduce a system of syllogistic proof, and prove a number of meta-level results about this system.

6.1 Language

In this section, we define a syllogistic language; the set of syntactically acceptable strings in this language (called well-formed formulas); and what counts as an argument in this language. First, we make a distinction in English between *categorematic* and *non-categorematic* terms.

Definition 6.1.1 (Categorematic term)**.** A *categorematic term* is a term that is significative (has meaning) on its own, in isolation from any other term.

Examples of categorematic terms in English include 'woman', 'dog', 'Greek', 'running', 'red', 'owns a donkey', 'taller than Fred', etc. Non-categorematic terms are everything else—prepositions, definite and indefinite articles, words like 'no' and 'none', or 'all' and 'some', etc.

Categorical terms are combined with connectives called 'copulae' into *categorical propositions*. There are four copulae that we are interested in here:

Definition 6.1.2 (Copulae)**.** A *copula* is an operator that connects two categorical terms. The four categorical copulae are the following:

- belongs to every
- belongs to no
- belongs to some
- doesn't belong to some

Definition 6.1.3. Let X and Y be arbitrary categorical terms. A *categorical proposition* is a statement having one of the following four forms:

a**:** X belongs to every Y.

e: X belongs to no Y.

i: X belongs to some Y.

o: X doesn't belong to some Y.

Thus, each categorical proposition has two terms connected by a copula. The term which follows the copula is called the 'subject' term and the term which precedes the copula is called the 'predicate' term.

Note that in ordinary English, the subject term usually comes before the copula and the predicate comes after, in reverse order to what we have specified above—where we have said above "X belongs to every Y", in ordinary, non-logical English we would usually say "Every Y is X". This non-standard order which inverts the order of the subject and the predicate is closer to Aristotle's original Greek presentation, and will allow us to articulate the syllogistic rules in a clearer and less confusing manner. However, since in ordinary English, the subject comes before the copula and the predicate comes after, in reverse order to what we have specified above, each of the statements above is equivalent in meaning to one of the following:

a: Every Y is a X.

e: No Y is a X.

i: Some Y is a X / Some Y are X.

o: Some Y is not a X / Some Y are not X.

The argument discussed in the introduction of this chapter would thus be properly rendered:

Mortal belongs to every man.
Man belongs to every Greek.
Therefore, mortal belongs to every Greek.

Converting sentences into the form of categorical propositions gives us a first step towards identifying the relevant structural features of arguments involving categorical propositions. We next go a step further by representing this syllogism in an abstract way by replacing the English categorematic terms with letters, e.g.:

O belongs to every M.
M belongs to every G.

O belongs to every G.

A second level of abstraction replaces the copulae with an abbreviation:

$$\frac{X a Y \quad Y a Z}{X a Z}$$

We have thus completely translated the syllogism from English into a logical notation. This notation forms the basis of our logical language, and our logical language will produce sentences at this second level of abstraction. We now make this precise.

Definition 6.1.4. A *term language* $\mathcal{L}_t$ consists in:

- A finite or infinite set of basic categorematic terms X, Y, Z, ...
- Four binary connectives, called 'copulae': a, e, i, o.

The basic terms represent categorematic terms in English. The basic terms are the *non-logical* symbols while the copulae are the *logical* symbols of the language.

The following defines categorical propositions, or well-formed formulas (wffs), within this language:

Definition 6.1.5 ($\mathcal{L}_t$ wffs)**.**

- If X and Y are distinct basic terms, then $X\,a\,Y$, $X\,e\,Y$, $X\,i\,Y$, and $X\,o\,Y$ are $\mathcal{L}_t$-wffs.
- Nothing else is an $\mathcal{L}_t$-wff.

Remember that the term which comes *first* in the categorical proposition represents the *predicate* in an ordinary English sentence, and the term which comes *second* represents the *subject* of the English sentence.

Each of the four types categorical proposition has two properties, quality and quantity, and each type is uniquely characterized by these two properties.

Definition 6.1.6. There are two qualities, *affirmative* and *negative*. Propositions with the copulae a and i are called affirmative; propositions with the copulae e and o are called negative.

Definition 6.1.7. There are two quantities, *universal* and *partial*[1]. Propositions with the copulae a and e are universal; propositions with the copulae i and o are partial.

Quality / : Quantity:	Affirmative	Negative
Universal	a	e
Partial	i	o

Thus we have the following characterizations of the four types:

The mnemonic letters to represent the four types of categorical proposition derive from the Latin words *affirmo* 'I affirm' and *nego* 'I deny', reflecting the quality of the propositions.

Our goal is to use this notation to translate sentences in English into the language $\mathcal{L}_t$, and then reason about the result. Many sentences in English fit neatly into the syntax of categorical propositions, e.g.,

- Every human is mortal.
- Some donut is not a vegetable.
- No peas are donuts.
- Some Greeks are not mortal.

Let $\mathcal{L}_t$ be a language consisting in the basic terms D, G, H, M, P, V (with D for 'donut', G for 'Greek', etc.). The above sentences are represented in $\mathcal{L}_t$ as follows:

- $M\, a\, H$
- $V\, o\, D$
- $D\, e\, P$
- $M\, o\, G$

There are also many sentences in English which do not immediately have the syntax of a categorical proposition, but which are equivalent to sentences that do. Consider the following:

1. All peas are vegetables.
2. Not all students run.
3. Some students attend every lecture.

[1] Other textbooks often call these 'particular'.

4. Some cat runs.

5. Some farmer doesn't own a donkey.

6. All gobonated bordures must have good contrast.

The first is not a categorical proposition because it uses 'All' instead of 'Every'; however, it is clearly equivalent to "Every pea is a vegetable", which is a categorical proposition. The second is not a categorical proposition, because 'Not all' is not a construction found in our list of copulae. It is, however, equivalent to "Some student does not run" (when we say that it is "equivalent", we mean that the two sentences always have the same truth-value; we will show that this is the case in the next section). This still is not a categorical proposition, because it, along with the remaining four, do not contain the verb 'is' or 'are'. However, again, each of these can be rewritten so that the verbs are converted to 'is' or 'are' and the predicate because 'thing which [original verb]s'. For example:

1. Every pea is a vegetable.

2. Some student is not a thing which runs.

3. Some students are students who attend every lecture.

4. Some cat is a thing which runs.

5. Some farmer is not a thing which owns a donkey.

6. Every gobonated bordure is a thing which must have good contrast.

Let $\mathcal{L}'_t$ be a language containing the basic terms $A, B, C, F, G^1, G^2, O, P, R, S, V$. Then we can represent these sentences as follows:

1. $V\,a\,P$

2. $R\,o\,S$

3. $A\,i\,S$

4. $R\,i\,C$

5. $O\,o\,F$

6. $G^2\,a\,G^1$

Every syllogism is constructed from categorical propositions. We define a syllogism as follows:[2]

Definition 6.1.8 (Syllogism)**.** A *syllogism* is a set Γ containing three categorical propositions consisting of three distinct terms, each of which occurs in exactly two of the propositions. Two of the propositions are the *premises* and the third is the *conclusion*.

Definition 6.1.9 (Major, minor, and middle terms)**.** The term which is the predicate of the conclusion is called the *major term*, and we often use P as a variable standing for the major term. The term which is the subject of the conclusion is called the *minor term*, and we often use S as a variable standing for the minor term. The term that doesn't occur in the conclusion but occurs in both premises is called the *middle term*, and we also often use M as a variable standing for the middle term.

Definition 6.1.10 (Major and minor premises)**.** Only one of the premises contains the major term. This one is called the *major* premise, the other one the *minor* premise.

It is a convention of the syllogistic that the major premise is always written first. How the terms are arranged in the premises—whether they are the subject or predicate—determines the figure of the syllogism.

Definition 6.1.11. The *figure* of a syllogism is the arrangement of the terms within the three categorical propositions. The four figures are depicted in Figure 6.1.

	Ist Figure	IInd Figure
Major premise:	P — M	M — P
Minor premise:	M — S :	M — S :
Conclusion:	P — S	P — S

	IIIrd Figure	IVth Figure
Major premise:	P — M	M — P
Minor premise:	S — M :	S — M :
Conclusion:	P — S	P — S

Figure 6.1: The Four Figures

[2]In this definition we follow medieval and modern commentators in identifying the syllogism with the premises *and conclusion*; this is in contradistinction to Aristotle who is some places appears to identify the syllogism with the pair of premises only, going on to then ask whether such syllogisms give rise to conclusions or not.

Definition 6.1.12. A syllogistic *mood* is a figure with three copulae inserted.

We have already seen one mood above, which can be represented in general form as follows:

$$\begin{array}{c} P\ a\ M \\ M\ a\ S \\ \hline P\ a\ S \end{array}$$

With four copulae and three slots, we get $4^3 = 64$ moods from each figure, i.e., $4 \times 64 = 256$ moods in total.

It should be clear that the language introduced here can handle only a very limited type of natural language expressions; there are a number of arguments which do not fit into this very narrowly defined structure. We say that this language and formal system has a limited *expressive power*: There are certain things that it can express, but there are many things which it cannot. We discuss the limitations of the expressive power of this term logic when we motivate propositional logic in §7.1.

6.2 Semantics

In this section we consider the semantics of the language introduced in the previous section. The semantical rules of a language govern how we are to interpret (give meaning to) the logical and non-logical symbols. Given a term language $\mathcal{L}_t$, the first component of the semantics will tell us how the basic terms are to be interpreted. We'll first give some informal examples of interpretations, before providing a precise semantics in terms of sets.

Example 6.2.1. Let $\mathcal{L}_1$ be a term language consisting of the four copulae and the non-logical basic terms A, B, C, D, and E. We let A stand for animals, B for bats, C for cats, D for dogs, and E for elephants.

Example 6.2.2. Let $\mathcal{L}_2$ be a term language consisting of the four copulae and the non-logical basic terms D, E, H, O, and W. We let D stand for dwarves, E for elves, H for hobbits, O for orcs, and W for wizards.

In these examples, we leave it to the reader's intuitive understanding of the English terms to know which items are picked out by which basic terms in the logical language. We can, however, be precise about what objects each basic term picks out, via sets.

Definition 6.2.3 (Interpretation)**.** Let $\mathcal{L}_t$ be a term language. An *interpretation of* $\mathcal{L}_t$ is a function I from the set of basic terms to non-empty[3] sets of objects. If the interpretation of a basic term X is the set Γ, we write $I(X) = \Gamma$.

Example 6.2.4. Let $\mathcal{L}_2$ be a term language consisting of the four copulae and the non-logical basic terms E, H, and W. We define I to be such that:

- $I(E)$ = {Elrond, Galadriel, Arwen, Legolas}
- $I(H)$ = {Frodo, Sam, Merry, Pippin, Bilbo}
- $I(W)$ = {Gandalf, Radagast, Saruman}

All of the above examples implicitly rely on real-world knowledge; but we are not required to consider the real world when coming up with interpretations. Because we are *defining* an interpretation, we are free to define it however we want, within the bounds of consistency. When we construct an interpretation, we are essentially building a world—a world where we are the designers and we get to choose what properties different objects have.

Example 6.2.5. Let $\mathcal{L}_3$ be a term language consisting of the four copulae and the non-logical basic terms A, B, and C. Let us define I to be such that:

- $I(A)$ = {Julia, Fred, Bob}
- $I(B)$ = {Julia}
- $I(C)$ = {Adelia, Lorn, Fred}

In this example, presumably you know nothing about who Adelia, Bob, Fred, Julia, and Lorn are (and you shouldn't; I've just made them up); you also don't have any idea what properties A, B, and C are *except* that A is a property shared by Julia, Fred, and Bob, B is a property that only Julia has out of the five, and C is a property that only Adelia, Lorn, and Fred have. What these properties *are* doesn't matter from the point of view of logic.

Whether we use an informal or a formal interpretation of the non-logical basic terms of a language, the interpretation of the logical symbols remains fixed. Informally, the truth conditions for each of the four types of categorical claims are exactly what you would expect

[3] On this restriction that the sets be non-empty, see the Digression below, p. 88.

them to be:

$X\,a\,Y$	is true iff	everything that is Y is also X
$X\,e\,Y$	is true iff	nothing is both Y and X
$X\,i\,Y$	is true iff	something is both Y and X
$X\,o\,Y$	is true iff	something is Y but not X

Formally, these conditions can be represented set-theoretically as follows (cf. §1.3.3):

Definition 6.2.6 (Truth conditions for categorical propositions)**.**

$X\,a\,Y$	is true iff	$I(Y) \subseteq I(X)$
$X\,e\,Y$	is true iff	$I(Y) \cap I(X) = \varnothing$
$X\,i\,Y$	is true iff	$I(Y) \cap I(X) \neq \varnothing$
$X\,o\,Y$	is true iff	$I(Y) \nsubseteq I(X)$

Let φ be any categorical proposition and I an interpretation of the non-logical symbols in φ. If φ is true on interpretation I, we write

$$I \vDash \varphi$$

Example 6.2.7. Let $\mathcal{L}_2$ and its interpretation be defined as in Example 6.2.4. On this interpretation, $H\,e\,E$ is true according to the truth-conditions in Definition 6.2.6, because there is no object that is both in the set containing Frodo, Sam, Merry, and Pippin and the set containing Elrond, Galadriel, Arwen, and Legolas.

In the above examples, we started with a language, gave an interpretation, and then determined which sentences were true on that interpretation. Sometimes, however, we want to *construct* an interpretation that make a specific sentence or group of sentences true:

Example 6.2.8. Let $\mathcal{L}_4$ be a term language consisting of the four copulae and the non-logical basic terms D, E, and F. Let us try to construct an interpretation that makes the following sentences true:

1. $F\,o\,D$
2. $D\,i\,F$
3. $E\,a\,F$
4. $E\,e\,D$

We'll begin with the first two, since partial claims are easier to make true than universal claims. In order for $F\,o\,D$ to be true, there has to be something which is in $I(D)$ but not in $I(F)$. Let's call that thing Thadia:

$$\begin{aligned} I(D) &= \{\text{Thadia}\} \\ I(F) &= \{\} \end{aligned}$$

For $D\,i\,F$ to be true, there needs to be something that is in both $I(D)$ and $I(F)$. Let us call that thing Mathanida:

$$\begin{aligned} I(D) &= \{\text{Mathanida, Thadia}\} \\ I(F) &= \{\text{Mathanida}\} \end{aligned}$$

Now we must consider the universal claims. The first says that everything in $I(F)$ must be in $I(E)$:

$$\begin{aligned} I(D) &= \{\text{Mathanida, Thadia}\} \\ I(E) &= \{\text{Mathanida}\} \\ I(F) &= \{\text{Mathanida}\} \end{aligned}$$

The last claim says that nothing is in both $I(E)$ and $I(D)$. Unfortunately, we already have something that is in both interpretations: Mathanida. Thus, the interpretation that we have constructed makes the first three claims true, but not the fourth.

Question 6.2.9. Is it possible to construct an interpretation that makes all four claims in Example 6.2.8 true?

Not every set of sentences has an interpretation that makes all the sentences true.

Definition 6.2.10 (Consistency)**.** A set Γ of categorical propositions is *consistent* if there is an interpretation I of the basic terms that makes all of the sentences in the set true. The set is *inconsistent* otherwise.

The truth values of the four types of propositions are related to each other by the relations of contrariness, subcontrariness, contradictoriness, subalternation, and superalternation.

Definition 6.2.11. Two categorical propositions are *contradictory* if the truth of one implies the falsity of the other, and vice versa.

If φ is an arbitrary categorical proposition, then we write $\bar{\varphi}$ for the contradictory of φ.

Corollary 6.2.12. *For any basic terms X and Y, $X\,a\,Y$ and $X\,o\,Y$ are contradictories, and $X\,e\,Y$ and $X\,i\,Y$ are contradictories.*

Proof. We prove only the first of these, leaving the other as an exercise for the reader.

We need to show that if $X\,a\,Y$ is true, then $X\,o\,Y$ must be false, and vice versa. Let I be an interpretation that makes $X\,a\,Y$ true, that is, $I \vDash X\,a\,Y$. It follows from Definition 6.2.6 that $I(Y) \subseteq I(X)$. Since $I(Y) \subseteq I(X)$, it is not true that $I(Y) \nsubseteq I(X)$. So it is not true that $I \vDash X\,o\,Y$. So $I \nvDash X\,o\,Y$, that is $X\,o\,Y$ is false on this interpretation.

Now, conversely, suppose that I is an interpretation that makes $X\,o\,Y$ true, that is $I \vDash X\,o\,Y$. It follows from Definition 6.2.6 that $I(Y) \nsubseteq I(X)$. Since $I(Y) \nsubseteq I(X)$, it is not true that $I(Y) \subseteq I(X)$. So it is not true that $I \vDash X\,a\,Y$. So $I \nvDash X\,a\,Y$, that is $X\,a\,Y$ is false on this interpretation, which is what we needed to show. □

The proofs of the following corollaries are similar.

Definition 6.2.13. Two propositions are *contrary* if they cannot both be true at the same time.

Corollary 6.2.14. *For any basic terms X and Y, $X\,a\,Y$ and $X\,e\,Y$ are contraries.*

Definition 6.2.15. Two propositions are *subcontrary* if they cannot both be false at the same time.

Corollary 6.2.16. *For any basic terms X and Y, $X\,i\,Y$ and $X\,o\,Y$ are subcontraries.*

Definition 6.2.17. One proposition is a *subaltern* of another if the truth of the latter implies the truth of the former. The *superaltern* relation is the converse of the subaltern relation.

Corollary 6.2.18. *For any basic terms X and Y, $X\,i\,Y$ is the subaltern of $X\,a\,Y$ while $X\,a\,Y$ is the superaltern of $X\,i\,Y$, and $X\,o\,Y$ is the subaltern of $X\,e\,Y$ while $X\,e\,Y$ is the superaltern of $X\,o\,Y$.*

These relations between the four types of categorical propositions are represented in the Categorical Square of Opposition in Figure 6.2 (the superalternation relation has been omitted).

We are now in a position to give our first definition of what counts as a good argument. In the context of categorical syllogisms, the characteristic of "goodness" is truth-preservation:

Definition 6.2.19. A *good* syllogism is one in which it is impossible for the premises to be true and the conclusion false, no matter what categorematic terms are used.

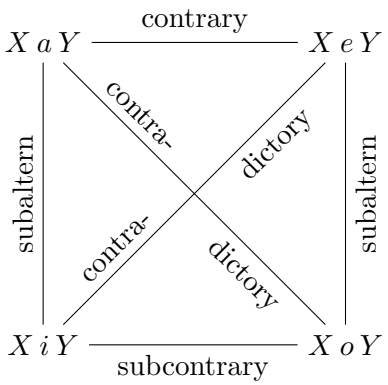


Figure 6.2: Categorical Square of Opposition

Such a syllogism is called *valid*. A syllogism which is not valid is called *invalid*. Whether a syllogism is valid or not depends not on the terms which go into the categorical propositions but on the form of the syllogism, that is, their figure and mood (cf. Definitions 6.1.11 and 6.1.12). This type of validity is called *formal* validity (related to the form of the argument) and is distinguished from *material* validity (related to the content, or the matter, of the argument). Of the possible 256 moods, 24 have been traditionally seen as valid. These 24 valid moods are listed in Figure 6.3.

Digression on existential import

In this chapter, we work with the assumption that the subject term of affirmative propositions are non-empty; that is, a implies i. From the other relations in the square, it follows then that e also implies o. This assumption is known as the assumption of "existential import", and it is required for the categorical propositions to have the truth conditions defined in Definition 6.2.6. However, this assumption is not an innocuous one—logicians often like to say that logic is ontologically neutral, in that it doesn't put any constraints on what can and cannot exist.[4]

The remainder of this subsection has not yet been written.

[4]However, we will see in §8.3 that there *is* one constraint that logics must impose upon the world, in order for any reasoning to get off the ground at all.

Ist figure	$P\,a\,M$	,	$M\,a\,S$	:	$P\,a\,S$	**Barbara**
	$P\,e\,M$	,	$M\,a\,S$	:	$P\,e\,S$	**Celarent**
	$P\,a\,M$	,	$M\,i\,S$	:	$P\,i\,S$	**Darii**
	$P\,e\,M$	,	$M\,i\,S$	:	$P\,o\,S$	**Ferio**
	$P\,a\,M$	,	$M\,a\,S$	:	$P\,i\,S$	**Barbari**
	$P\,e\,M$	,	$M\,a\,S$	:	$P\,o\,S$	**Celaront**
IInd figure	$M\,e\,P$	,	$M\,a\,S$	:	$P\,e\,S$	**Cesare**
	$M\,a\,P$	,	$M\,e\,S$	:	$P\,e\,S$	**Camestres**
	$M\,e\,P$	,	$M\,i\,S$	:	$P\,o\,S$	**Festino**
	$M\,a\,P$	,	$M\,o\,S$	:	$P\,o\,S$	**Baroco**
	$M\,e\,P$	,	$M\,a\,S$	:	$P\,o\,S$	**Cesaro**
	$M\,a\,P$	,	$M\,e\,S$	:	$P\,o\,S$	**Camestrop**
IIIrd figure	$P\,a\,M$	,	$S\,a\,M$	:	$P\,i\,S$	**Darapti**
	$P\,i\,M$	,	$S\,a\,M$	:	$P\,i\,S$	**Disamis**
	$P\,a\,M$	,	$S\,i\,M$	:	$P\,i\,S$	**Datisi**
	$P\,e\,M$	,	$S\,a\,M$	:	$P\,o\,S$	**Felapton**
	$P\,o\,M$	,	$S\,a\,M$	:	$P\,o\,S$	**Bocardo**
	$P\,e\,M$	,	$S\,i\,M$	:	$P\,o\,S$	**Ferison**
IVth figure	$M\,a\,P$	,	$S\,a\,M$	:	$P\,i\,S$	**Bramantip**
	$M\,a\,P$	,	$S\,e\,M$	:	$P\,e\,S$	**Camenes**
	$M\,i\,P$	,	$S\,a\,M$	:	$P\,i\,S$	**Dimaris**
	$M\,e\,P$	,	$S\,a\,M$	:	$P\,o\,S$	**Fesapo**
	$M\,e\,P$	,	$S\,i\,M$	:	$P\,o\,S$	**Fresison**
	$M\,a\,P$	,	$S\,e\,M$	:	$P o S$	**Camenop**

Figure 6.3: The 24 valid moods

6.3 Proof theory: Natural deduction

Our goal in this section is to develop a reliable method to determine of any given syllogism whether or not it is formally valid according to Definition 6.2.19. It is easy to demonstrate that some syllogism is not formally valid by producing an interpretation of the terms which makes the premises true but not the conclusion.

Example 6.3.1. The following syllogism is invalid:

$$\begin{array}{l} P\,e\,M \\ M\,e\,S \\ \hline \end{array}$$

$$P\,e\,S$$

We can show that this is invalid via an informal argument: Let I be an interpretation such that $I(P)$ is the set of all cats, $I(M)$ is the set of all dogs, and $I(S)$ is the set of things that purr. Then the major premise is true, because nothing is both a cat and a dog; the minor premise is true, because nothing both purrs and is a dog; but it is not true that nothing both purrs and is a cat, so the conclusion is false.

But as we noted earlier, we don't want to rely on our intuitions about the extensions of certain sets. A clearer demonstration of the invalidity of this syllogism is given in the following proof:

Proof. Let:

$$\begin{aligned} I(P) &:= \{b\} \\ I(M) &:= \{c\} \\ I(S) &:= \{b\} \end{aligned}$$

From this it is clear that $I \vDash PeM$ because $I(P) \cap I(M) = \varnothing$; $I \vDash Mes$ because $I(M) \cap I(S) = \varnothing$; but $I \nvDash P\,e\,S$, because $I(P) \cap I(S) = \{b\}$. □

But demonstrating that a syllogism is valid requires showing that there is no possible combination of basic terms that result in the premises being true and the conclusion false. Doing this would take an infinite amount of time, given that there are infinitely many ways in which the three terms of a syllogism can be interpreted. This is thus not an *effective* method. Instead, we introduce a method of proof that allows us to derive conclusions from pairs of premises. This method of proof will be such that we will never be able to derive false conclusions from true premises, and thus, if we can establish the existence of a proof of a conclusion from a specific pair of premises, we will be assured that the syllogism is valid.

The general method of proof is as follows:

1. Start with a list of obviously valid moods ("axioms"), then
2. Identify a list of appropriate transformation rules, then
3. Derive all valid moods from the perfect syllogisms by these transformations,
4. And find counterexamples for all other moods.

The proof method that we introduce here is based on the method of proof that Aristotle himself outlines in Book I, chapters i–ii, iv–vii of the *Prior Analtyics* [Aristotle,]. Here, Aristotle defines a notion of "perfect syllogism":

> I now call a syllogism perfect that needs nothing else beyond what is contained in it to become obvious [Aristotle, , I.i].[5]

Aristotle discusses first figure syllogisms in Book I chapter iv, and there he identifies the moods **Barbara**, **Celarent**, **Darii** and **Ferio** as *perfect*:

> It is also obvious that these all in themselves are perfect syllogisms.

These perfect moods we take as the axioms of our proof system; they are assumed as obviously correct and do not admit of any proof. We then introduce rules of proof which are *validity-preserving*, that is, if the perfect syllogisms are valid, then any syllogism which can be reduced to a perfect syllogism by means of these validity-preserving rules or transformations will also be valid. A proof system which is validity-preserving in this way is called *sound* (cf. §6.4).

Definition 6.3.2. A *syllogistic proof* is a finite list of numbered formulas such that every formula is annotated with one of our syllogistic proof rules or an axiom. If there is a proof of a categorical proposition φ from premises ψ_1 and ψ_2, we write $\psi_1, \psi_2 \vdash \varphi$.

We already have our axioms. We define how to apply these axioms in §6.3.3, and we also define two bookkeeping rules (§6.3.1), two types of transformations of categorical propositions, Simple Conversion (*conversio simpliciter*) and Accidental Conversion (*conversio per accidens*) (§6.3.2), and two methods of hypothetical proof, *Reductio Ad Absurdem* and Ecthesis (§6.3.4) Every proof will start from the assumption of the premises, and then proceed by applying the transformation rules to these premises, until the desired conclusion is obtained.

6.3.1 'Bookkeeping' rules

There are two bookkeeping rules, Premise and Reiteration.

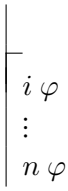
Rule 6.3.3 (Premise)**.** If ψ_1 is the major premise of a syllogism and ψ_2 the minor premise, then ψ_1 can be written on line 1 and ψ_2 line 2. The annotation for line n is Premise, and the resulting proof looks like this:

1 ψ_1	Premise
2 ψ_2	Premise
⊢	

[5] Τέλειον μὲν οὖν καλῶ συλλογισμὸν τὸν μηδενὸς ἄλλου προσδεόμενον παρὰ τὰ ἐιλημμένα πρὸς τὸ φανῆναι τὸ αναγκαῖον.

The vertical line drawn on the left is called a *scope line*. It indicates that every formula to the right of the vertical line depends on the assumption of the two premises. The assumptions themselves are set apart from the other propositions by the perpendicular line extending from the vertical line.

Rule 6.3.4 (Reiteration)**.** For any lines i and n, if φ occurs on line i of a proof, $i < n$, and n is within the scope of i, then φ can be written on line n. The annotation for line n is Reiteration, i.



$i\ \varphi$
$\vdots$
$n\ \varphi$ Reiteration

Reiteration is used to bring formulas within the scope of new assumptions. It can never be used to move formulas outside of the scope of the assumptions they depend on. It is also used to ensure that the premises of a syllogism are in the right order, so that the axioms can be applied (see §6.3.3).

6.3.2 Transformation rules

To introduce the transformation rules, it will be useful to be able to refer to the parts of a categorical proposition without specifying the specific terms or copulae. We introduce a generic way of representing these propositions that allows us to do so, by using the letters t_i for terms and the letter k for the copula, so that we can indicate the parts of a categorical proposition via this form:

$$t_1\, k\, t_2$$

Example 6.3.5. In the categorical proposition $M\ a\ S$, $t_1 = M$, $k = a$, and $t_2 = S$.

Both transformation rules come in two versions:

Rule 6.3.6 (Simple Conversion)**.**

1. For any lines i and n, if $t_1\ e\ t_2$ occurs on line i of a proof, $i < n$, and n is within the scope of i, then $t_2\ e\ t_1$ can be written on line n. The annotation for line n is

Simple Conversion, i.

$i\ t_1\ e\ t_2$
$\vdots$
$n\ t_2\ e\ t_1$ Simple Conversion, i

2. For any lines i and n, if $t_1\ i\ t_2$ occurs on line i of a proof, $i < n$, and n is within the scope of i, then $t_2\ i\ t_1$ can be written on line n. The annotation for line n is Simple Conversion, i.

$i\ t_1\ i\ t_2$
$\vdots$
$n\ t_2\ i\ t_1$ Simple Conversion, i

That is, if the copula is either e or i, you can swap the subject and predicate terms.

Rule 6.3.7 (Accidental Conversion)**.**

1. For any lines i and n, if $t_1\ a\ t_2$ occurs on line i of a proof, $i < n$, and n is within the scope of i, then $t_2\ i\ t_1$ can be written on line n. The annotation for line n is Accidental Conversion, i.

$i\ t_1\ a\ t_2$
$\vdots$
$n\ t_2\ i\ t_1$ Accidental Conversion, i

2. For any lines i and n, if $t_1\ e\ t_2$ occurs on line i of a proof, $i < n$ and n is within the scope of i, then $t_2\ o\ t_1$ can be written on line n. The annotation for line n is Accidental Conversion, i.

$i\ t_1\ e\ t_2$
$\vdots$
$n\ t_2\ o\ t_1$ Accidental Conversion, i

That is, if the copula is universal, i.e., a or e, you can always make the weaker, partial statement, i.e., i or o, so long as the subject and predicate terms are swapped.

Summary of transformation rules:

- a-claims can be converted only accidentally.
- e-claims can be converted both simply and accidentally.
- i-claims can be converted only simply.
- o-claims cannot be converted.

6.3.3 Applying axioms

The four perfect syllogisms are our axioms; they allow us to write down new propositions in a proof which are not transformations of the premises.

Rule 6.3.8 (Barbara)**.** For any lines $i, i+1 < n$, if $t_1\, a\, t_2$ occurs on line i and $t_2\, a\, t_3$ occurs on line $i+1$, then $t_1\, a\, t_3$ can be written on line n. The annotation for line n is Barbara$, i, i+1$.

$$
\begin{array}{|l}
\vdash \\
\quad i\ t_1\ a\ t_2 \\
\quad i+1\ t_2\ a\ t_3 \\
\quad \vdots \\
\quad n\ t_1\ a\ t_3 \qquad\qquad \mathsf{Barbara}, i, i+1
\end{array}
$$

Rule 6.3.9 (Celarent)**.** For any lines $i, i+1 < n$, if $t_1\, e\, t_2$ occurs on line i and $t_2\, a\, t_3$ occurs on line $i+1$, then $t_1\, e\, t_3$ can be written on line n. The annotation for line n is Celarent$, i, i+1$.

$$
\begin{array}{|l}
\vdash \\
\quad i\ t_1\ e\ t_2 \\
\quad i+1\ t_2\ a\ t_3 \\
\quad \vdots \\
\quad n\ t_1\ e\ t_3 \qquad\qquad \mathsf{Celarent}, i, i+1
\end{array}
$$

Rule 6.3.10 (Darii)**.** For any lines $i, i+1 < n$, if $t_1\, a\, t_2$ occurs on line i and $t_2\, i\, t_3$ occurs on line $i+1$, then $t_1\, i\, t_3$ can be written on line n. The annotation for line n is Darii$, i, i+1$.

$i\ t_1\ a\ t_2$	
$i+1\ t_2\ i\ t_3$	
$\vdots$	
$n\ t_1\ i\ t_3$	Darii$, i, i+1$

Rule 6.3.11 (Ferio)**.** For any lines $i, i+1 < n$, if $t_1\, e\, t_2$ occurs on line $i < n$ and $t_2\, i\, t_3$ occurs on line $i+1$, then $t_1\, o\, t_3$ can be written on line n. The annotation for line n is Ferio$, i, i+1$.

$i\ t_1\ e\ t_2$	
$i+1\ t_2\ i\ t_3$	
$\vdots$	
$n\ t_1\ o\ t_3$	Ferio$, i, i+1$

At this point, we are now able to prove all but two of the non-axiom valid syllogisms, using only the bookkeeping rules and the transformation rules. We give examples of two proofs:

Example 6.3.12. A proof of the mood Darapti (IIIrd figure):

$1\ P\, a\, M$	Premise
$2\ S\, a\, M$	Premise
$3\ P\, a\, M$	Reiteration, 1
$4\ M\, i\, S$	Accidental Conversion, 2
$5\ P\, i\, S$	Darii, 3, 4

Example 6.3.13. A proof of the mood Camestrop (IInd figure):

$1\ M\, a\, P$	Premise
$2\ M\, e\, S$	Premise
$3\ S\, e\, M$	Simple Conversion, 2
$4\ M\, a\, P$	Reiteration, 1
$5\ S\, e\, P$	Celarent, 3, 4
$6\ P\, o\, S$	Accidental Conversion, 5

6.3.4 Hypothetical methods

The two hypothetical methods introduced in this section allow us to make assumptions beyond what is warranted by the premises. We will indicate the subsidiary nature of these assumptions by introducing a new scope line for them.

In *Reductio Ad Absurdem*, the method of proof involves assuming the contradictory of the categorical proposition you wish to establish, and showing that that assumption allows you to prove some proposition and its contradictory. Because no proposition and its contradictory can both be true together, if our proof rules are sound we should not be able to prove both a proposition and its contradictory. Assuming that are rules *are* sound (we prove that they are in Theorem 6.4.1), this means that the original assumption cannot be true; if it cannot be true, it must be false, and if it is false, then its contradictory is true.

Rule 6.3.14 (RAA)**.** If from an assumption of φ on line i (with the annotation Assumption) of a proof it is possible to prove ψ at line j and the contradictory of ψ at line k, $i < j < n$ and $i < k < n$, then the contradictory of φ can be written on line n, with the scope line initiated at line i terminated at line $n-1$. The annotation for line n is RAA, i–k, and lines i–k are called a 'subproof'.

i	φ	Assumption
$\vdots$		
j	ψ	
$\vdots$		
k	$\bar{\psi}$	
n	$\bar{\varphi}$	RAA, i–k

(Recall that $\bar{\psi}$ is the contradictory of ψ.) That is, if from the assumption that a categorical proposition φ is true you can prove both some proposition and its contradictory, then you know that your initial assumption was wrong and that φ must be false. The assumption made at line i is said to be 'discharged' at line n. We stipulate that RAA can only be used in a given proof *once*.

This method of proof is needed only for two of the valid syllogisms in Figure 6.3: Baroco

and Bocardo. We give the proof of Bocardo here:

1 $P\,o\,M$	Premise
2 $S\,a\,M$	Premise
3 $P\,a\,S$	Assumption
4 $S\,a\,M$	Reiteration, 2
5 $P\,a\,M$	Barbara, 3, 4
6 $P\,o\,M$	Reiteration, 1
7 $P\,o\,S$	RAA, 3–6

The second hypothetical method is the method of *Ecthesis* or "setting out". For this method, we will introduce new notation that is not, strictly speaking, part of our syllogistic language, but which we allow within the context of proofs alone. Aristotle in the *Prior Analytics* says:

> The demonstration [of Darapti] can also be carried out through the impossible or by setting out [*ecthesis*]. For if both terms belong to all S, and one chooses one of the Ss, namely, s, then both P and R will belong to it, so that P will belong to some R (28a24–6).

The method of ecthesis involves two parts: The ecthetic assumption itself, namely that if some term belongs to all S, then we can pick an arbitrary object that is S, call it s, and reason about it. Thus, we must introduce notation that allows us to pick out individual objects, and not just sets of objects. In the context of the method of ecthesis, we will continue to use majuscule Roman letters to stand for basic terms; but we will now allow the use of miniscule Roman letters to stand for individual objects, in any place where we would otherwise have expected to find a basic term.

The second part of ecthesis is what has been called the principle of Expository Syllogism. This principle is founded on something that Aristotle says in the *Sophistical Refutations*, namely:

> Things that are the same as one and the same thing are the same as one another (168b32).

This is a backhanded way of saying that identical objects are identical, and hence will have identical properties. So let us assume that we can show that R belongs to particular object,

s, and that P also belongs to it. Then, since s is the same thing as itself, this means that there is some object (namely, s) which is both P and R. This outlines the positive case of the expository syllogism; we can make similar inferences involving negative claims. If there is something that is not both P and S, then this means that there is something that is S and not P (or P and not S); this, again, will be the same object in both cases. The expository syllogism is what allows us to make these deductions formally.

Before giving a general definition of the method of ecthesis, we will give an example where it is applied. This is an alternative proof of Darapti (cf. Example 6.3.12):

Example 6.3.15.

1 $P\ a\ M$	Premise
2 $S\ a\ M$	Premise
3 $M\ a\ s$	Assumption
4 $P\ a\ s$	Assumption
5 $S\ a\ M$	Reiteration, 2
6 $M\ a\ s$	Reiteration, 3
7 $S\ a\ s$	Barbara, 5, 6
8 $P\ a\ s$	Reiteration, 4
9 $S\ a\ s$	Reiteration, 7
10 $P\ i\ S$	Expository Syllogism, 8, 9

Definition 6.3.16 (Ecthesis)**.** To be added.

6.3.5 A few notes about natural deduction proofs

- A proof according to the rules set out above will always begin with two applications of the rule Premise, writing down the two premises.
- Pay attention to your scope lines!
- Always double check that you can in fact apply the rule you've applied. Simple conversion can only be performed on e and i claims. Accidental conversion can only be performed on a and e claims. (Note that there is thus no transformation rule for o claims.) The axioms cannot be applied unless the minor premise directly follows after the major premise.

- Every line must be annotated. If you do not have an annotation, you do not have a proof.

6.4 Metalogical results

In the Middle Ages, a system of mnemonics was developed to help the beginning logic student remember the valid syllogisms and their proofs. The earliest written record we have of this mnemonic is in William of Sherwood's *Introduction to Logic*, written in the middle of the 13th century [of Sherwood, 1995, p. 76], but the mnemonic almost certainly did not originate with him. Here is one version of the mnemonic (not William's).

> *Barbara, Celarent, Darii, Ferioque prioris,*
> *Césare, Cámestrés, Festíno, Baróco secúndae.*
> *Tértia Dáraptí, Disámis, Datísi, Felápton,*
> *Bocárdo, Feríson habét. Quárta ínsuper áddit*
> *Brámantíp, Camenés, Dimáris, Fesápo, Fresíson.*

(Note that not all 24 of the valid syllogisms in Figure 6.3 are included in this mnemonic.) Both James Welton and Francis Bowen, two 19th-century logicians, quote Augustus DeMorgan as describing this mnemonic as "the magic words which are more full of meaning than any that were ever made" [Bowen, 1895, p. 201], [Welton, 1891, p. 385]. The reason for this that from each name not only do we know that the syllogism is valid, but we can reconstruct both its form and its proof from one of the perfect syllogisms.

- The vowels correspond to the copulae of the major and minor premises and the conclusion (in that order).

- The first letter indicates to which one of the four perfect moods the mood is to be reduced: 'B' to Barbara, 'C' to Celarent, 'D' to Darii, and 'F' to Ferio.

- The letter 's' after a vowel indicates that the proposition with that vowel as a copula has to be simply converted.

- The letter 'p' after a vowel indicates either that the proposition with that vowel as a copula has to be accidentally converted ("*per accidens*"). or that accidental conversion will be used on another proposition in order to generate the proposition with that vowel as a copula.

- The letter 'c' after the first or second vowel indicates that the mood has to be proved indirectly by proving the contradictory of the corresponding premise in the course of using *reductio ad absurdum*.
- The letter 'm' indicates that the premises have to change order (be "moved"). This may be done via Reiteration or simply by applying the other rules.
- All other letters have only aesthetic purposes.

We noted earlier that there are 256 different syllogistic moods, of which only 24 are seen to be valid. If we can give proofs for the 20 moods which are not perfect, we have only completed half of our task—we still need to show that the remaining 232 moods are not provable. However, it would be very tedious to construct counterexamples for each of these. Instead, we will introduce a method that allows us to show that large classes of syllogisms are not valid. In the previous two sections we outlined a method of proof that allows us to prove results *within* a logical system. We can also prove things *about* the logical system; such results and proofs are the focus of this section. Because we are proving things *about* the logical system rather than *within* the logical system, our proofs will be written in English. Results *about* a system are called *meta-level* results (cf. §1.2), and the language we use to prove meta-level results is the *metalanguage*.

The first result we prove connects the proof rules to the semantics in §6.2, showing that any interpretation that makes the premises of a proof true will also make the conclusion of that proof true. This property—that the rules are truth-preserving—is called *soundness*.

Theorem 6.4.1. *The proof theory introduced in §6.3 is sound with respect to the semantics defined §6.2.*

Before we prove this, we first prove some simpler results.

Lemma 6.4.2 (Validity of the four basic moods)**.** *The four basic moods, Barbara, Celarent, Darii, and Ferio, are valid. That is, there is no interpretation which makes the premises true and the conclusion false.*

Proof. We prove each case:

Barbara Let us try to construct an interpretation I that makes the premises true and the conclusion, $P\,a\,S$, false. To make $P\,a\,S$ false, we must make its contradictory, $P\,o\,S$, true. That is, (1) there must be something in $I(S)$ that is not in $I(P)$; call it x:

$$\begin{aligned} I(P) &= \{\} \\ I(S) &= \{x\} \end{aligned}$$

By the minor premise, we know that (2) everything which is in $I(S)$ is also in $I(M)$:

$$\begin{aligned} I(P) &= \{\} \\ I(S) &= \{x\} \\ I(M) &= \{x\} \end{aligned}$$

By the major premise, we know that (3) everything that is in $I(M)$ is also in $I(P)$. But this means that x must be in $I(P)$, and this contradicts (1).

Celarent Let us try to construct an interpretation I that makes the premises true and the conclusion, $P\,e\,S$, false. To make $P\,e\,S$, we must make its contradictory, $P\,i\,S$, true. That is, (1) there must be something that is in both $I(P)$ and $I(S)$; call it x:

$$\begin{aligned} I(P) &= \{x\} \\ I(S) &= \{x\} \end{aligned}$$

From the minor premise, we know that (2) everything which is in $I(S)$ is in $I(M)$:

$$\begin{aligned} I(P) &= \{x\} \\ I(S) &= \{x\} \\ I(M) &= \{x\} \end{aligned}$$

But then there is something which is in both $I(P)$ and $I(M)$, which contradicts the major premise, which says that (3) there is nothing which is in both $I(P)$ and $I(M)$.

Darii Let us try to construct an interpretation I that makes the premises true and the conclusion false. To make the minor premise true, (1) there must be something which is in both $I(M)$ and $I(S)$; call it x:

$$\begin{aligned} I(M) &= \{x\} \\ I(S) &= \{x\} \end{aligned}$$

From the major premise, we know that (2) everything in $I(M)$ is also in $I(P)$:

$$\begin{aligned} I(M) &= \{x\} \\ I(S) &= \{x\} \\ I(P) &= \{x\} \end{aligned}$$

But in order to make the conclusion, $P\,i\,S$, false, there must be nothing which is in both $I(P)$ and $I(S)$, which is impossible.

Ferio Let us try to construct an interpretation I that makes the premises true and the conclusion false. To make the minor premise true, (1) there must be something which is in both $I(M)$ and $I(S)$; call it x:

$$\begin{aligned} I(M) &= \{x\} \\ I(S) &= \{x\} \end{aligned}$$

To make the conclusion, $P\,o\,S$, false, we must make its contradictory, $P\,a\,S$, true. That is, (2) everything that is in $I(S)$ must also be in $I(P)$:

$$\begin{aligned} I(M) &= \{x\} \\ I(S) &= \{x\} \\ I(P) &= \{x\} \end{aligned}$$

But this contradicts the major premise, which says that (3) there is nothing which is in both $I(P)$ and $I(M)$.

□

Lemma 6.4.3 (Soundness of the transformation rules)**.** *Simple Conversion and Accidental Conversion are truth-preserving.*

Proof.

Simple Conversion We prove that Simple Conversion of an e-claim is truth-preserving. That is, we show that any interpretation that makes $t_1\,e\,t_2$ true also makes $t_2\,e\,t_1$ true as well. Let I be an interpretation that makes $t_1\,e\,t_2$ true but is such that it makes $t_2\,e\,t_1$ false. For $t_2\,e\,t_1$ to be false, its contradictory, $t_2\,i\,t_1$, must be true. That is, I must be such that there is some object, call it x, where:

$$\begin{aligned} x &\in I(t_2) \\ x &\in I(t_1) \end{aligned}$$

But any such interpretation will make $t_1 e t_2$ false as well, since there is overlap between $I(t_1)$ and $I(t_2)$.

Accidental Conversion We prove that Accidental Conversion of an a-claim is truth-preserving. That is, we show that any interpretation that makes $t_1\,a\,t_2$ true also makes $t_2\,i\,t_1$ true as well. Let I be an interpretation that makes $t_1\,a\,t_2$ true. Then we know that there is at least one object, call it x, that is in $I(t_2)$, given our requirement (cf. Definition 6.2.3) that interpretations of the terms be non-empty:

$$x \in I(t_2)$$

For this interpretation to make $t_1\, a\, t_2$ true, any object in $I(t_2)$ must also be in $I(t_1)$:

$$\begin{array}{ll} \text{If} & x \in I(t_2) \\ \text{then} & x \in I(t_1) \end{array}$$

But then it follows that $I \vDash t_2\, i\, t_1$, since there is something that is in both $I(t_2)$ and $I(t_1)$, namely x, so $I(t_2) \cap I(t_1) \neq \varnothing$.

The cases of Simple Conversion of i-claims and Accidental Conversion of e-claims are left as exercises for the reader. □

Lemma 6.4.4. *RAA is truth-preserving.*

Proof. We want to show that any interpretation that makes the premises of a proof true will also make the result of an application of RAA true. Let I be an interpretation such that ψ_1 and ψ_2, the two premises are true. Now, let n be a line of the proof annotated with RAA, i–k, where $i \leq n$, $k \leq n$, and $i \neq k$. That is, the formula at line n is of the form $\bar{\varphi}$ where φ occurs on line i with the annotation Assumption, and at some j, k, where $i < j < n$ and $i < k < n$ there occur formulas ψ_3 and $\bar{\psi}_3$. Because the subproof from i to k does not contain any instance of RAA, we know that every line of it is annotated with either an axiom, a bookkeeping rule, or a transformation rule, all of which we've proven above are truth-preserving. Thus, if the assumption at line i is true, namely φ, then the interpretation also makes both ψ_3 and $\bar{\psi}_3$ true. But contradictories by definition cannot both be true at the same time. Therefore the assumption that the interpretation makes φ true is incorrect, and hence it must make φ false. But that is the same as making $\bar{\varphi}$ true, which is the formula on line n. □

Proof of Theorem 6.4.1. Immediate from Lemmas 6.4.2, 6.4.3, and 6.4.4 along with the observation that Reiteration is truth-preserving (if a sentence is true on an interpretation at line i, writting it down again at line n will not change its truth value), and that our premises are taken to be true by assumption. □

Theorem 6.4.5 (Aristotle)**.** *No conclusion can be proven from two negative premises. That is, if φ_1 and φ_2 are both negative, then for any ψ,*

$$\varphi_1, \varphi_2 \nvdash \psi$$

This gets rid of 64 invalid moods. The remaining invalid moods can be excluded by the following metatheoretic results:

Theorem 6.4.6. *No conclusion can be proven from two partial premises (Exercise 42).*

Theorem 6.4.7. *No affirmative conclusion can be proven from a negative premise.*

Theorem 6.4.8. *No negative conclusion can be proven from two affirmative premises.*

Theorem 6.4.9. *No universal conclusion can be proven from a partial premise.*

All of these theorems are proven in a similar fashion, based on how proofs can be constructed.

First, note that every proof *must* include an application of one of the axioms; this is because the conclusion of any syllogism must involve both the major and the minor term, but since every premise of a syllogism contains the middle term, neither will have both the major and minor (since every categorical proposition only has two terms, and every syllogism has three distinct terms).

Next, note that:

- None of Barbara, Celarent, Darii, or Ferio have two negative premises.
- None of Barbara, Celarent, Darii, or Ferio have two partial premises.
- Both Celarent and Ferio have a negative conclusion.
- Both Barbara and Darii have an affirmative conclusion.
- Both Darii and Ferio have a partial conclusion.

This means that any proof from a pair of premises that are (a) both negative or (b) both partial will contain more than just the application of an axiom: We must also use the other rules to convert the premises into a form that an axiom can be applied to.

There are four possible rules we could apply, Reiteration, Simple Conversion, Accidental Conversion, and RAA. Reiteration simply copies a previously written down proposition, without change. Thus if the premises are both negative or both partial, no amount of application of Reiteration will change that. Simple Conversion changes the order of the terms, but does not change the copulae; thus, if one of the premises has been simply converted, it will still be either negative or partial (though not both, since *o*-claims cannot be simply converted), no matter how many times it is simply converted. Accidental Conversion converts universal copulae into partial copulae, but does not change the quality of the proposition. Thus, applying Accidental Conversion to either of two negative claims will still result in two negative claims. Accidental Conversion cannot be applied to partial claims.

This leaves only the case of RAA. Whenever RAA is applied, one of the premises will be reiterated into the subproof to be used in the application of an axiom along with the

proposition assumed at the start of the subproof. The other premise will be reiterated into the subproof to be the contradictory of the proposition resulting from the application of the premise. If both premises are negative, then in order to apply an axiom in the subproof, the assumed proposition must be positive (for no axiom can be applied to two negative claims). But if one premise is negative and the other affirmative, then the only axioms that could be applied are Celarent and Ferio, both of which result in a negative conclusion. This conclusion would need to be contradicted by the other premise; but two negative propositions will never be contradictories. The case is similar if both premises are partial; if they are, then the assumed proposition must be universal in order for an axiom to applied. We leave the rest of the details to the reader to work out, and the other cases follow in a similar fashion.

6.5 Summary

In this chapter we have introduced a formal language for categorical syllogisms, semantic information about the relationships between the truth values of categorical propositions, and a proof theory which is sound and complete with respect to these semantics.

Chapter 7

Propositional logic [last modified 17 Mar 26]

7.1 Introduction

The language presented in the previous chapter is limited in its expressive power, that is, in what English-language sentences we are able to represent in the logical language. This is a feature in so far as it presented us with a simple finitistic system regarding which it was easy to introduce and prove basic meta-theoretic results. But it is a drawback in so far as some very straightforward argument types cannot be straightforwardly accommodated in the system. For example, neither of the two following arguments can be represented as categorical syllogisms:

> Every natural number is even or odd.
>
> Every even number is divisible by itself.
>
> Every odd number is divisible by itself.
>
> Therefore every natural number is divisible by itself.

This cannot be represented as a categorical syllogism because it has three premises, not two, and too many distinct terms.

> Every person is mortal.
>
> Socrates is a person.
>
> Therefore Socrates is mortal.

This cannot be represented as a categorical syllogism because neither the minor premise nor the conclusion are quantified.[1]

In this chapter, we introduce an alternative to the term-based logic of the syllogistic which has greater expressive power, but which is still finitistic in nature, the system of *propositional logic*. Propositional logic is concerned with four truth-functionals connectives: Negation, Conjunction, Disjunction, and Implication. We briefly discuss these connectives in English, and then motivate our study of them via looking at Stoic logic; for related material, see also the discussion of syncategorematic terms in §3.4.

7.1.1 Negation

Negations in English can be formed in a number of ways:

- not
- It is not the case that
- It is false that
- non-

7.1.2 Conjunctions

The most commonly used conjunction in English is 'and', but the following connectives can also function in the same way as conjunction:

- also
- but
- however
- yet
- still
- moreover
- although
- nevertheless

[1]Note that it is possible to extend the syllogistic to allow indefinite quantifiers, null quantifiers, and proper names.

7.1.3 Disjunctions

The most commonly used disjunction in English is 'or', which can either be taken inclusively (more than one of the disjuncts can be true and the entire statement still be true) or exclusively (exactly one of the disjuncts can be true for the entire statement to be true). We are interested in the inclusive version here (for the exclusive version, see [Fisk, 1965].). It can also be used to represent 'unless', given that every disjunction is equivalent to a conditional (which we'll show below).

7.1.4 Conditionals/implications

There are a variety of ways of expressing simple (i.e., non-counterfactual) conditionals in English.[2] The most common is with "If..., then...", but a variety of other expressions can be used:

- only if
- in case
- therefore
- provided that
- given that
- on condition that
- if

Conditionals are also used to indicate necessary and sufficient conditions. It is important to know that 'p if q' and 'p only if q' do not mean the same thing. In the first, q is a sufficient condition for p; the in second, q is a necessary condition for p. If the two are combined, 'p if and only if q', then q is both necessary and sufficient for p. 'if and only if' is commonly abbreviated to 'iff'.

Unless

A tricky English connective is 'unless'. Consider the sentence "I won't do the dishes unless you vacuum". This indicates that your vacuuming is a necessary condition for my doing the dishes; that is, if I do the dishes, then you vacuumed. Thus, '... unless...' incorporates not one but *two* logical connectives: negation and implication.

[2] We discuss counterfactuals in Chapter 17.

7.2 History

In this section, we discuss some of the historical developments of logical systems that take connectives like those listed above as basic. We begin with looking at the logic developed by the Stoics. The best introductory source for Stoic logic is [Mates, 1961]; all of our example sentences are taken from this source.

In contrast with Aristotelian syllogistics, which takes terms, quantifiers, and copulae as the basic building blocks of logic, the basic building block of Stoic logic is something called the 'λεκτον' (*lekton*, pl. *lekta*), "what the Barbarians do not understand when they hear Greek words spoken" [Mates, 1961, p. 11]. Examples of lekton include statements such as "It is day," "He is moving," "The earth flies," "You will marry a beautiful woman," etc. Each of these lekta is taken to be a single unit, not further analysable. Note that each Aristotelian categorical proposition is also a Stoic lekton: "All cats are animals" can be treated as a single, unanalysable unit, ignoring the copula, the terms, its quality, quantity, etc. We can introduce a notion of interpretation for these lekta. Because the internal structure of a lekton is not relevant to determining its truth, we can simply stipulate which lekta are true and which are false:

Definition 7.2.1 (Interpretation of lekta)**.** Given a set of lekta, an *interpretation* of the lekta is a truth-value assignment, that is, a list of which lekta in the set are true and which are false.

To every individual lekton there is a corresponding negation. For example, "It is not day", "He is not moving", "The earth does not fly", and "You will not marry a beautiful person" are all negations of the example lekta above. These negations can also be formed in a more explicit fashion, by prefixing the negation, e.g., "It is not the case that it is day" or even more simply, "Not: It is day". If a given lekton is true on some truth-value assignment, then its negation is false, and vice versa.

There are three important ways in which lekta and their negations can combined with each other:

Implication

Two lekta can be combined to form an implication. Examples of such implications include the following:

- If there are gods, then the universe is conducted according to divine foresight.
- If the earth is flying, then the earth exists.

- If the earth is flying, then it has wings.
- If he is moving, then he is walking.
- If it is day, it is light.
- If it is night, it is dark.
- If it is day, then I am conversing.
- If it is night, then I am conversing.
- If it is night, then it is day.

Disjunction

Two lekta can be combined to form a disjunctive. Two examples of disjunctions are given below; the first one serves as the first premise of a complex argument involving disjunctions, implications, and basic lekta.

- Either you will marry a beautiful woman or you will marry an ugly one.

  If she is beautiful, you will share her with others.

  If ugly, she will be a punishment.

  But neither of these things is desirable.

  Therefore, do not marry.

- Pleasure is either good or bad or neither good nor bad.

Conjunction

- Scipio was the son of Paulus and was twice consul and triumphed and was censor and was colleague in the censorship of L. Mummius.
- Scipio was the son of Paulus and was twice consul and triumphed and was censor and was colleague in the censorship of L. Mummius and he overcame Hannibal in Africa.

It is also possible to conjoin only two sentences together, e.g., "Scipio was the son of Paulus and was twice consul".

The two questions we want to answer are (a) when are each of these complex statements true? and (b) what forms of argument are going to be considered 'good'?

It is straightforward to determine the truth value of a conjunction: If every one of the conjuncts is true, as it is in the first example of a conjunction above, then the entire statement is true. If at least one of them is false (as is the case in the second example, since Scipio was in fact defeated by Hannibal), then the entire conjunction is false. We make this precise in §7.4.

For Stoics, disjunctions were exclusive: A disjunction is true if exactly one of the lekta it is composed of is true. This is in contrast with ordinary English 'or', which can (but doesn't have to be) read inclusively: "It is either red or colored" is true of an object which is both red and colored. (The system of propositional logic that we develop below will follow English, rather than Stoic, conventions, and adopt the inclusive 'or'.)

There was no unified Stoic position on the answer to (a) for conditionals (implications). Three important Stoic figures, Philo, Diodorus Cronus, and Chrysippus, each had differing views.

Philo A conditional is true if it is not the case that the "if" part is true and the "then" part is false.

Diodorus A conditional is true if it is never the case that the "if" part is true and the "then" part is false.

Chrysippus The view attributed to Chrysippus is articulated by Sextus Empiricus, and says that a conditional is true whenever the contradictory of the consequent conflicts with its antedecent (with "conflicts" being in need for further exposition). Chrysippus's view of conditionals is very similar to by Diogenes Laertius's view of the validity of arguments; he says in *Vitae* VII, 77 that an argument is valid when the negation of the conclusion is incompatible with the conjunction of its premises.

(Hájek notes that there was apparently a fourth Stoic conditional, but that it "is obscure, and seemingly of little historical interest" [Hájek, 2009, p. 206].)

If we add to these three ways of combining two lekta into one complex statement a means of negating complex sentences, then we can see that some of the combinations are redundant, in that they are equivalent to other ones. For example, on a Philonian account of the conditional, "If it is day, then it is not night" is equivalent to "Not both it is day and it is night". We will also show the interdefinability of the combinations in §7.4, and define a minimal set of connectives.

The standard of 'goodness' of argument for the Stoics was the same as for Aristotle, namely, necessary preservation of truth value.

Definition 7.2.2 (Validity of Stoic arguments)**.** A Stoic argument is *valid* iff there is no interpretation which makes the premises true and the conclusion false.

Like Aristotle, the Stoics took certain argument forms as basic or undemonstrated. To be precise, they identified five types of arguments as undemonstrated. Using the Stoic locutions for describing the argument forms, these five undemonstrated types are:

(1) If the first, then the second.
The first.
Therefore the second.

(2) If the first, then the second.
Not the second.
Therefore not the first.

(3) Not both the first and the second.
The first.
Therefore not the second.

(4) Either the first or the second.
The first.
Therefore not the second.

(5) Either the first or the second.
Not the first.
Therefore the second.

Note that the validity of (4) depends crucially on the fact that 'or' is read exclusively.

In each of these basic forms, 'the first' and 'the second' are indexicals, place-holders for any lekta whatsoever. This is the first step towards representing ordinary natural language arguments in a formal language. In the next section, we take this a step forward, and introduce a wholly symbolic language.

7.3 Language

In this section we define the language of propositional logic and define rules for the construction of well-formed formulas.

Definition 7.3.1. A *proposition* is that which is capable of having a truth value (in classical propositional logic, is either true or false).

Definition 7.3.2. A *propositional language* $\mathcal{L}_p$ consists in:

- An infinite set of atomic propositional letters p, q, r, ...
- Four logical connectives: $\neg$ (unary), $\vee, \wedge, \rightarrow$ (binary).
- Punctuation: (,).

The set of well-formed formulas (wffs) is defined recursively:

Definition 7.3.3 (Propositional wffs)**.**

- Every atomic proposition is an $\mathcal{L}_p$-wff.
- If φ is a $\mathcal{L}_p$-wff, then so is $\neg\varphi$.
- If φ and ψ are $\mathcal{L}_p$-wffs, then so are $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, and $(\varphi \rightarrow \psi)$.

When no ambiguity will result, we will omit the outermost set of parentheses.

Definition 7.3.4 (Subformulas)**.**

- φ is a *subformula* of $\neg\varphi$, and $\neg\varphi$ is the *negation* of φ.
- φ and ψ are *subformulas* of $\varphi \wedge \psi$. $\varphi \wedge \psi$ is called a *conjunction*, and φ and ψ are both *conjuncts*.
- φ and ψ are *subformulas* of $\varphi \vee \psi$. $\varphi \vee \psi$ is called a *disjunction*, and φ and ψ are both *disjuncts*.
- φ and ψ are *subformulas* of $\varphi \rightarrow \psi$. $\varphi \rightarrow \psi$ is called a *conditional*, a *hypothetical*, or an *implication* and φ is the *antecedent* and ψ the *consequent*.

If φ is a subformula of ψ, and ψ is a subformula of χ, then φ is also a subformula of χ.

Definition 7.3.5. The *main connective* of a wff is the connective that is in the least number of parentheses, that is, the connective which has been applied last in the construction of the formula.

Note 7.3.6. An atomic wff has no main connective.

To every wff there corresponds a parse tree which indicates how the sentence is constructed.

Definition 7.3.7 (Propositional parse trees)**.** Parse trees are composed of *roots*, *branches*, and *leaves*. The wff to be parsed is the *root*, and the atomic formulas occurring in the wff are the *leaves*. The wffs that occur on the branches are called 'nodes' (these include the root and the leaves). All leaves are atoms (and all atoms are leaves), and the branches themselves are generated in a recursive fashion starting from the root:

1. If the root node is an atom, then the parse tree consists of the leaf/root alone, and nothing more occurs.

2. If the root node is not a leaf (i.e., not an atom), then it is of the form $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, or $\varphi \to \psi$.

   a) If the node is of the form $\neg\varphi$, then there is one branch going out from the node, and the node at the end of the branch is labeled φ.

   b) If the node is of the form $\varphi \wedge \psi$, $\varphi \vee \psi$, or $\varphi \to \psi$, then there are two branches going out from the node. The node at the end of the left branch is labeled φ and the node at the end of the right branch is labeled ψ.

3. We now have a new set of nodes at the top (or bottom, depending on what direction your branches grow) of our tree. These new nodes can each be treated as the "root" of a new tree. Each of these nodes is either a leaf or not. If it is a leaf, go to (1). If it is not a leaf, then go to (2). the tree is extended according to steps (1) and (2).

The construction terminates when every node is a leaf.

An example parse tree, for the formula $\neg p \vee (q \to (r \to (p \wedge q)))$, is given in Figure 7.1. Further examples are found in Figures 7.2 and 7.3.

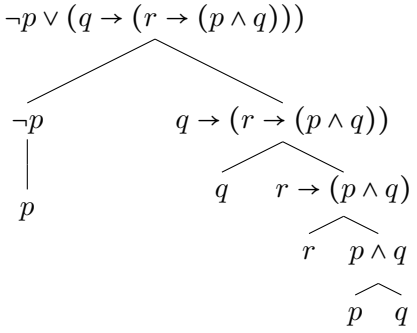


Figure 7.1: An example parse tree.

Lemma 7.3.8 (Unique readability)**.** *For every formula φ in $\mathcal{L}_p$, exactly one of the following holds:*

1. *φ is atomic.*

2. *There is a unique ψ such that φ is $\neg\psi$. In this case, φ is called the 'negation of ψ'.*

3. *There are unique ψ, χ such that $\varphi = \psi \wedge \chi$. In this case, φ is called a 'conjunction', and ψ and χ are both called 'conjuncts'.*

4. *There are unique ψ, χ such that $\varphi = \psi \vee \chi$. In this case, φ is called a 'disjunction', and ψ and χ are both called 'disjuncts'.*

5. *There are unique ψ, χ such that $\varphi = \psi \rightarrow \chi$. In this case, φ is called a 'conditional' or an 'implication', ψ is called the 'antecedent', and χ is called the 'consequent'.*

Proof. By inspection of the parse trees and Definition 7.3.3. If φ is atomic, the fact is clear. If it is not atomic, then suppose that there are distinct ψ, χ such that $\varphi = \neg\psi = \neg\chi$. But this is impossible if $\psi \neq \chi$, since adding $\neg$ to the front does not change either ψ or χ. The case for when there are distinct ψ,χ, and ψ',χ' such that $\varphi = \psi@\chi = \psi'@\chi'$ (for @ = one of $\vee, \wedge \rightarrow$) is similar. □

Essentially, the only way to get identical strings is to put identical things together in an identical way. Note that the scope of a negation will always be as narrow as syntactically possible; e.g., $\neg\varphi \wedge \psi$ is to be read as the conjunction of one negated conjunct, $\neg\varphi$, and one non-negated conjunct, ψ. To force the negation to apply to the entire conjunction, and not just a single conjunct, parentheses are necessary, e.g., $\neg(\varphi \wedge \psi)$.

7.4 Semantics

Definition 7.4.1. A *truth-functional proposition-forming function* is any function which

- takes propositions as input and returns propositions as output, and

- is such that the truth-value of the output-proposition is wholly determined by the truth-value of the input-propositions.

These truth-functional proposition-forming functions are often called 'connectives' or 'operators'.

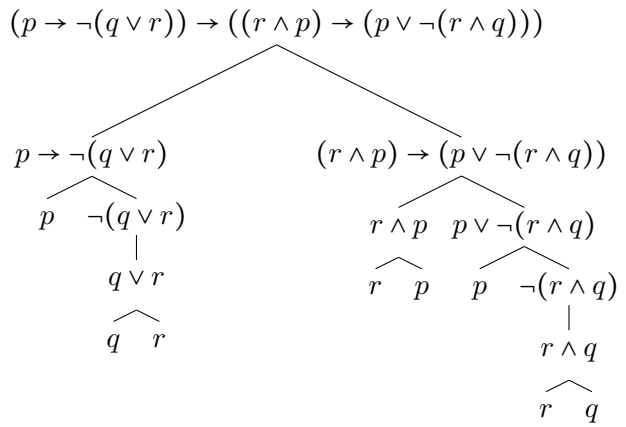


Figure 7.2: Another example parse tree.

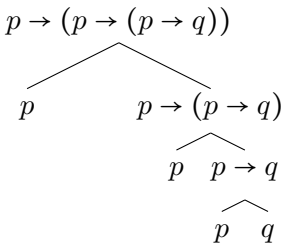


Figure 7.3: Another example parse tree.

7.4.1 Truth tables

The classical semantics for the propositional logical connectives are truth-functional.

Definition 7.4.2. An *interpretation* is an assignment of truth vlaues to atomic propositions.

Definition 7.4.3. A *classical interpretation* is an interpretation which respects the following two principles (cf. Chapter 5):

1. **Law of Excluded Middle**: Every proposition is assigned either T (true) or F (false).
2. **Principle of Non-Contradiction**: No proposition is assigned both T and F.

In this chapter, we will *only* use classical interpretations. For information on non-classical propositional logics, see Chapter 21.

An interpretation is also called a 'truth-value assignment' (for obvious reasons). The assignment of truth values to the atomic propositions is entirely arbitrary: There are no

restrictions (other than LEM and PNC, in the case of classical interpretations), as to how the truth values can be assigned to the atoms. The assignment of truth values to the complex (non-atomic) formulas is, in contrast, governed by the following four truth tables:

Definition 7.4.4 (Negation)**.**

φ	$\neg\varphi$
T	F
F	T

The table for negation states that φ and $\neg\varphi$ will always have the opposite values.

Definition 7.4.5 (Conjunction)**.**

φ	ψ	$\varphi \wedge \psi$
T	T	T
T	F	F
F	T	F
F	F	F

The only time a conjunction is true is when both conjuncts are true; in any other situation, the conjunction is false.

Definition 7.4.6 (Disjunction)**.**

φ	ψ	$\varphi \vee \psi$
T	T	T
T	F	T
F	T	T
F	F	F

The only time a disjunction is false is when both conjuncts are false; in any other situation, the disjunction is true. This is an *inclusive* disjunction, where the possibility of having *both* conjuncts true at the same time is not excluded. This is in contrast with the Stoic account of disjunction (cf. p. 111).

Definition 7.4.7 (Conditional)**.**

φ	ψ	$\varphi \to \psi$
T	T	T
T	F	F
F	T	T
F	F	T

The only time a conditional is false is when the antecedent is true and the consequent is false. This account of the conditional follows Philo (as opposed to Diodorus), and is an account of the *material* conditional. The material conditional is one of the weakest types of conditional found in the English language, and there are many cases (e.g., counterfactuals, scientific and mathematical reasoning, etc.) where the account given by the material conditional will be inadequate for the analysis of natural language. In particular, many people object to the material conditional as an adequate analysis of English-language conditionals because they give rise to the so-called "paradoxes of material implication." These paradoxes are discussed in §20.1, and alternative, non-truthfunctional, accounts of the conditional are provided there.

These truth tables can be combined to generate the truth values of wffs of arbitrary complexity: Once we know the truth values of the subformulas, we can generate the truth values of the more complex formulas. For instance, to calculate the truth-table of the wff $(p \to q) \to ((\neg r \vee q) \vee \neg p)$, we must first consider all possible combination of truth values for the three atoms, p, q, and r. Because there are three atoms and each has two possible truth values, the complete table for $(p \to q) \to ((\neg r \vee q) \vee \neg p))$ will involve $2^3 = 8$ rows (and one column for every unique subformula), as shown in Figure 7.4.

p	q	r
T	T	T
T	T	F
T	F	T
T	F	F
F	T	T
F	T	F
F	F	T
F	F	F

Figure 7.4: All possible truth value combinations for three atoms.

Next, we calculate the values of the subformulas involving atoms, as shown in Figure 7.5.

p	q	r	$\neg p$	$\neg r$	$p \rightarrow q$
T	T	T	F	F	T
T	T	F	F	T	T
T	F	T	F	F	F
T	F	F	F	T	F
F	T	T	T	F	T
F	T	F	T	T	T
F	F	T	T	F	T
F	F	F	T	T	T

Figure 7.5: Combining atomic truth values into complex ones.

In order to generate the values of the next most complex wffs, we can treat the three right-most columns of Figure 7.5 as if they were atomic formulas. This is demonstrated in Figures 7.6, 7.7, and 7.8.

q	$\neg r$	$\neg r \vee q$
T	F	T
T	T	T
F	F	F
F	T	T
T	F	T
T	T	T
F	F	F
F	T	T

Figure 7.6: Calculating truth values of complex wffs (1).

$\neg p$	$\neg r \vee q$	$(\neg r \vee q) \vee \neg p$
F	T	T
F	T	T
F	F	F
F	T	T
T	T	T
T	T	T
T	F	T
T	T	T

Figure 7.7: Calculating truth values of complex wffs (2).

$p \to q$	$(\neg r \vee q) \vee \neg p$	$(p \to q) \to ((\neg r \vee q) \vee \neg p)$
T	T	T
T	T	T
F	F	T
F	T	T
T	T	T
T	T	T
T	T	T
T	T	T

Figure 7.8: Calculating truth values of complex wffs (3).

The subformulas need not all be represented in unique columns in the truth table; it is possible to simply calculate the values under each connective or atom in the complex formula. This, less verbose, method is illustrated by the truth table which is given in Figure 7.9, for the formula in Figure 7.1. The column for the main connective highlighted in red.

p	q	r	$($	$\neg$	p	$\vee$	$($	q	$\to$	$($	r	$\to$	$($	p	$\wedge$	q	$)))$
T	T	T		F	T	T		T	T		T	T		T	T	T	
T	T	F		F	T	T		T	T		F	T		T	T	T	
T	F	T		F	T	T		F	T		T	F		T	F	F	
T	F	F		F	T	T		F	T		F	T		T	F	F	
F	T	T		T	F	T		T	F		T	F		F	F	T	
F	T	F		T	F	T		T	T		F	T		F	F	T	
F	F	T		T	F	T		F	T		T	F		F	F	F	
F	F	F		T	F	T		F	T		F	T		F	F	F	

Figure 7.9: The truth table for $\neg p \vee (q \to (r \to (p \wedge q)))$

It is also possible to use the parse trees of Definition 7.3.7 to determine how to combine the truth tables. Every unique node in the tree will become a column in the truth table. We start with the leaves, and assign truth values to them such that for every possible combination of truth values amongst the leaves, there is at least one row of the truth table that corresponds to it. We then generate the truth values of each node from simplest to most complex.

We highlight two important classes of propositional wffs, the valid and the contradictory (or inconsistent).

Definition 7.4.8. An $\mathcal{L}_p$-wff φ is *valid* iff it is true on every assignment of truth values to the atomic propositions. If φ is valid, we write $\vDash \varphi$.[3]

That is, the column under the main connective will contain only T's. Valid $\mathcal{L}_p$-wffs are also called *tautologies*. The wff in Figure 7.1 is a tautology, as Table 7.9 demonstrates. A list of commonly used propositional tautologies is given in Table 7.10.

Definition 7.4.9. An $\mathcal{L}_p$-wff is *inconsistent* iff it is false on every assignment of truth values to the atomic propositions.

That is, the column under the main connective will contain only F's. Inconsistent $\mathcal{L}_p$-wffs are also called *contradictions*.

1	$(p \wedge q) \to p$	
2	$(p \wedge q) \to q$	
3	$(p \to q) \to ((p \to r) \to (p \to (q \wedge r)))$	Composition
4	$p \to (q \to (p \wedge q))$	Adjunction
5	$(p \to q) \to ((q \to p) \to (p \leftrightarrow q))$	
6	$(p \to q) \to ((q \to r) \to (p \to r))$	Transitivity/Syllogism
7	$(p \to (q \to r)) \to ((p \wedge q) \to r)$	Importation
8	$(p \to q) \to ((q \to (r \to s)) \to ((p \wedge r) \to s))$	
9	$p \to (p \vee q)$	
10	$q \to (p \vee q)$	
11	$(p \to q) \to ((r \to q) \to ((p \vee r) \to q))$	
12	$p \leftrightarrow \neg\neg p$	Double Negation (Lemma 7.4.35)
13	$(p \vee q) \leftrightarrow \neg(\neg p \wedge \neg q)$	DeMorgan's Law
14	$(p \wedge q) \leftrightarrow \neg(\neg p \vee \neg q)$	DeMorgan's Law
15	$(p \to q) \to (\neg q \to \neg p)$	Contraposition/Transposition
16	$(p \vee q) \leftrightarrow (q \vee q)$	Commutativity (Lemma 7.4.32)
17	$(p \wedge q) \leftrightarrow (q \wedge p)$	Commutativity (Lemma 7.4.31)
18	$((p \vee q) \vee r) \leftrightarrow (p \vee (q \vee r))$	Associativity (Lemma 7.4.21)
19	$((p \wedge q) \wedge r) \leftrightarrow (p \wedge (q \wedge r))$	Associativity (Lemma 7.4.20)
20	$p \leftrightarrow (p \vee p)$	
21	$p \leftrightarrow (p \wedge p)$	

Table 7.10: Some common propositional tautologies and their names

Sometimes it is useful to refer to an arbitrary tautology or arbitrary contradiction without specifying which one. To do so, we will use the symbols $\top$ ('top') and $\bot$ ('bottom'),

[3] '$\vDash$' is not a symbol *in* our language, but a symbol that we use to say something *about* our language.

respectively. While strictly speaking these symbols are not a part of our language, we can introduce them as short-hand notation. With them in hand, we can capture the fact that tautologies and contradictions are set apart from other wffs in that their truth-values are always *constant*: Their value does not depend on the assignment of values to the atomic propositions.

Definition 7.4.10 (Constant wff)**.** A *constant* wff is any wff that is constructed from $\top$, $\bot$, and the truth-functional operators.

Corollary 7.4.11. *Every constant wff is either valid or inconsistent.*

Definition 7.4.12. A sentence is *consistent* if there is at least one assignment of truth values to the atomic propositions that makes it true.

Corollary 7.4.13. *(1) Every tautology is consistent.*

(2) If a sentence is (merely) consistent, then there is at least one truth-value assignment that makes it false and at least one that makes it true.

Above we have been talking about wffs only, whereas if we take a step back and look at the broader picture, what we are really interested in is arguments. Recall our definition of argument from Chapter 1 (Def. 1.1.2), in which an argument consists in a set of premises, Γ, and another sentence, φ, the conclusion.

Our definition of a "good argument" in propositional logic is the same as in the syllogistics: An argument is *good* if it is valid. The only difference is that here, we define validity in terms of truth value assignments instead of interpretations:

Definition 7.4.14 (Validity)**.** An $\mathcal{L}_p$-argument is *valid* iff every assignment of truth values that makes the premises true also makes the conclusion true. We write $\Gamma \vDash \varphi$ for a valid argument whose premises are members of Γ and whose conclusion is φ.

It then follows that an argument is invalid ($\Gamma \nvDash \varphi$) if there is at least one truth value assignment that makes all members of Γ true and φ false.

Validity is the central focus of propositional logic and its extensions, but the astute reader may complain that there is a way of making such a good argument *even better*: For nothing in the definition of validity requires that the premises actually *be* true. If the premises are true, then we say that the argument is sound:

Definition 7.4.15 (Soundness)**.** An $\mathcal{L}_p$-argument is *sound* if it is valid and the premises are in fact true.

There is no notation that corresponds to the notion of a sound argument, in the way that we introduced the double-turnstile for a valid argument. This is in keeping with the fact that logicians don't actually care about whether or not certain sentences are true, they simply care about the relationship between the truth of some sentences and the truths of others.

There is an important connection between the notion of validity used for arguments and the notion of an $\mathcal{L}_p$-valid wff:

Fact 7.4.16. An $\mathcal{L}_p$-argument is valid iff the wff constructed by taking the conjunction of the premises as the antecedent of a conditional and the conclusion as the consequent of the conditional is an $\mathcal{L}_p$-valid wff.

Example 7.4.17. Consider the following complex (i.e., non-basic) Stoic argument:

> If the first and the second, then the third.
> Not the third.
> The first.
> Therefore not the second.

The individual sentences can be represented:

$(p \wedge q) \rightarrow r$
$\neg r$
p
$\neg q$

This argument is valid iff the wff $(((p \wedge q) \rightarrow r) \wedge (\neg r \wedge p)) \rightarrow \neg q$ is valid.

Certain syntactically distinct wffs turn out to be equivalent to each other, in the sense that they will always generate the same truth tables:

Definition 7.4.18 (Semantic equivalence)**.** Two wffs φ and ψ are *semantically equivalent* if they have the same truth value on every truth-value assignment.

We sometimes say that semantically equivalent sentences are "logically equivalent." If two sentences are semantically equivalent, we write $\varphi \leftrightarrow \psi$. This is new notation that we introduce and define to be the same as $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$.

Corollary 7.4.19. *If two sentences φ and ψ are semantically equivalent, then $\varphi \leftrightarrow \psi$ will be a tautology.*

We can exploit certain logical equivalences to simplify notation by allowing us to drop parentheses.

Lemma 7.4.20 (Associativity of $\wedge$)**.** $(\varphi_1 \wedge \varphi_2) \wedge \varphi_3$ *and* $\varphi_1 \wedge (\varphi_2 \wedge \varphi_3)$ *are semantically equivalent.*

Proof. By inspection of the truth tables. □

Lemma 7.4.21 (Associativity of $\vee$)**.** $(\varphi_1 \vee \varphi_2) \vee \varphi_3$ *and* $\varphi_1 \vee (\varphi_2 \vee \varphi_3)$ *are semantically equivalent.*

Proof. By inspection of the truth tables. □

As a consequence of Lemmas 7.4.20 and 7.4.21, we will henceforth simply write $(\varphi_1 \wedge \varphi_2 \wedge \varphi_3)$ rather than either $(\varphi_1 \wedge \varphi_2) \wedge \varphi_3$ or $\varphi_1 \wedge (\varphi_2 \wedge \varphi_3)$, and similarly for disjunctions.

With the notion of semantic equivalence between wffs, we can also define the notion of a *normal form* of a wff. First, note that for any formula φ, φ and $\varphi \wedge \varphi$ are equivalent. Thus we can consider φ itself to be a conjunction, albeit a degenerate one containing only one conjunct. Similarly, note that for any formula φ, φ and $\varphi \vee \varphi$ are equivalent. Thus we can consider φ itself to be a disjunction, albeit a degenerate one containing only one disjunct.

Definition 7.4.22 (Conjunctive Normal Form)**.** A wff φ is said to be in *conjunctive normal form* if it is a conjunction (possibly degenerate) in which each conjunct is a disjunction (possibly degenerate) of atomic propositions and their negations.

Definition 7.4.23 (Disjunctive Normal Form)**.** A wff φ is said to be in *disjunctive normal form* if it is a disjunction (possibly degenerate) in which each disjunct is a conjunction (possibly degenerate) of atomic propositions and their negations.

Example 7.4.24. The following formulas are all in CNF:

$$p$$
$$p \vee \neg q$$
$$(p \vee r) \wedge q$$
$$(p \vee q \vee \neg r) \wedge r \wedge (\neg p \vee r)$$

Example 7.4.25. The following formulas are all in DNF:

$$p \vee q$$
$$p \vee \neg q$$
$$(p \wedge \neg p) \vee q \vee r$$
$$(p \wedge p) \vee (r \wedge \neg s) \vee (p \wedge q)$$

Corollary 7.4.26. *Some formulas are in both CNF and DNF at the same time.*

The main result concerning CNF is the following:

Theorem 7.4.27. *Every wff φ is equivalent to a wff φ' where φ' is in CNF.*

Corollary 7.4.28. *Every formula φ is equivalent to a formula φ' that uses only the connectives $\neg$ and $\wedge$.*

To prove Theorem 7.4.27, we first prove two distributivity laws.

Lemma 7.4.29 (Distribution of $\wedge$ over $\vee$)**.** $\varphi_1 \wedge (\varphi_2 \vee \varphi_3)$ *and* $(\varphi_1 \wedge \varphi_2) \vee (\varphi_1 \wedge \varphi_3)$ *are semantically equivalent.*

Proof. By inspection of the truth tables:

φ_1	φ_2	φ_3	(φ_1	$\wedge$	(φ_2	$\vee$	φ_3)	(φ_1	$\wedge$	φ_2)	$\vee$	(φ_1	$\wedge$	φ_3)
T	T	T	T	T	T	T	T	T	T	T	T	T	T	T
T	T	F	T	T	T	T	F	T	T	T	T	T	F	F
T	F	T	T	T	F	T	T	T	F	F	T	T	T	T
T	F	F	T	F	F	F	F	T	F	F	F	T	F	F
F	T	T	F	F	T	T	T	F	F	T	F	F	F	T
F	T	F	F	F	T	T	F	F	F	T	F	F	F	F
F	F	T	F	F	F	T	T	F	F	F	F	F	F	T
F	F	F	F	F	F	F	F	F	F	F	F	F	F	F

□

Lemma 7.4.30 (Distribution of $\vee$ over $\wedge$)**.** $\varphi_1 \vee (\varphi_2 \wedge \varphi_3)$ *and* $(\varphi_1 \vee \varphi_2) \wedge (\varphi_1 \vee \varphi_3)$ *are semantically equivalent.*

Proof. See Exercise . □

We make use of two laws of commutativity:

Lemma 7.4.31 (Commutativity of $\wedge$)**.** $p \wedge q$ *and* $q \wedge p$ *are semantically equivalent.*

Lemma 7.4.32 (Commutativity of $\vee$)**.** $p \vee q$ *and* $q \vee p$ *are semantically equivalent.*

And two forms of the so-called DeMorgan's laws:

Lemma 7.4.33. $\neg(p \wedge q)$ *and* $\neg p \vee \neg q$ *are semantically equivalent.*

Lemma 7.4.34. $\neg(p \vee q)$ *and* $\neg p \wedge \neg q$ *are semantically equivalent.*

And the redundancy of double negations:

Lemma 7.4.35 (Law of Double Negation)**.** *p is equivalent to $\neg\neg p$.*

Proof. These five lemmas are proved by inspection of the truth tables. □

With the addition of the equivalences given in Table 7.11, we are now in a position to prove Theorem 7.4.27:

Proof of Theorem 7.4.27. We give an algorithm that converts any wff φ which is not in CNF into a wff φ' which is:

1. Eliminate all occurrences of $\rightarrow$ by applying the equivalence illustrated in Table 7.11.
2. Apply DeMorgan's laws and the law of double negation until all negations are attached to atoms.
3. Apply the distribution and commutativity laws to distribute disjunctions over conjunctions.

□

Example 7.4.36. We convert $\neg(p \vee (\neg q \wedge r)) \rightarrow \neg((\neg p \rightarrow q) \wedge (\neg r \wedge s))$ into CNF.

1. Eliminate conditionals: $\neg\neg(p \vee (\neg q \wedge r)) \vee \neg((\neg\neg p \vee q) \wedge (\neg r \wedge s))$.
2. Eliminate double negations: $(p \vee (\neg q \wedge r)) \vee \neg((p \vee q) \wedge (\neg r \wedge s))$.
3. Apply DeMorgan's: $(p \vee (\neg q \wedge r)) \vee (\neg(p \vee q) \vee \neg(\neg r \wedge s))$.
4. Apply DeMorgan's again: $(p \vee (\neg q \wedge r)) \vee (\neg(p \vee q) \vee (\neg\neg r \vee \neg s))$.
5. Eliminate double negations: $(p \vee (\neg q \wedge r)) \vee (\neg(p \vee q) \vee (r \vee \neg s))$.
6. Apply DeMorgan's again: $(p \vee (\neg q \wedge r)) \vee (\neg p \wedge \neg q) \vee (r \vee \neg s))$.
7. Distribute $\vee$ over $\wedge$: $((p \vee \neg q) \wedge (p \vee r)) \vee (\neg p \wedge \neg q) \vee (r \vee \neg s))$.
8. Associativity: $((p \vee \neg q) \wedge (p \vee r)) \vee (((\neg p \wedge \neg q) \vee r) \vee \neg s)$.
9. Distribute $\vee$ over $\wedge$: $((p \vee \neg q) \wedge (p \vee r)) \vee (((\neg p \vee r) \wedge (r \vee \neg q)) \vee \neg s)$.
10. Distribute $\vee$ over $\wedge$: $((p \vee \neg q) \wedge (p \vee r)) \vee ((\neg p \vee r \vee \neg s) \wedge (r \vee \neg q \vee \neg s))$.

11. Distribute $\vee$ over $\wedge$:

$$((p \vee \neg q) \vee ((\neg p \vee r \vee \neg s) \wedge (r \vee \neg q \vee \neg s))) \wedge ((p \vee r) \vee ((\neg p \vee r \vee \neg s) \wedge (r \vee \neg q \vee \neg s)))$$

12. Distribute $\vee$ over $\wedge$:

$$(\neg p \vee r \vee \neg s \vee p \vee \neg q) \wedge (r \vee \neg q \vee \neg s \vee p \vee \neg q) \wedge ((p \vee r) \vee ((\neg p \vee r \vee \neg s) \wedge (r \vee \neg q \vee \neg s)))$$

13. Distribute $\vee$ over $\wedge$:

$$(\neg p \vee r \vee \neg s \vee p \vee \neg q) \wedge (r \vee \neg q \vee \neg s \vee p \vee \neg q) \wedge (p \vee r \vee \neg p \vee r \vee \neg s) \wedge (p \vee r \vee r \vee \neg q \vee \neg s)$$

The wff:

$$(\neg p \vee r \vee \neg s \vee p \vee \neg q) \wedge (r \vee \neg q \vee \neg s \vee p \vee \neg q) \wedge (p \vee r \vee \neg p \vee r \vee \neg s) \wedge (p \vee r \vee r \vee \neg q \vee \neg s) \quad (7.1)$$

is in CNF.

Note that the resulting wff contains duplicative disjuncts. We can simplify the wff to the following by removing such disjuncts, e.g.:

$$(\neg p \vee r \vee \neg s \vee p \vee \neg q) \wedge (r \vee \neg q \vee \neg s \vee p) \wedge (p \vee r \vee \neg p \vee \neg s) \wedge (p \vee r \vee \neg q \vee \neg s) \quad (7.2)$$

Both (7.1) and (7.2) are in CNF, showing that there is not necessarily a unique wff in CNF that a given wff will reduce to.

Theorem 7.4.37. *Every wff φ is equivalent to a wff φ' where φ' is in DNF.*

Proof. Analogous to the proof of Theorem 7.4.27 and left as an exercise to the reader. □

Corollary 7.4.38. *Every formula φ is equivalent to a formula φ' that uses only the connectives $\neg$ and $\vee$.*

There is a useful connection between CNF/DNF and tautologies and contradictions:

Lemma 7.4.39. *A wff φ in CNF is a tautology iff every conjunction contains some atom p and its negation.*

Lemma 7.4.40. *A wff φ in DNF is a contradiction iff every disjunction contains some atom p and its negation.*

Proof. Exercise for the reader. □

There are three reasons why we care about semantic equivalence as defined above. The first arises when we are translating from English into a propositional language. English contains many constructions that are not easily and immediately mappable onto the basic propositional connectives. One example of this is the English connective 'unless'. It is not immediately clear what the logical form of the English sentence "I don't do the dishes unless you vacuum" is, though it is clear that it is a combination of the negation "I don't do the dishes" and "You vacuum". What, then, is the best way to represent the connective 'unless'? It is clearly *not* 'and': "I don't do the dishes and you vaccuum" does not, intuitively, say the same thing as "I don't do the dishes unless you vacuum". What about the following?

1. Either I don't do the dishes or you vacuum.

2. If I do the dishes, then you vacuum.

3. If you don't vacuum, I don't do the dishes.

All of these, namely, $\neg\varphi \vee \psi$, $\varphi \rightarrow \psi$, and $\neg\psi \rightarrow \neg\varphi$, are intuitively equivalent, even though they are syntatically distinct strings. As it turns out, they are all in fact semantically equivalent (have the same meaning), as is illustrated by Table 7.11.

φ	ψ	$\neg\varphi$	$\neg\psi$	$\neg\varphi$	$\vee$	ψ	φ	$\rightarrow$	ψ	$\neg\psi$	$\rightarrow$	$\neg\varphi$
T	T	F	F		T			T			T	
T	F	F	T		F			F			F	
F	T	T	F		T			T			T	
F	F	T	T		T			T			T	

Table 7.11: Equivalence of disjunction, implication + negation, and contraposition

From this we can see two things: First, that it does not matter which of these three syntactically different wffs we use to render the English-language sentence, because they all have the same truth value. Second, that we did not have to introduce $\rightarrow$ into our language as a primitive; we could instead *define* $\varphi \rightarrow \psi$ to be $\neg\varphi \vee \psi$. Many of the connectives are interdefinable in this way (a fact that the Stoic logicians themselves recognized): Any one of the binary connectives along with negation is sufficient to define the other two.

A second reason the notion of equivalence is valuable is because, as noted above, it allows us to simplify notation and drop the internal parentheses in conjunctions with more than two conjuncts and disjunctions with more than two disjunctions.

A third reason is that it gives us an insight into what sorts of things must be true (or valid), given that other things are true (or valid). We will use semantic equivalences as a

guide when we develop our proof theory for propositional logic, in §7.5. If two formulas are semantically equivalent, we will introduce derived rules that allow us to incorporate these equivalencies directly into our proofs.

A fourth reason for being interested in normal forms and equivalences is because of applications in automated reasoning. I will add some references here sometime in the future, since this is not a topic to be covered in this book.

7.4.2 Shortened truth-table method

If φ contains n distinct atoms, then the full truth table for φ will contain 2^n lines. This means that once $n \geq 4$, calculating a full truth table for φ by hand becomes quite tedious. It is not necessary, however, to produce a full truth table to determine whether φ is a tautology (or a contradiction). We introduce the shortened truth-table method, which is based on trying to construct a falsifying (or verifying) instance.

This method starts from the assumption that one *can* make a given wff false (or true), and then determines how truth values to the subformulas would have to be assigned in order for this to be the case. If the assumption is born out—that is, if it is possible to make the wff false (or true) then you can conclude that it is not a tautology (or a contradiction). If, however, your assumption leads you to assign a single atomic formula both 'true' and 'false' then you know that the formula cannot be falsified (or verified). Let us go through some examples.

Example 7.4.41. Is $p \to ((p \wedge \neg p) \to q)$ a tautology?

If it is, then there is no truth value assignment that makes it false. Suppose that there is in fact one. The only way that a conditional is false is when the antecedent is true and the consequent false:

p	$\to$	$((p$	$\wedge$	$\neg$	$p)$	$\to$	$q)$
T	F					F	

Thus, we already know what value one of the atoms has on this truth-value assignment: $V(p) = T$. The consequent is itself a conditional. The only way in which the consequent is false is when its antecedent is true and its consequent is false:

p	$\to$	$((p$	$\wedge$	$\neg$	$p)$	$\to$	$q)$
T	F		T			F	F

We have now uniquely determined the truth values of the two atoms, p and q: $V(p) = T$ and $V(q) = F$. But this presents us with difficulty when trying to make the antecedent of the second conditional true: For if p is true, the $\neg p$ must be false, but that means $p \wedge \neg p$ is false, rather than true as we require. Our assumption that we could find a falsifying instance has proven contradictory, and from this we can conclude that there is no way to make this wff false. Thus, it is a tautology.

Example 7.4.42. Is $(j \to (k \to l)) \to (k \to (j \to l))$ a tautology?

If it is, then there is no truth value assignment that makes it false. Suppose that there is in fact one. The only way that a conditional is false is when the antecedent is true and the consequent false:

j	$\to$	$(k$	$\to$	$l))$	$\to$	$(k$	$\to$	$(j$	$\to$	$l))$
	T				F		F			

There are many different ways for a conditional to be true, so let's concentrate on the one that must be false. For it to be false, again, the antecedent must be true and the consequence false:

j	$\to$	$(k$	$\to$	$l))$	$\to$	$(k$	$\to$	$(j$	$\to$	$l))$
	T				F	T	F		F	

Now that we know that k must be true, we can fill this in throughout the table:

j	$\to$	$(k$	$\to$	$l))$	$\to$	$(k$	$\to$	$(j$	$\to$	$l))$
	T	T			F	T	F		F	

We again have a conditional that needs to be false: Thus, we can make the antecedent true and the consequent false:

j	$\to$	$(k$	$\to$	$l))$	$\to$	$(k$	$\to$	$(j$	$\to$	$l))$
	T	T			F	T	F	T	F	F

If l is false, then $k \to l$ will be false:

j	$\to$	$(k$	$\to$	$l))$	$\to$	$(k$	$\to$	$(j$	$\to$	$l))$
	T	T	F	F	F	T	F	T	F	F

But if $k \to l$ is false, j needs to be false in order for $j \to (k \to l)$ to be true. However, we have already assigned j true. This is a contradiction; so as in the previous example, our assumption that there is a falsifying instance of $(j \to (k \to l)) \to (k \to (j \to l))$ was proven contradictory. So, this is a tautology.

Example 7.4.43. Is $((q \to p) \wedge (\neg q \to r)) \to \neg(p \vee r)$ a tautology?

If it is, then there is no truth value assignment that makes it false. Suppose that there is one. Then both conjuncts in the antecedent need to be true and the consequent needs to be false:

q	$\to$	p	$\neg$	q	$\to$	r	$\neg$	$(p$	$\vee$	$r)$
	T				T		F			

For the negation of the disjunction to be false, the disjunction must be true:

q	$\to$	p	$\neg$	q	$\to$	r	$\neg$	$(p$	$\vee$	$r)$
	T				T		F		T	

There are three different ways in which this disjunction can be true:

q	$\to$	p	$\neg$	q	$\to$	r	$\neg$	$(p$	$\vee$	$r)$
	T	T			T	T	F	T	T	T
	T	F			T	T	F	F	T	T
	T	T			T	F	F	T	T	F

Let us consider the case where p is false. When p is false, q must also be false in order for $q \to p$ to remain true:

q	$\to$	p	$\neg$	q	$\to$	r	$\neg$	$(p$	$\vee$	$r)$
	T	T			T	T	F	T	T	T
F	T	F		F	T	T	F	F	T	T
	T	T			T	F	F	T	T	F

If q is false, then $\neg q$ is true:

q	$\rightarrow$	p	$\neg$	q	$\rightarrow$	r	$\neg$	$(p$	$\vee$	$r)$
	T	T			T	T	F	T	T	T
F	T	F	T	F	T	T	F	F	T	T
	T	T			T	F	F	T	T	F

This is consistent with $\neg q \rightarrow r$ being true, as required. Thus, we have filled in all the columns in the truth table in a line that makes the two conjuncts of the antecedent true (and hence the entire conjunction true) and the consequent false, where $V(p) = F$, $V(q) = F$, and $V(r) = T$. Thus, this is not a tautology.

7.5 Proof theory

The proof system for a logic always consists of two parts: Axioms and rules of inference. These two components are related to each other in an important way: The more axioms you have, the fewer rules of inference you need, and the more rules of inference you have, the fewer axioms you need. In §6.3, we saw an example of a proof system where there are roughly as many axioms as there are rules of inference, so there is roughly parity. In that system, we could have removed the reductio ad absurdum rule from the proof rules if we added both Baroco and Bocardo to the list of axioms.

Axiomatic proof systems for propositional logic go towards one of the extremes: They use numerous axioms but only two rules of inference, *modus ponens* and substitution; we discuss such systems in §7.5.2. In the next section, we go to the other end of the extreme: The proof system introduced there will have ten basic rules (in §7.5.1 we derive some non-basic rules) and zero axioms.

7.5.1 Natural deduction

The general process behind our proof system is that we begin with complex sentences (the premises of our argument) and then break them down into simpler sentences which we can then recombine to build up new complex sentences, ultimately ending with our conclusion. This process of breaking down and rebuilding sentences is done via *elimination* and *introduction* rules: We have one rule of each type for each connective, plus two rules which are 'book-keeping' rules.

This general process covers a multitude of very bad proof systems. We don't want to be able to break down and recombine our sentences in any old way, we would rather like

to do so in a useful way. In the context of logic, 'useful' here is *truth-preserving*: We want our rules to be such that by applying them we are never lead from truth to falsity. If our rules of inference are truth-preserving, then we say that they are *sound* with respect to the semantics outlined in §7.4. We prove that the rules introduced in this section are sound in §7.6.

Definition 7.5.1 (Proof)**.** A *proof* (or a *derivation*) is a finite list of numbered formulas such that every formula is annotated with one of our proof rules (basic or derived). If there is a proof of φ from assumptions $\psi_1, \psi_2, \ldots, \psi_n$, we write $\psi_1, \psi_2, \ldots, \psi_n \vdash \varphi$. If there is a proof of φ from no assumptions, we write $\vdash \varphi$.

Definition 7.5.2. If $\vdash \varphi$, then we call φ a *theorem*.

Note that the single turnstile, $\vdash$, should not be confused with the double turnstile, $\vDash$, introduced in Definition 7.4.8. Both are symbols used to say something about a logical language, but the single turnstile is concerned with proof and the double turnstile with truth. In §7.6, we show how these two distinct definitions actually capture the *same* notion of "good argument".

'Bookkeeping' rules

There are two bookkeeping rules, Assumption and Reiteration.

Rule 7.5.3 (Assumption)**.** If φ is a wff, then φ can be written on line n. The annotation for line n is Assumption.

$n\ \varphi$ Assumption

This rule is justified by the fact that one can always reason hypothetically, that *if* something is the case, *then* something else must be the case. Every assumption has a *scope line*, a vertical line drawn on the left to indicate that every formula to the right of the vertical line depends on that assumption. Every formula is said to *depend* on the assumptions located on every scope line to the formula's left, and one formula, φ, is *under* or *within the scope* of another formula, ψ, if every scope line to the left of ψ is also to the left of φ. The assumption itself is indicated by a perpendicular line extending from the vertical line.

Every proof starts off by the assumption of one or more wffs, the premises. If more than one premise is given, then all of the premises can be assumed together in the first step

(albeit each on their own separate line of the proof for ease of reference). After that, each new assumption is given a new scope line of its own. These scope lines continue until a rule that closes the scope line is applied. We then say that the assumption is "discharged". In a theorem, there will be no undischarged assumptions. The only rules that discharge assumptions and close scope lines are Rule 7.5.10 ($\rightarrow$ I), Rule 7.5.12 ($\neg$I), Rule 7.5.13 ($\neg$E), and Rule 7.5.9 ($\vee$E), all defined below.

The final line of a proof must be either within the scope of the *only* the initial assumptions, or within the scope of no assumptions whatsoever. If some line of a proof is within the scope of more assumptions than the initial one(s), then it is not yet a completed proof.

Rule 7.5.4 (Reiteration)**.** If φ occurs on line i of a proof, $i < n$, and n is within the scope of i, then φ can be written on line n. The annotation for line n is Reiteration, i.

$i\ \varphi$	
$\vdots$	
$n\ \varphi$	Reiteration

Reiteration is used to bring formulas within the scope of new assumptions. It can never be used to move formulas outside of the scope of the assumptions they depend on. This is the same Reiteration rule as we had in the syllogistic proof theory (cf. Rule 6.3.4.)

Note 7.5.5. Each of the introduction and elimination rules for the connectives can only be applied to lines that have the same scope lines to their left.

Rules for $\wedge$

Rule 7.5.6 ($\wedge$I)**.** If φ occurs on line i of a proof, ψ occurs on line j of a proof, $i < n$, and $j < n$, and n within the scope of i and j, then $\varphi \wedge \psi$ can be written on line n. The annotation for line n is $\wedge$I, i, j.

$i\ \varphi$	
$\vdots$	
$j\ \psi$	
$\vdots$	
$n\ \varphi \wedge \psi$	$\wedge$I, i, j

That is, if you can prove two wffs independently, then you can prove the conjunction of them. The intuitive justification for this rule is that if two wffs are true, then their conjunction is also true.

Rule 7.5.7 ($\wedge$E)**.** If $\varphi \wedge \psi$ occurs on line i of a proof and $i < n$, then φ can be written on line n. The annotation for line n is $\wedge\mathsf{E}, i$. Alternatively, if $\varphi \wedge \psi$ occurs on line i of a proof and $i < n$, then ψ can be written on line n. The annotation for line n is $\wedge\mathsf{E}, i$.

$$
\begin{array}{|l|l}
\hline
\\
\vdash & \\
\quad i\ \varphi \wedge \psi & \\
\quad \vdots & \\
\quad n\ \varphi & \wedge\mathsf{E}, i
\end{array}
$$

or

$$
\begin{array}{|l|l}
\hline
\\
\vdash & \\
\quad i\ \varphi \wedge \psi & \\
\quad \vdots & \\
\quad n\ \psi & \wedge\mathsf{E}, i
\end{array}
$$

That is, if you can prove a conjunction of two wffs, then you can prove each of the conjuncts individually. The intuitive justification for this rule is that if a conjunction is true, then both conjuncts are also true.

Rules for $\vee$

Rule 7.5.8 ($\vee$I)**.** If φ occurs on line i of a proof and $i < n$, then $\varphi \vee \psi$ can be written on line n. The annotation for line n is $\vee\mathsf{I}, i$. Alternatively, if φ occurs on line i of a proof and $i < n$, then $\psi \vee \varphi$ can be written on line n. The annotation for line n is $\vee\mathsf{I}, i$.

$$
\begin{array}{|l|l}
\hline
\\
\vdash & \\
\quad i\ \varphi & \\
\quad \vdots & \\
\quad n\ \varphi \vee \psi & \vee\mathsf{I}, i
\end{array}
$$

or

$$
\begin{array}{|l}
\hline
i\ \varphi \\
\vdots \\
n\ \psi \vee \varphi \qquad\qquad \vee\mathsf{I}, i
\end{array}
$$

That is, if you can prove a wff, then you can prove that wff disjoined with any other wff, regardless of whether that other wff has previously been proved or not. The intuitive justification for this rule is that if a wff is true, then the disjunction of that wff and anything else will also be true.

Rule 7.5.9 ($\vee\mathsf{E}$)**.** If $\varphi \vee \psi$ occurs on line i of a proof and from an the assumption of φ at line j you can prove γ at line k, and from the assumption of ψ at line l you can prove γ at line m, where $i, j, k, l, m < n$, $j < k$, and $l < m$, then γ can be written on line n within the same scope lines as line i. The annotation for line n is $\vee\mathsf{E}, i, j$–k, l–m.

$$
\begin{array}{|l}
\hline
i\ \varphi \vee \psi \\
\vdots \\
\quad \begin{array}{|ll}
\hline
j\ \varphi & \text{Assumption} \\
\hline
\quad \vdots & \\
k\ \gamma &
\end{array} \\
\quad \begin{array}{|ll}
\hline
l\ \psi & \text{Assumption} \\
\hline
\quad \vdots & \\
m\ \gamma &
\end{array} \\
n\ \gamma \qquad\qquad \vee\mathsf{E}, i, j\text{–}k, l\text{–}m
\end{array}
$$

That is, if you have proven a disjunction, and you can prove another wff from the assumption of the first disjunct *and* from the assumption of the second disjunct, then you can prove that wff. The intuitive justification for this rule is that if a disjunction is true, then anything that follows from both of the disjuncts must also be true. The assumptions made at lines j and l are said to be 'discharged' at line n.

Rules for →

Rule 7.5.10 (→ I)**.** If from an assumption of φ on line i of a proof it is possible to prove ψ at line j, $i < j < n$, then $\varphi \to \psi$ can be written on line n, with the new scope line introduced at line i terminated at line j. The annotation for line n is $\to \mathsf{I}, i\text{–}j$.

That is, if from the assumption of a wff φ you can prove another wff ψ, then you can prove the implication of ψ from φ. The intuitive justification for this rule is that this is just what "if…then…" means. The assumption made at line i is said to be 'discharged' at line n.

Rule 7.5.11 (→ E)**.** If $\varphi \to \psi$ occurs on line i of a proof, φ occurs on line j, $j < n$, and $i < n$, then ψ can be written on line n. The annotation for line n is $\to \mathsf{E}, i, j$.

That is, if you can prove a conditional, and you can prove the antecedent, then you can prove the consequent. The intuitive justification for this rule is that if a conditional and its antecedent are both true, then the consequent of the conditional must also be true.

Rules for ¬

Rule 7.5.12 (¬I)**.** If from an assumption of φ on line i of a proof it is possible to prove ψ at line j and $\neg\psi$ at line k, $i < j < n$ and $i < k < n$, then $\neg\varphi$ can be written on line n, with

the scope line initiated at line i terminated at line k. The annotation for line n is $\neg\mathsf{I}, i\text{–}k$.

That is, if from the assumption that a wff φ is true you can prove both some wff and its negation (an impossible or contradictory situation), then you know that your initial assumption was wrong and that φ must be false. The assumption made at line i is said to be 'discharged' at line n.

Rule 7.5.13 ($\neg\mathsf{E}$)**.** If from an assumption of $\neg\varphi$ on line i of a proof it is possible to prove ψ at line j and $\neg\psi$ at line k, $i < j < n$ and $i < k < n$, then φ can be written on line n with the scope line initiated at line i terminated at line k. The annotation for line n is $\neg\mathsf{E}, i\text{–}k$.

That is, if from the assumption that a wff φ is false you can prove both some wff and its negation (an impossible or contradictory situation), then you know that your initial assumption was wrong and that φ must be true. The assumption made at line i is said to be 'discharged' at line n.

Examples

We are now in a position to give some example proofs, to show how these rules can be put together. In the next section (§7.5.1) we'll discuss strategies for generating proofs.

Example 7.5.14. $(p \wedge q) \to r, \neg r, p \vdash \neg q$:

$1\ (p \wedge q) \to r$	Assumption
$2\ \neg r$	Assumption
$3\ p$	Assumption
⊢	
$\quad 4\ q$	Assumption
$\quad$ ⊢	
$\quad 5\ p$	Reiteration,3
$\quad 6\ p \wedge q$	∧I,4,5
$\quad 7\ (p \wedge q) \to r$	Reiteration,1
$\quad 8\ r$	→ E, 6, 7
$\quad 9\ \neg r$	Reiteration,2
$10\ \neg q$	¬I, 4–9

Example 7.5.15. $p \to q, p \to r \vdash p \to (q \wedge r)$.

$1\ p \to q$	Assumption
$2\ p \to r$	Assumption
⊢	
$\quad 3\ p$	Assumption
$\quad$ ⊢	
$\quad 4\ p \to q$	Reiteration,1
$\quad 5\ p \to r$	Reiteration,2
$\quad 6\ q$	→ E, 3, 4
$\quad 7\ r$	→ E, 3, 5
$\quad 8\ q \wedge r$	∧I, 6, 7
$9\ p \to (q \wedge r)$	→ I, 3–8

Example 7.5.16. $(g \vee h) \to (s \wedge t), (t \vee u) \to (c \wedge d) \vdash g \to c$.

1 $(g \vee h) \to (s \wedge t)$	Assumption
2 $(t \vee u) \to (c \wedge d)$	Assumption
3 g	Assumption
4 $g \vee h$	$\vee$I, 3
5 $(g \vee h) \to (s \wedge t)$	Reiteration, 1
6 $s \wedge t$	$\to$ E, 4, 5
7 t	$\wedge$E, 6
8 $t \vee u$	$\vee$I, 7
9 $(t \vee u) \to (c \wedge d)$	Reiteration, 2
10 $c \wedge d$	$\to$ E, 8, 9
11 c	$\wedge$E, 10
10 $g \to c$	$\to$ I, 3–11

Example 7.5.17. $f \vee (d \to t), \neg f, d \vdash t$ We give two different proofs for this. The first proves t directly:

1 $f \vee (d \to t)$	Assumption
2 $\neg f$	Assumption
3 d	Assumption
4 f	Assumption
5 $\neg(d \to t)$	Assumption
6 $\neg f$	Reiteration, 2
7 f	Reiteration, 4
8 $d \to t$	$\neg$E, 5–7
9 $d \to t$	Assumption
10 $d \to t$	Reiteration, 9
11 $d \to t$	$\vee$E, 1, 4–8, 9–10
12 t	$\to$E, 3, 11

The second proves t following the recommendation of the proof strategies outlined in §7.5.1, step 4.

1 $f \vee (d \to t)$	Assumption
2 $\neg f$	Assumption
3 d	Assumption
4 $\neg t$	Assumption
5 f	Assumption
6 $\neg(d \to t)$	Assumption
7 $\neg f$	Reiteration, 2
8 f	Reiteration, 5
9 $d \to t$	$\neg$E, 6–8
10 $d \to t$	Assumption
11 $d \to t$	Reiteration, 10
12 $d \to t$	$\vee$E, 1, 5–9, 10–11
13 t	$\to$E, 3, 12
14 t	$\neg$E, 4–13

What you can see here is that the assumption of $\neg t$ at line 4 doesn't really play any role in the proof: The contradiction that we obtained was with $\neg t$ itself, which means that we were able to prove t directly without needing to invoke line 4. Nevertheless, this is still a correct proof, albeit one with superfluous lines.

Proof strategies

The basic method of proof is to use the elimination rules to break down complex formulas into simpler ones, and then use the introduction rules to build them back up into different complex formulas.

Of course, this doesn't say much about *when* to break down complex formulas into simpler ones or *how* to build them back up into more complex formulas. Strategies for proofs can be developed through practice, and by recognising repeating patterns. However, because developing these strategies requires practice, here is a technique that can be used to get started in proofs:

1. Do you have any assumptions? If yes, go to (1a); if no, go to (1b).

    a) Use as many elimination rules as you can. Then go to (1b).

    b) What is the main connective of what you are trying to prove?

        i. It is a conjunction. Go to (2).

        ii. It is a disjunction. Go to (3).

        iii. It is a negation or an atom. Go to (4).

        iv. It is a conditional. Go to (5).

        v. I am trying to prove a contradiction, but don't know *which* contradiction. Go to (8).

2. If the wff you are trying to prove is a conjunction, then you can obtain it via $\wedge$I (Rule 7.5.6) if you can obtain both conjuncts. Consider each conjunct individually (return to (1)).

3. If the wff you are trying to prove is a disjunction, then you can obtain it via $\vee$I (Rule 7.5.8) if you can obtain a disjunct. Does either disjunct contain any of the atoms that appear in any of your premises? If yes, then try to prove that disjunct (return to 1b)).

4. If the wff you are trying to prove is an atom or a negation, then you can obtain it by $\neg$E (Rule 7.5.13) or $\neg$I (Rule 7.5.12), respectively. That is, if you are trying to prove an atom, make its negation a new assumption; if you are trying to prove a negated atom or complex formula, make the atom or complex formula (without the negation) an assumption. Then return to (1).

5. If the wff you are trying to prove is a conditional, then you can obtain it by $\rightarrow$ I (Rule 7.5.10). That is, make the antecedent a new assumption, and try to prove the consequent (return to (1)).

6. Good luck, this is one context in which no algorithm can reliably tell you what to do. Rely on the intuitions that you develop through practice, but also take a look at the next hints:

What to do when you get stuck:

- Look at any assumptions you haven't yet used, and see if you can now use them.

- Apply derived rules (see §7.5.1) wherever you can.

Derived rules

Proofs which use only the ten rules introduced above are often long and clunky, involving intricate subproofs that cannot be generated by the above strategy, and which are unintuitive to a beginner. In this section, we introduce the notion of *derived rules*. These are rules which are not taken as basic, but can be considered 'short-cuts', in that each application of a rule is justified by appeal to a proof from the premise(s) to the conclusion that uses only the ten basic rules.

There are two types of derived rules that we consider. The first type consists in three rules which function very similarly to the basic rules of the previous section, in that they take two wffs from previous lines and allow you to write down a third, new wff on a new line. The three rules are modus tollens, disjunctive syllogism, and transitivity of $\rightarrow$. We shall simply call these three rules "derived rules" without any further modifer.

The second type of derived rule are the so-called *rules of replacement*. These do not work quite in the same way as the basic rules or the derived rules of the other type, in that it is allowed to apply them not only to wffs written on earlier lines of a proof, but also to *subformulas* of those wffs; we will explain this further below. The rules of replacement we introduce in this section are contraposition, DeMorgan's laws, double negation, and exportation.

The first derived rules that function like the basic rules is modus tollens:

Rule 7.5.18 (Modus tollens)**.**

$$\begin{array}{|l}
\\
\hline
i\ \varphi \rightarrow \psi \\
\vdots \\
j\ \neg\psi \\
\vdots \\
n\ \neg\varphi \qquad\qquad\qquad\qquad\qquad\qquad \mathsf{MT}, i, j
\end{array}$$

That is, if you can prove both a conditional and the negation of its consequent, then you can prove the negation of the antecedent. The intuitive justification is that if the consequent of a conditional is false, then the only way the conditional itself can be true is if the antecedent is also false.

Lemma 7.5.19. *Anything you can prove with* MT *you can also prove without it.*

Proof. Suppose you have a proof with $\varphi \rightarrow \psi$ on line i and $\neg\psi$ on line j, and you wish to prove $\neg\varphi$ on some line following both i and j. The following proof will suffice for any φ and

ψ:

$i\ \varphi \to \psi$	
$\vdots$	
$j\ \neg\psi$	
$\quad k\ \varphi$	Assumption
$\quad k+1\ \psi$	$\to$E, i, k
$\quad k+2\ \neg\psi$	Reiteration, j
$k+3\ \neg\varphi$	$\neg$I, k–$k+2$

□

The second derived rule is disjunctive syllogism. It comes in two variants, like $\wedge$E and $\vee$I:

Rule 7.5.20 (Disjunctive syllogism)**.**

$i\ \varphi \vee \psi$	
$\vdots$	
$j\ \neg\psi$	
$\vdots$	
$n\ \varphi$	DS, i, j

or

$i\ \varphi \vee \psi$	
$\vdots$	
$j\ \neg\varphi$	
$\vdots$	
$n\ \psi$	DS, i, j

That is, if you can prove both a disjunction and the negation of one of its disjuncts, then you can prove the other disjunct. The intuitive justification is that if a disjunction is true, at least one of the disjuncts is true. If you know that one of the disjuncts is false,

then it has to be the case that the other one is true. As with modus tollens, anything you can prove with DS you can prove without it, using only the basic rules. We leave the proof of this to the reader.

Once you have DS, its use will almost wholly eclipse the use of $\vee$E.

The third derived rule of this type is the transitivity of $\rightarrow$:

Rule 7.5.21 (Transitivity of $\rightarrow$)**.**

$$\begin{array}{|l} \\ \vdash \\ \quad i\ \varphi \rightarrow \psi \\ \quad \vdots \\ \quad j\ \psi \rightarrow \chi \\ \quad \vdots \\ \quad n\ \varphi \rightarrow \chi \end{array} \qquad \mathsf{Trans}\rightarrow, i, j$$

That is, if you can prove two conditionals, where the consequent of one is the antecedent of the other, then you can prove a third conditional from the antecedent of the first to the consequent of the second. (This rule was often taken as one of the basic proof rules by logicians in the early 14th century; the rule was known as "From the first to the last".) The intuitive justification of this is that anything that follows from something that follows from something follows from that thing directly.

Finally, recall that we defined, above, $\varphi \leftrightarrow \psi$ as short-hand for $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$. Because this is a definition using only $\rightarrow$ and $\wedge$, strictly speaking $\leftrightarrow$ does not need its own set of introduction and elimination rules. However, it can be a useful short-cut to have them, so we will add $\leftrightarrow$I and $\leftrightarrow$E to our list of derived rules:

Rule 7.5.22 ($\leftrightarrow$I)**.**

$$\begin{array}{|l} \\ \vdash \\ \quad i\ \varphi \rightarrow \psi \\ \quad \vdots \\ \quad j\ \psi \rightarrow \varphi \\ \quad \vdots \\ \quad n\ \varphi \leftrightarrow \psi \end{array} \qquad \leftrightarrow \mathsf{I}, i, j$$

That is, if you can prove two conditionals, where the antecedent of one is the consequent of the other, and the consequent of the first is the antecedent of the second, then you can

write down the biconditional. The intuitive justification for this is that if you can prove both conditionals, then you could prove their conjunction as well (via $\wedge$I), then you can rewrite the conjunctions of the two conditionals via the introduction of the new notation.

Rule 7.5.23 ($\leftrightarrow$ E)**.**

$$
\begin{array}{|l|l}
\hline
i\ \varphi \leftrightarrow \psi & \\
\vdots & \\
j\ \varphi & \\
\vdots & \\
n\ \psi & \leftrightarrow \mathsf{E}, i, j \\
\hline
\end{array}
$$

or

$$
\begin{array}{|l|l}
\hline
i\ \varphi \leftrightarrow \psi & \\
\vdots & \\
j\ \psi & \\
\vdots & \\
n\ \varphi & \leftrightarrow \mathsf{E}, i, j \\
\hline
\end{array}
$$

or

$$
\begin{array}{|l|l}
\hline
i\ \varphi \leftrightarrow \psi & \\
\vdots & \\
j\ \neg\varphi & \\
\vdots & \\
n\ \neg\psi & \leftrightarrow \mathsf{E}, i, j \\
\hline
\end{array}
$$

or

$$
\begin{array}{|l|l}
\hline
i\ \varphi \leftrightarrow \psi & \\
\vdots & \\
j\ \neg\psi & \\
\vdots & \\
n\ \neg\varphi & \leftrightarrow \mathsf{E}, i, j \\
\hline
\end{array}
$$

This is simply a restatement of →I and MT for both conditionals.

Rules of Replacement

The rules of replacement that we introduce in this section function somewhat differently from the fifteen basic and derived rules defined above. Instead of taking a single wff or a pair of wffs and using them as justification for writing down some wholly new wff (though usually involving some of the subformulas of the wffs used to generate it!), we take a wff and modify it by replacing it, or a subformula of it, with something which is semantically equivalent (see Definition 7.4.18), hence their name, "rules of replacement". Thus these rules, *and these rules alone*, can be applied not only to the full wff, but also to any part of a wff that has the requisite form.

The first rule of replacement is contraposition. This rule is very closely related to MT, and in fact can be derived using →I and MT. The intuitive justification of this rule is that if you can prove a conditional, then you can also prove that if you have the negation of the consequent of the conditional you can also prove the negation of the antecedent.

Rule 7.5.24 (Contraposition)**.** If $\varphi \to \psi$ occurs on line i of a proof, whether as a wff or as a subformula, it can be replaced with $\neg\psi \to \neg\varphi$ on line n, or vice versa, with justification Contra., i.

Rather than giving a schematic representation of the rule (which, if we tried to represent it in full generality would actually be more confusing than less), we will give two examples of it the rule, one where we apply it to an entire wff, and one where we apply it to a subformula:

⋮	
$i\ \gamma \to (\varphi \to \psi)$	
⋮	
$m\ \gamma \to (\neg\psi \to \neg\varphi)$	Contra., i
⋮	
$n\ \neg(\varphi \to \psi) \to \neg\gamma$	Contra., i

The second rule of replacement are the proof-rule correlates of DeMorgan's laws.

Rule 7.5.25 (DeMorgan's)**.** There are two versions of this law, one for negated conjunctions and one for negated disjunctions. (1) If $\neg(\varphi \wedge \psi)$ occurs on line i of a proof, whether as

a wff or as a subformula, it can be replaced with $\neg\varphi \vee \neg\psi$ on line n, or vice versa, with justification DeM, i. (2) If $\neg(\varphi \vee \psi)$ occurs on line i of a proof, whether as a wff or as a subformula, it can be replaced with $\neg\varphi \wedge \neg\psi$ on line n, or vice versa, with justification DeM, i.

These rules are introduced because of the semantic equivalence between conjunctions of negations and negated disjunctions, and between disjunctions of negations and negated conjunctions, which was proven in Lemmas 7.4.33 and 7.4.34.

Rule 7.5.26 (Double negation)**.** If φ occurs on line i of a proof, whether as a wff or as a subformula, it can be replaced with $\neg\neg\varphi$ on line n, or vice versa, with justification DN, i.

That is, for any formula, you can either take two negations off it, or add two negations to it. The intuitive justification is that if a formula is true, then its negation is false, and the negation of the negation is true.

Rule 7.5.27 (Exportation)**.** If $(\varphi \wedge \psi) \rightarrow \gamma$ occurs on line i of a proof, whether as a wff or as a subformula, it can be replaced with $\varphi \rightarrow (\psi \rightarrow \gamma)$ on line n, or vice versa, with justification Exp, i.

That is, if you can prove that from a conjunction you can prove some other wff, then you can prove that if you have the first conjunct, then if you have the second one, you can get the target wff; and if you can prove that if you have some wff φ, then if you have another one ψ, you can get a target formula, then if you have the conjunction of φ and ψ together, you can also get the target formula. The intuitive justification comes from the fact that it doesn't matter whether you're given premises all at once or if you're given them one at a time.

The following example illustrates some of the derived rules.

Example 7.5.28. $\neg(p \to q) \vdash p \wedge \neg q$

1 $\neg(p \to q)$	Assumption
2 $\neg(p \wedge \neg q)$	Assumption
3 $\neg p \vee \neg\neg q$	DeM, 2
4 p	Assumption
5 $\neg\neg p$	DN, 4
6 $\neg\neg q$	DS, 3, 5
7 q	DN, 6
8 $p \to q$	→I, 4–7
9 $\neg(p \to q)$	Reit, 1
10 $p \wedge \neg q$	¬E, 2–9

The derived rules allow us to significantly shorten and simplify proofs. However, they do not allow us to prove anything *more*; just as with the other derived rules, anything that you can prove with the rules of replacement, you can also prove without. This will be easier to demonstrate after we have proven that the basic rules are a sound and complete proof system for classical propositional logic, which we do in §7.6, so we will defer the proof of this until then. However, since anything that we can prove with the derived rules we can prove without, learning the derived rules is optional, as there will never be a proof where you are *required* to use them.

7.5.2 Axiomatic propositional logic

In this section, we look at a different proof system for propositional logic, one that is more similar to the approach we took to the syllogistic, involving both axioms and rules of inference. But while in the syllogistic we had roughly as many axioms as we had rules of inference, in this proof system we will have ten axioms, and only two rules of inference.

Definition 7.5.29. The propositional axiomatic calculus consists in the following:

- Axioms:
  1. $p \to (q \to p)$
  2. $(p \to (q \to r)) \to ((p \to q) \to (p \to r))$
  3. $(p \wedge q) \to p$

4. $(p \wedge q) \to q$
5. $p \to (q \to (p \wedge q)$
6. $p \to (p \vee q)$
7. $q \to (p \vee q)$
8. $(p \to r) \to ((q \to r) \to ((p \vee q) \to r))$
9. $(p \wedge \neg p) \to q$
10. $p \vee \neg\, p$

- Rules of inference:

  Modus ponens (MP) If φ is a theorem and $\varphi \to \psi$ is a theorem, then so too is ψ.

  Uniform Substitution (US) The result of uniformly replacing atoms $p_1, \ldots, p_n$ by wffs $\beta_1, \ldots, \beta_n$ in a theorem is also a theorem.

The keen-eyed reader will note that many of the axioms mimic the introduction and elimination rules of the previous proof system.

this section has not been completed

7.6 Soundness and completeness

In this section, we connect the truth-tables of §7.4 and the proof-theory of §7.5.

While it is intuitively clear that we have chosen our proof systems so that they match up nicely to the truth-tables for the propositional connectives, we have not yet made precise the way in which they match up. There are two desiderata for any proof system which is constructed: That it be *sound* and that it be *complete*. A proof system is sound with respect to a given semantics if every theorem of the system is a tautology. A sound proof system will never derive a contradiction; a system which is sound will prove *only* true things. Soundness is a necessary requirement for a proof system: Any unsound proof system is useless and should be junked.

Completeness expresses the converse relation: A proof system is complete with respect to a given semantics if every tautology is a theorem of the system. Completeness is a goal for any adequate proof-theory, but it is a goal that cannot always be obtained: There are semantics for which it is impossible to provide any useful complete proof-system. An example of an incomplete system is given in Chapter 10.

7.6.1 Soundness and completeness of natural deduction

Before we state the soundness and completeness results formally, and prove them, we first prove a useful fact about the propositional proof theory:

Theorem 7.6.1 (Deduction theorem)**.** *Let* Γ *be a set of* $\mathcal{L}_p$ *formulas. Then*

$$\Gamma \vdash \varphi \to \psi \text{ iff } \Gamma, \varphi \vdash \psi$$

Proof. Because this is a biconditional, we must prove both directions.

($\Rightarrow$) Suppose that $\Gamma \vdash \varphi \to \psi$. We need to construct a proof of ψ from the assumption of φ along with all the members of Γ, i.e., a proof of the following form:

$i\ \varphi$	Assumption
$\vdots$	
$n\ \psi$	

We can fill in the dots by taking the proof of $\varphi \to \psi$ and inserting it after line i, renumbering the lines of the original proof and its annotations appropriately. Let j be line on which $\varphi \to \psi$ occurrs. Then, we can annotate line n with $\to \mathsf{E}, i, j$. Thus we have shown that $\Gamma, \varphi \vdash \psi$.

($\Leftarrow$) Suppose that $\Gamma, \varphi \vdash \psi$, that is, we have a proof beginning with the assumption of all the members of Γ, and the assumption of φ on line i and ending with ψ on line n, e.g., the proof contains the following:

$i\ \varphi$	Assumption
$\vdots$	
$n\ \psi$	

We then transform this into a proof of $\varphi \to \psi$ by terminating the scope line and introducing the conditional $\varphi \to \psi$ on line $n+1$ with annotation $\to \mathsf{I}, i$–n. We have then shown that $\Gamma \vdash \varphi \to \psi$. □

A consequence of the Deduction Theorem is that any time we can prove a formula φ on the basis of some assumptions, it will always be possible to prove, without any assumptions, a hypothetical sentence with the assumptions in the antecedent. As a result, every proof can be associated with a proof containing no assumptions. We make use of this fact when we prove soundness.

Theorem 7.6.2 (Soundness)**.**

$$\textit{If } \vdash \varphi \textit{ then } \vDash \varphi$$

Instead of proving Theorem 7.6.2 directly, we instead prove the following:

Lemma 7.6.3.

$$\textit{If } \Gamma \vdash \varphi \textit{ then } \Gamma \vDash \varphi$$

That is, if there is a proof of φ which starts from assuming the formulas in Γ, then every truth value assignment that makes all members of Γ true makes φ true. We prove this fact by showing that every rule of inference is truth-preserving: That is, if we assume that the premises of the rule are true, according to the semantics, then after we've applied a proof rule, the resulting new line of the proof will also be true. Because our specific proof system does not have any axioms, this is sufficient to demonstrate the desired result for any arbitrary theorem φ. We will prove that they are using *proof by induction*, by showing that if the target result holds for simpler proofs it will also hold of proofs which are one-step more complex. In the context of proofs, "simplicity" is interpreted as proof length.

Before we begin the proof, it is important to note the following:

Fact 7.6.4.

1. The first line of every proof will always be an assumption.
2. It is always possible to turn a proof into a proof in which all assumptions have been discharged.

Proof. (1) follows from the fact that every rule other than Assumption requires the existence of lines of proof prior to the application of the rule, so the only way to get a proof started is via an assumption. For proofs which contain rules other than Assumption, (2) follows directly from the $\Leftarrow$ direction of Theorem 7.6.1. Suppose we have a proof that contains *only* an assumption, φ. From this, we can construct a new proof by reiterating the assumption on line 2, and then applying $\rightarrow \mathsf{I}$ on line 3 to obtain $\varphi \rightarrow \varphi$. (It is easy to check that $\varphi \rightarrow \varphi$ is a tautology.) □

Corollary 7.6.5. *If $\vdash \varphi$, then φ is not an assumption.*

Proof of Lemma 7.6.3. Suppose that $\Gamma \vdash \varphi$. Then the annotation of φ is one of the basic rules other than Assumption, and φ occurs on some line n, where $n \geq 2$. We consider four cases: where the annotation of φ is Reiteration, where the annotation of φ is $\wedge\mathsf{I}$, where it is $\rightarrow \mathsf{I}$, and where it is $\neg\mathsf{E}$. The cases of $\wedge\mathsf{E}$, $\neg\mathsf{I}$, $\vee\mathsf{E}$, $\vee\mathsf{I}$, and $\rightarrow \mathsf{E}$ are all left as exercises for the reader, and are proven in a similar fashion.

Reiteration Let n be a line of a proof annotated with Reiteration, i, where $i < n$. We need to show that if the formula at line i is true, then the formula at line n. Because line n is annotated with Reiteration, i, it means that line n has exactly the same formula on it that line i does. Thus, if the formula on line i is true, since the formula on line n is identical to the formula on line i, it too must be true, by LEM and PNC.

$\wedge$**I** Let n be a line of a proof annotated with $\wedge$I, i, j for $i < n$ and $j < n$ and $i \neq j$. That is, the formula at line n is of the form $\varphi \wedge \psi$, where φ occurs on line i and ψ occurs on line j. Assume that the formulas on lines i and j are true; that is, φ is true and ψ is true. It follows from the truth-table for $\wedge$ that $\varphi \wedge \psi$ is also true. But that is the formula on line n. So the formula on line n is true.

$\rightarrow$**I** Let n be a line of a proof annotated with $\rightarrow$ I, i–j, where $i < n$, $j < n$, and $i \neq j$. That is, the formula at line n is of the form $\varphi \rightarrow \psi$ where φ occurs as a new assumption on line i and ψ occurs on line j within the scope of the assumption of i. Because the subproof from i to j is less complex than the entire proof including line n, we assume that the subproof is sound; that is, if the formula on line i (that is, φ) is true, then the formula on line j (that is, ψ) must also be true. We now must show that the formula on line n (that is, $\varphi \rightarrow \psi$) is true. First, consider the case where φ is false. Then $\varphi \rightarrow \psi$ is true, and thus the formula on line n is true. Next, consider the case where φ is true. By the assumption of the soundness of the subproof from i to j, we know that ψ is also true. But then, so is $\varphi \rightarrow \psi$, and thus the formula on line n is true.

$\neg$**E** Let n be a line of a proof annotated with $\neg$E, i–j, where $i < n$, $j < n$, and $i \neq j$. That is, the formula at line n is of the form φ where $\neg\varphi$ occurs as a new assumption on line i and at some k, l, where $i \leq k \leq j$ and $i \leq l \leq j$ there occur formulas ψ and $\neg\psi$. Because the subproof from i to j is less complex than the entire proof including line n, we assume that the subproof is sound; that is, if the formula on line i (that is, $\neg\varphi$) is true, then all formulas up to line j, including lines k and l (that is, ψ and $\neg\psi$), must also be true. However, if the formulas on both lines k and l are true, then their conjunction must also be true (see $\wedge$I case above). But $\psi \wedge \neg\psi$ is a contradiction and can never be true. Thus, there is no case in which the assumption at line i can be true; it must be false. But if the formula at line i (that is, $\neg\varphi$) is false, then the formula on line n (that is, φ) must be true, by LEM.

The rest of the rules are left as exercises for the reader. □

What can be seen from this is that the proofs of the soundness of individual rules all follow a similar pattern: (1) It is assumed that a proof with an application of the relevant rule at some line n exists. (2) From the rule being applied, information about the structure of the proof prior to line n is inferred. (3) Because this part of the proof is simpler than the proof that includes line n, we can apply the inductive hypothesis and assume it is sound. (4) We state what we need to show (e.g., that every truth value assignment that makes the premises of the rule true also make the formula on line n true). (5) We appeal to truth tables in order to show that the desired result holds.

Proof of Theorem 7.6.2. Theorem 7.6.2 follows directly from Lemma 7.6.3 and Theorem 7.6.1. □

A number of readers might struggle with the proof of Lemma 7.6.3 because it hardly seems like there is anything to prove in each case: Of course the introduction and elimination rules track the truth tables for the connectives, because we defined the rules so that they would! But that is precisely the point of this proof: To show that we have in fact defined the rules correctly. To drive this point home, let us define a rule which is *unsound*, and then show exactly where the proof of soundness breaks down. We call the rule $\vee\mathsf{E}^*$; compare it with Rule 7.5.9, which is sound.

Rule 7.6.6 ($\vee\mathsf{E}^*$)**.** If $\varphi \vee \psi$ occurs on line i of a proof and $i < n$, then φ can be written on line n. The annotation for line n is $\vee\mathsf{E}^*, i$.

$$\begin{array}{|l} \\ \hline i\ \varphi \vee \psi \\ \vdots \\ n\ \varphi \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vee\mathsf{E}^*, i \end{array}$$

It should be intuitively clear that this rule is *not* sound, because just because a disjunction is true, it doesn't mean we know anything about the truth values of either disjunct. We will nevertheless try to prove that it is sound, and show exactly where the proof breaks down:

Attempted proof of the soundness of Rule 7.6.6.

$\vee\mathbf{E}^*$ Let n be a line of a proof annotated with $\vee\mathsf{E}^*, i$ for $i < n$. That is, the formula at line n is of the form φ, where $\varphi \vee \psi$ occurs on line i. Assume that the formula on line i is true. Now, we need to show that the formula on line n is true. But if we look at the truth tables for $\vee$, it is possible for a disjunction to be true while the left-hand

disjunct is false: Let φ be false and ψ be true; then $\varphi \vee \psi$, the formula on line i, is true. But φ is the formula on line n. Therefore, it *doesn't* follow, per the truth tables, that if the formula on line i is true that the formula on line n is also true. Therefore, this rule is not sound.

□

Theorem 7.6.7 (Completeness)**.**

$$\textit{If } \vDash \varphi \textit{ then } \vdash \varphi$$

Proof of Theorem 7.6.7. to be written □

7.6.2 Soundness and completeness of axiomatic propositional logic

Chapter 8

Predicate logic [last modified 5 March 26]

8.1 Introduction

The topic of this chapter is predicate logic, aka quantifier logic, aka first-order logic, aka FOL.

In Chapter 7, we introduced the propositional connectives $\neg$, $\wedge$, $\vee$, and $\rightarrow$ in order to express certain types of inference that could not be rendered in a syllogistic fashion. But in doing so, we actually *lost* some expressive power. Consider the following sentences:

- Every number is either even or odd.
- For every number there is a larger prime.
- There is a prime larger than every number.
- This table is brown.
- The wall is red.
- Socrates is mortal.

All of these statements would be rendered by a single atomic propositional letter in propositional logic—including the first one. One might be tempted to render it as $p \vee q$, where $p =$"Every number is even" and $q =$"Every number is odd", but it is immediately clear that "Every number is even or every number is odd" is *not* synonymous with "Every number is either even or odd"; for the latter sentence is true while the former is not.

Rendering each of these as atomic propositions means that any internal structure that these sentences has is lost. And yet, there is quite a bit of structure in these sentences that one might feel is relevant to determining the truth value of the sentences and their inferential relationships with other sentences. For example, these sentences contain:

- Comparatives and other relations.
- Properties.
- Quantities.
- Proper names and definite descriptions.

Intuitively, all of these are relevant for reasoning, and hence should be relevant for logic. It is for this reason that we introduce predicate logic. Predicate logics are more expressive than both the syllogistic (which can handle quantification but not propositional combinations) and propositional logic (which can handle propositional combinations but not quantification). In fact, predicate languages are extensions of propositional languages, but instead of the atomic statements being unanalyzable, they will make complex assertions involving objects and their properties.

8.2 Language

What is a property? The metaphysicians will give you many answers, but we will not be concerned with the metaphysics of properties. For our purposes, a property is the sort of thing that can be used to describe an individual object or a group of objects. The simplest type of property are the properties expressed in English by simple nouns, adjectives, or verbal phrases, such as "is a chair", "is red", or "is running". (That is, they are very similar to the categorical terms that can be expressed in the syllogistic.) However, English also has more complex properties that involve more than one object—such properties are called relations. If the relation relates two objects, then it is a *binary relation*. An example of a binary relation is the relation "is married to". If the relation relates three objects, then it is a *ternary relation*. An example of a ternary relation is the relation "between". Although there are not many naturally-expressible relations in English that involve more than three objects, we do not put any constraint on the number of objects that may be related to each other by a given property in a quantified language.

There are two ways that we can talk about objects in a quantified language. We can name a specific object through the use of a *constant*, such as a proper name, or a *function*, such as "wife of". We can also designate an arbitrary or variable object through the use of

individual variables. It is the variables that are connected to the quantifiers and allow us to make statements about "everything" or "something". We make this precise by introducing the following notation:

Definition 8.2.1. A *quantified language* $\mathcal{L}_q$ consists in:

- Propositional logical connectives: $\neg, \vee, \wedge, \rightarrow$.
- A set of individual variables: x, y, z,...
- Predicate logical operators, called quantifiers: $\forall$ (universal), $\exists$ (existential).
- A set of constants: a, b, c, ...
- A set $\mathcal{F}$ of function symbols: f, g, h, ... Every function is associated with a natural number n, its *arity*.
- A set $\mathcal{R}$ of relation symbols: R, S, T, ... Every relation is associated with a natural number n, its *arity*. We call unary relations *predicates*.[1]
- Punctuation: (,).

It is a standard convention, both in this book and elsewhere, that variables are always indicated by a single letter (or letter plus sub- or superscript). Constants are single letters by default, but can be strings of letters if explicitly defined.

A quantified language $\mathcal{L}_q$ is often augmented with $=$, identity. In what follows we assume, unless stated explicitly otherwise, that our languages include identity.

We now define what counts as a well-formed formula in a quantified language. While in propositional logic we can go immediately from our atomic formulas to more complex formulas, in quantified languages we make a distinction between *terms*, which pick out objects in the language, and *formulas*, which make statements about the objects.

Definition 8.2.2 (Terms)**.** τ is a *quantified term* if one of the following holds:

- τ is a variable.
- τ is a constant.
- There is an n-ary function symbol f and terms $\tau_1, \ldots, \tau_n$ such that τ is $f(\tau_1, \ldots, \tau_n)$.

[1] When it is important that the arity of a function or relation be specified, we indicate it with a superscript, e.g., F^n or R^n. In most cases, the arity will be omitted as it can be determined from context.

We say that variables and constants are *atomic* terms, while the application of a function to a term results in a *complex* term.

Example 8.2.3. Let $\mathcal{L}_q$ be a quantified language containing the constants j and a, one binary function symbol, f, and one binary relation symbol, B. j and a are both atomic terms of $\mathcal{L}_q$. $f(\mathsf{j},\mathsf{j})$ is an example of a complex term of $\mathcal{L}_q$.

We next define how to create atomic formulas. As noted above, atomic formulas in predicate logic have an internal structure, as opposed to atomic formulas in propositional logic which are not further analysable. Atomic formulas can either express the identity between two terms, or that a predicate or relation holds of a set of terms.

Definition 8.2.4 (Atomic formula)**.** φ is an *atomic quantified formula* if one of the following holds:

- There is an n-ary relation symbol R and terms $\tau_1, \ldots, \tau_n$ such that φ is $R\tau_1 \ldots \tau_n$.
- There are two terms τ_1 and τ_2 such that φ is $\tau_1 = \tau_2$.

Example 8.2.5. Let $\mathcal{L}_q$ be defined as in Example 8.2.3. $\mathsf{a} = \mathsf{j}$, $B\mathsf{aj}$, and $B\mathsf{ja}$ are all atomic formulas of $\mathcal{L}_q$.

Note that order matters: $B\mathsf{aj}$ and $B\mathsf{ja}$ are not the same formula.

Definition 8.2.6 (Formula)**.** φ is a *quantified well-formed formula* if one of the following holds:

- φ is an atomic formula.
- There are formulas ψ, χ such that one of the following holds:
  - φ is $\neg\psi$.
  - φ is $(\psi \wedge \chi)$.
  - φ is $(\psi \vee \chi)$.
  - φ is $(\psi \rightarrow \chi)$.
- There is a formula ψ and a variable x such that one of the following holds:
  - φ is $\forall x\psi$.
  - φ is $\exists x\psi$.

For more information on how to "read" these quantified formulas, see §8.2.1.

Note 8.2.7 (Scope)**.** The quantifiers work similarly to negation in terms of scope (cf. Lemma 7.3.8). In the formula $\exists x Px \wedge Sx$, the scope of the existential quantifier is the first conjunct; in the formula $\exists x(Px \wedge Sx)$, the scope is the entire conjunction.

Note 8.2.8 (Simplified notation)**.** In formulas with successive occurrences of the same quantifier, we sometimes omit all but the first quantifier and retain the variables in their original order. E.g., sometimes we write $\forall x \forall y \varphi$ as $\forall xy \varphi$.

Definition 8.2.9 (Main connective)**.** The operator or connective which has been applied last in the construction of the formula is the *main connective*.

Given Definition 8.2.6, it is possible to extend Definition 7.3.7 from propositional to quantified languages. The atomic terms occurring in the leaves are formulas of the form defined in Definition 8.2.4, and clauses for $\forall x$ and $\exists x$ are added in analogy to the clause for $\neg$:

Definition 8.2.10 (Predicate parse trees)**.** Parse trees for predicate languages are composed of *roots*, *branches*, and *leaves*. The wff to be parsed is the *root*, and the atomic formulas occurring in the wff are the *leaves*. The branches are generated in a recursive fashion. The wffs that occur on the branches are called 'nodes' (these include the root and the leaves). We begin with the root. If the root node is a leaf, then the parse tree consists of the leaf/root alone. If the root node is not a leaf, then it is of the form $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \rightarrow \psi$, $\forall x \varphi$, or $\exists x \varphi$.

1. If the node is of the form $\neg\varphi$, $\forall x \varphi$, or $\exists x \varphi$, then there is one branch going out from the node, and the node at the end of the branch is labeled φ.

2. If the node is of the form $\varphi \wedge \psi$, $\varphi \vee \psi$, or $\varphi \rightarrow \psi$, then there are two branches going out from the node. The node at the end of the left branch is labeled φ and the node at the end of the right branch is labeled ψ.

We now have a new set of nodes at the top (or bottom, depending on what direction your branches grow) of our tree. Each of these nodes is either a leaf or not. If it is a leaf, then we do nothing. If it is not a leaf, then the tree is extended according to steps (1) and (2). Continue until every node is a leaf.

We now introduce a number of notions relating to variables. First, we show how to identify the set of variables that occur in a term or a formula, and whether each occurrence of a variable is *free* or *bound* (cf. §1.3 for a review of the set-theoretic notation used).

Definition 8.2.11. For every term $\tau \in \mathcal{L}$, $\mathrm{Var}(\tau)$ is the set of variables occurring in τ. This set is defined recursively as follows:

- $\mathrm{Var}(c) = \varnothing$, where c is a constant in $\mathcal{L}$.
- $\mathrm{Var}(x) = \{x\}$, where x is a variable in $\mathcal{L}$.
- $\mathrm{Var}(f(\tau_1, \ldots, \tau_n)) = \mathrm{Var}(\tau_1) \cup \cdots \cup \mathrm{Var}(\tau_n)$, where f is a function symbol and $\tau_1, \ldots, \tau_n$ are terms in $\mathcal{L}$.

In a wff, some occurrences of the variables will be bound by a quantifier, and some will not be. The following definition recursively defines the variables in a formula, the bound variables in a formula, and the unbound (or free) variables in a formula:

Definition 8.2.12. Let φ be a wff of $\mathcal{L}$. Then:

- If φ is atomic, then $\mathrm{Bound}(\varphi) = \varnothing$ and $\mathrm{Var}(\varphi) = \mathrm{Free}(\varphi) = \mathrm{Var}(\tau_1) \cup \cdots \cup \mathrm{Var}(\tau_n)$, where $\tau_1, \ldots, \tau_n$ are the terms occurring in φ.
- If φ is $\neg\psi$, then $\mathrm{Var}(\varphi) = \mathrm{Var}(\psi)$, $\mathrm{Bound}(\varphi) = \mathrm{Bound}(\psi)$, and $\mathrm{Free}(\varphi) = \mathrm{Free}(\psi)$.
- If φ is $\psi \wedge \chi$, $\psi \vee \chi$, or $\psi \to \chi$, then $\mathrm{Var}(\varphi) = \mathrm{Var}(\psi) \cup \mathrm{Var}(\chi)$, $\mathrm{Bound}(\varphi) = \mathrm{Bound}(\psi) \cup \mathrm{Bound}(\chi)$, and $\mathrm{Free}(\varphi) = \mathrm{Free}(\psi) \cup \mathrm{Free}(\chi)$.
- If φ is $\forall x\psi$ or $\exists x\psi$, then $\mathrm{Var}(\varphi) = \mathrm{Var}(\psi) \cup \{x\}$, $\mathrm{Bound}(\varphi) = \mathrm{Bound}(\psi) \cup \{x\}$, and $\mathrm{Free}(\varphi) = \mathrm{Free}(\psi) \setminus \{x\}$.

That is, every variable occurs freely *unless* it is under the scope of a quantifier binding the same variable. An easy method of determining which variables are bound by which quantifiers is given by the parse trees. A quantifier binds all occurrences of its variable which occur above it in the tree that are not already bound by another quantifier.

Example 8.2.13. Let $\varphi = \forall x \exists y Rxyz \wedge \exists x Rxyy$. Then

$$\mathrm{Var}(\varphi) = \mathrm{Var}(\forall x \exists y Rxyz) \cup \mathrm{Var}(\exists x Rxyy)$$

Taking each subset individually, the first is:

$$\begin{aligned} \mathrm{Var}(\forall x \exists y Rxyz) &= \mathrm{Var}(\exists y Rxyz \cup \{x\}) \\ &= \mathrm{Var}(Rxyz) \cup \{y\} \cup \{x\} \\ &= \{x, y, z\} \cup \{y\} \cup \{x\} \\ &= \{x, y, z\} \end{aligned}$$

and the second is:

$$\begin{aligned}\mathrm{Var}(\exists x Rxyy) &= \mathrm{Var}(Rxyy) \cup \{x\} \\ &= \{x, y\} \cup \{x\} \\ &= \{x, y\}\end{aligned}$$

Thus, $\mathrm{Var}(\varphi) = \{x, y, z\} \cup \{x, y\} = \{x, y, z\}$.

The free occurrences of variables in this wff are in **bold**:

$$\forall x \exists y Rxy\mathbf{z} \wedge \exists x Rx\mathbf{yy}$$

Example 8.2.14. If φ is the formula $\forall x \exists y Rxyz \wedge \exists z \forall y Ryzx$, then $\mathrm{Var}(\varphi) = \{x, y, z\}$, $\mathrm{Free} = \{x, z\}$, and $\mathrm{Bound}(\varphi) = \{x, y, z\}$.

Lemma 8.2.15. *For every formula φ,* $\mathrm{Var}(\varphi) = \mathrm{Bound}(\varphi) \cup \mathrm{Free}(\varphi)$. *That is, every variable occurs either bound or free (or both).*

Proof. By induction on the complexity of φ.

Basis case: φ is an atomic formula. Then by definition $\mathrm{Bound}(\varphi) = \varnothing$, so $\mathrm{Bound}(\varphi) \cup \mathrm{Free}(\varphi) = \mathrm{Free}(\varphi)$. By definition, $\mathrm{Free}(\varphi) = \mathrm{Var}(\varphi)$, so it follows that $\mathrm{Var}(\varphi) = \mathrm{Bound}(\varphi) \cup \mathrm{Free}(\varphi)$.

Inductive step: Assume that $\mathrm{Var}(\psi) = \mathrm{Bound}(\psi) \cup \mathrm{Free}(\psi)$ and $\mathrm{Var}(\chi) = \mathrm{Bound}(\chi) \cup \mathrm{Free}(\chi)$. We leave the cases where $\varphi = \neg\psi$, $\varphi = \psi \wedge \chi$, $\varphi = \psi \vee \chi$, $\varphi = \psi \to \chi$, and $\exists x \psi$ as exercises for the reader and only prove the case where $\varphi = \forall x \psi$.

$$\begin{aligned}\mathrm{Var}(\varphi) &= \mathrm{Var}(\psi) \cup \{x\} && \text{(by def.)} \\ &= \mathrm{Bound}(\psi) \cup \mathrm{Free}(\psi) \cup \{x\} && \text{(by ind. ass.)} \\ &= \mathrm{Bound}(\psi) \cup \{x\} \cup \mathrm{Free}(\psi) && \text{(order doesn't matter)} \\ &= \mathrm{Bound}(\varphi) \cup \mathrm{Free}(\psi) && \text{(by def.)}\end{aligned}$$

Now, $\mathrm{Free}(\varphi) \subseteq \mathrm{Free}(\psi)$. When $x \notin \mathrm{Free}(\psi)$, $\mathrm{Free}(\varphi) = \mathrm{Free}(\psi)$, by definition, and hence

$$\mathrm{Var}(\varphi) = \mathrm{Bound}(\varphi) \cup \mathrm{Free}(\varphi)$$

When $x \in \mathrm{Free}(\psi)$, $x \notin \mathrm{Free}(\varphi)$, but rather $x \in \mathrm{Bound}(\varphi)$. So all that has happened is that x has moved from being free to being bound, and nothing else has changed. This gives us the desired result. □

Definition 8.2.16 (Sentence)**.** A wff with no free (occurrences of) variables is called a *sentence* or a *closed formula*.

We now identify a concept of 'substitution' (defining which will involve the notions of free and bound occurrences of variables that we have just defined). The general idea is that if we have a formula $\varphi(x)$ where x possibly occurs freely in φ, then we can define another formula $\varphi(\tau/x)$, which is the result of replacing all free occurrences of x by some term τ in φ. Informally, if $\varphi(x) \coloneqq \exists y Rxy$ then $\varphi(\mathsf{a}/x) \coloneqq \exists y R\mathsf{a}y$, and if $\varphi(x) \coloneqq \exists x Rxy$ then $\varphi(\mathsf{a}/x) \coloneqq \exists x Rxy$.

We begin with terms. Suppose we have a term τ in which some term τ_1 occurs. If we have another term τ_2, then we can construct a new term by replacing all occurrences of τ_1 in τ with τ_2; we say that the new term constructed is a *substitution instance* of the first. For example, let $\tau = f(x, y)$ and $\tau_2 = \mathsf{c}$. Then $f(x, \mathsf{c})$ and $f(\mathsf{c}, y)$ are both substitution instances of τ; in the first, $\tau_1 = y$ and in the second $\tau_1 = x$. More formally:

Definition 8.2.17 (Substitution (terms))**.** Let τ and τ_1 be terms and x a variable. We define $\tau(\tau_1/x)$, *the substitution of* τ_1 *for* x *in* τ, recursively as follows:

- If τ is a constant, then $\tau(\tau_1/x) = \tau$.
- If τ is variable, then $\tau(\tau_1/x) = \tau$ if $x \neq \tau$, and $\tau(\tau_1/x) = \tau_1$ if $x = \tau$.
- If $\tau = f(\sigma_1, \ldots, \sigma_n)$, then $\tau(\tau_1/x) = f(\sigma_1(\tau_1/x), \ldots, \sigma_n(\tau_1/x))$.

We can then define substitution instances of formulas.

Definition 8.2.18 (Substitution (formulas))**.** Let φ be a formula, x a variable, and τ a term. We define $\varphi(\tau/x)$, the *substitution of* τ *for* x *in* φ, recursively as follows:

- If φ is $R\tau_1, \ldots, \tau_n$, then $\varphi(\tau/x) = R\tau_1(\tau/x), \ldots, \tau_n(\tau/x)$, and similarly if φ is $\tau_1 = \tau_2$.
- If φ is $\neg\psi$, then $\varphi(\tau/x)$ is $\neg\psi(\tau/x)$.
- If φ is $\psi \wedge \chi$, $\psi \vee \chi$, or $\psi \to \chi$, then $\varphi(\tau/x)$ is $\psi(\tau/x) \wedge \chi(\tau/x)$, $\psi(\tau/x) \vee \chi(\tau/x)$, or $\psi(\tau/x) \to \chi(\tau/x)$, respectively.
- If φ is $\forall y\psi$ or $\exists y\psi$, where $y \neq x$, then $\varphi(\tau/x)$ is $\forall y\psi(\tau/x)$ or $\exists y\psi(\tau/x)$, respectively.
- If φ is $\forall x\psi$ or $\exists x\psi$, then $\varphi(\tau/x) = \varphi$.

The basic rule of thumb is that substitution can only be performed on free variables: To substitute τ for x in a formula φ is to erase all free occurrences of x from φ and replace them with τ. When no confusion will result, we will omit x from $\varphi(\tau/x)$ and simply write $\varphi(\tau)$. In each case, we substitute the term τ in for *every* free occurrence of variable x. Thus, the relationship 'is a substitution instance of' is not symmetric:

Example 8.2.19. $\forall x(Px \wedge Fx)$ is a substitution instance of $\forall x(Px \wedge Fy)$, but not vice versa.

Further examples of formulas that are, and are not, substitution instances of other formulas are given in Table 8.1.

Formula	**Substitution instances**	**Not substitution instances**
$x = x$	$y = y$ $\mathsf{a} = \mathsf{a}$	$\mathsf{a} = x$ $\mathsf{a} = \mathsf{b}$
$x = y \rightarrow y = x$	$x = z \rightarrow z = x$ $y = y \rightarrow y = y$	$x = y \rightarrow x = x$ $x = y \rightarrow z = x$
$\forall x(x = y \rightarrow y = x)$	$\forall x(x = z \rightarrow z = x)$ $\forall x(x = \mathsf{a} \rightarrow \mathsf{a} = x)$ $\forall x(x = x \rightarrow x = x)$	$\forall x(x = y \rightarrow x = x)$ $\forall x(\mathsf{a} = y \rightarrow y = \mathsf{a})$

Table 8.1: Examples of substitution instances

Not all substitutions are created equal, however; in Example 8.2.19 we saw a substitution wherein a variable that occurred freely in the original formula ended up bounded in the substitution. While syntactically, such substitution is allowed, it will turn out that such substitutions are not necessarily truth-preserving. When we substitute terms in for variables, we want to ensure that by doing so we don't end up with any previously free occurrences of variables becoming bound. We therefore specify when it is allowed to substitute a term in for a variable:

Definition 8.2.20 (Allowed substitution)**.** If φ is a formula, τ is a term, and x a variable of $\mathcal{L}$, then τ can be substituted for x in φ in the following cases:

- If φ is atomic, then τ may be substituted in for x in φ.
- If φ is $\neg\psi$, then τ may be substituted in for x in φ if it can be substituted in for x in ψ.
- If φ is $\psi \wedge \chi$, $\psi \vee \chi$, or $\psi \rightarrow \chi$, then τ may be substituted in for x in φ if it can be substituted in for x in both ψ and χ.

- If φ is $\forall y\psi$ or $\exists y\psi$, then τ can be substituted in for x in φ if either

  1. $x \notin \text{Free}(\varphi)$, or
  2. τ can be substituted in for x in ψ and $y \notin \text{Var}(\tau)$.

The basic idea here is that if a substitution results in certain occurrences of variables being bound that were free in the original formula, the substitution is not allowable.

All of the substitution instances in Table 8.1 are allowed. We give an example of a substitution that is not allowed.

Example 8.2.21. Let φ be $\forall y(x \neq y)$. $\varphi(y/x)$ is not allowed. This is because:

1. φ is of the form $\forall y\psi$;
2. $x \in \text{Free}(\varphi)$;
3. $y \in \text{Var}(\psi)$.

If we *did* substitute y in for x, the result, $\forall y(y \neq y)$ would be necessarily false.

We can also define the syntactic substitution of constants for constants, and variables for constants, by changing the first two bullet points of Definition 8.2.17. Let τ and τ_1 be terms, and c a constant. Then:

- If τ is a variable, then $\tau(\tau_1/\mathsf{c}) = \tau$.
- If τ is a constant, then $\tau(\tau_1/\mathsf{c}) = \tau$ if $\mathsf{c} \neq \tau$, and $\tau(\tau_1/\mathsf{c}) = \tau_1$ if $\tau = \mathsf{c}$.

In general, we will only be allowed to substitute co-referring constants (e.g., we can only substitute c' for c when $\mathsf{c}' = \mathsf{c}$), but this will fall out of our proof rules.

8.2.1 Translating from logic to English and back

From a logical language into English

In a given quantified lanuage $\mathcal{L}_q$ the constants, function symbols, and relation symbols will correspond to certain English words or phrases. The interpretations of these symbols are often given informally, e.g., "R is the unary predicate 'is red'." The propositional connectives retain the same interpretation that they previously had—$\wedge$ is conjunction, $\vee$ is disjunction, etc.—thus most of the difficulty comes from finding idiomatic ways of rendering the quantifiers.

The universal quantifier $\forall x$ corresponds to the English locutions 'for all', 'every', 'all', 'any', and the like. The existential quantifier $\exists x$ corresponds to the English locutions 'there is', 'some', 'exists', 'there exists', and similar. For example, $\forall x Px$ is read "Every x is P", and $\exists y \exists z Ryz$ is read "There is a y and a z such that y is related to z by R". If P stands for the predicate 'is a pear', and R stands for the relation 'runs from', then these two formulas can be read more idiomatically as "Everything is a pear" and "Something runs from something." (Note that it is not required *by the syntax alone* that y and z pick out distinct objects!) Here we can see that bound variables are easy to translate: They are already taken care of by the quantifiers themselves. Free variables, on the other hand, are generally more difficult to render neatly in English. Because they lack a quantifier, we don't want our translations to make it seem like there is a quantifier, implicitly or explicitly. As a rule of thumb, you can think of a free variable x as simply representing 'a thing'. (Hence, free variables are how we can represent sentences with indefinite subjects, such as "A man runs.")

In general, translating from logic into English can be done in a two-step process: In the first step, one literally reads the formula off, substituting in English phrases for their symbolic representations. Then, in the second step, the literal English translation is modified into something that sounds more natural. (The second step is not always possible.)

Example 8.2.22. Let $\mathcal{L}'_q$ be a quantified language containing no constants or function symbols, and containing the unary predicate symbols P 'is a pear', F 'is a fruit', R 'is red', T 'is a table', and the binary relation symbol O 'is on'. With this language we can say things like the following:

1. Px "A thing is a pear."

2. $Px \wedge Rx$ "A thing is a pear and it is red." (Note that here x has to be interpreted as the same thing in both conjuncts.)

3. $\exists x(Px \wedge Rx)$ "There exists an x such that x is a red pear," or "There is a red pear," or "Something is a red pear."

4. $\exists x Tx$ "Something is a table."

5. $Tx \wedge Py$ "A thing is a table and a thing is a pear." (Note again that there is no guarantee that x and y are different things.)

6. $Tx \wedge Py \wedge x \neq y$ "A thing is a table and another thing is a pear." (Here we do know that x and y are not the same thing.)

7. $\exists x \exists y(Tx \wedge Py \wedge Oyx)$ "There is a pear on some table."

8. $\exists x Fx$ "Something is a fruit."

9. $\forall x Fx$ "Everything is a fruit."

10. $\forall x(Px \rightarrow Fx)$ "Every pear is a fruit."

11. $\forall x(\neg Px \vee Fx)$ "Everything is either not a pear or is a fruit."

12. $\forall x(\neg(Px \vee Fx))$ "Everything is not a pear or a fruit."

Note 8.2.23. Order matters: It is one thing to say $\exists x \exists y(Tx \wedge Py \wedge Oyx)$ "There is a pear on some table," and quite another thing to say $\exists x \exists y(Tx \wedge Py \wedge Oxy)$ "There is a table on a some pear."

These symbolisations are all straightforward with perhaps the exception of number 10: What would be a universal affirmative predication in the syllogistic is rendered in predicate logic as a universally quantified implication. (We will see in §8.3 some consequences of this definition of universal predication that some people have found counterintuitive.)

From English into a logical language

When translating from English into symbolic form, we generally must first construct the language that we are translating in to. That is, we must identify the functions, constants, predicates, and relations that occur in the English sentences and identify appropriate symbols to represent each of them.

It can be difficult, sometimes, to determine whether a certain English word or phrase corresponds to a predicate or to a relation. For example, does the sentence "John is the brother of Alice" contain a unary predicate 'brother of Alice' or a binary relation 'brother of'? Does the phrase "the red chair" pick out a single predicate, or does it pick out a conjunction of two predicates, 'red' and 'chair'? If context makes it clear that there is only one person who is a client, is "the client" a constant, or a predicate?

In general, the best rule of thumb to use is to try to make the language as generally applicable as possible. The predicate 'brother of Alice' applies to fewer objects than the relation 'brother of' does; the former is less general. Similarly, even if in a particular context there is only one client, 'client' is something that can potentially apply to many people, and thus it is preferable to treat it as a predicate rather than as a constant. On the other hand, a proper name such as 'Alice' will only ever pick out a single individual; it may do so ambiguously, if there is more than one person named Alice under consideration, but only

in very rare cases does it make sense to use 'Alice' in a collective or distributive sense (for example, "There are as many girls as there are Jameses in my daughter's class.")

In English, we do not generally make simple bare existence statements like "There is a pear" or "Something is red", or genuine universal statements like "Everything is a pear" or "Everything is running". More often, we are making predications which link multiple properties to the same object or objects. For example, instead of saying "There is a pear", we might say "There is a pear on the table", or "The pear is red". Alternatively, instead of saying "Everything is a fruit", we might say "All pears are fruit". Sentences of the former type—predicating something of a single thing of a specific kind—are symbolised by the existential quantifier and a conjunction of the two relevant predicates or relations. Sentences of the latter type—saying that everything of a specific kind is also of another kind—are represented via universally quantified implications.

The best way to understand this, why $\forall x(Px \to Fx)$ is the right way to represent "All pears are fruit" is to think intuitively about what makes "All pears are fruit" true—or, more simply, what could make it false. The only way that "All pears are fruit" is false is if there is some pear that is not a fruit—that is, if there is something for which the antecedent of the conditional is true and the consequent is false. From this we can see that any if there isn't anything which is both a pear and not a fruit, then everything that is a pear is a fruit.

Example 8.2.24. Consider the following text:

> Julian: "Are you going to the party tomorrow? Everyone else is."
>
> Bethany: "That's not true. Alice will only go if Max's sister does. But if either Nick or Osman go, Max won't go, and Max's sister won't go without him. So either Alice won't be there or Nick and Osman won't, so not everyone else is going.
>
> Julian: "Okay, so not everyone else is. Who cares. Are *you* going?"
>
> Bethany: "No, I'm having too much fun doing my logic homework."

First, identify the constants: Alice, Bethany, Max, Nick, Osman. We'll represent these with a, b, m, n, and o.

Next, identify any functions. Recall that functions take as input objects and give back *other objects* (as opposted to sentences). We have one function: 'sister of', which we'll represent with s. This is a unary function, taking one constant or variable as input.

Third, identify any predicates or relations. Here, there is only one characteristic or property that is under discussion, namely the property of 'going to the party tomorrow'. We have quite a bit of flexiblity as to how we represent this. We can take 'going to the party

tomorrow' as a complex unary predicate that applies to single people. Or, we could identify a binary relation 'going to tomorrow', which relates people to parties (in which case, we *also* need to have unary predicates 'person' and 'party'). Finally, we could understand it instead a *ternary* relation, of 'x is going to y at time t' (in which case we'd have to introduce another unary predicate, 'time').

The final option, requiring us to grapple with 'tomorrow', introduces a level of complexity due to the fact that 'tomorrow' is an indexical. We'll discuss how indexicals can be accounted for in another chapter. For now, for simplicity's sake, we will take the first option. For this, we simply need a single relation symbol, G 'is going to the party tomorrow'. Now we can translate the indicative sentences; the sentences and their symbolisations are given in Table 8.2.

English	**Symbolisation**
Everyone else is going to the party	$\forall x(x \neq \mathsf{b} \to Gx)$
Not everyone else is going to the party	$\neg\forall x(x \neq \mathsf{b} \to Gx)$
Alice will only go if Max's sister does	$G\mathsf{a} \to Gs(\mathsf{m})$
If either Nick or Osman go, Max won't	$(G\mathsf{n} \vee G\mathsf{o}) \to \neg G\mathsf{m}$
Max's sister won't go without him	$Gs(\mathsf{m}) \to G\mathsf{m}$
Either Alice won't be there or Nick and Osman won't	$\neg G\mathsf{a} \vee (\neg G\mathsf{n} \wedge \neg G\mathsf{o})$
I'm not going	$\neg G\mathsf{b}$

Table 8.2: Some example translations

Example 8.2.25. Consider the set of shapes in Figure 8.3, and all the possible true things that could be said of these shapes, including:

1. There is a red square.
2. A blue square is between an orange square and a purple square.
3. There are no green shapes.
4. The orange triangle is not a square.
5. No triangle is a square.
6. The purple square is bigger than the orange square.
7. Every shape exists in red.
8. There are no purple circles.

9. Only red things are circles.

10. All blue things are squares.

11. Some triangle is the same color as some circle.

12. There are two distinct triangles.

13. Some shapes are bigger than others.

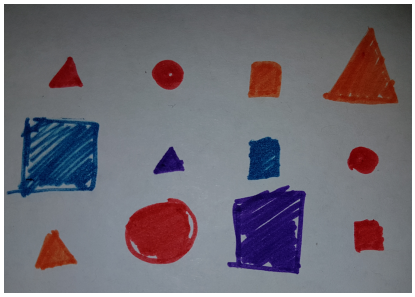


Figure 8.3: Domain for Example 8.2.25

In these sentences, there are no constants and no functions, therefore we only need to identify any relevant predicates and relations. There are five colors: R, B, G, O, and P. There are three shapes: T, S, C. There is also the general property of 'being a shape', which we will represent by F (for 'figure', since we've already used S for 'square'). And then we have one binary relation $>$ ('bigger than') and one ternary relation W ('between', since we've already used B for 'blue'). Using this language, these sentences are symbolised as follows:

1. $\exists x(Rx \wedge Sx)$

2. $\exists x \exists y \exists z((Bx \wedge Sz) \wedge (Oy \wedge Sy) \wedge (Pz \wedge Sz) \wedge Wxyz)$.

3. $\neg \exists x(Fx \wedge Gx)$

4. $\exists x(Ox \wedge Tx \wedge \neg Sx)$

5. $\forall x(Tx \rightarrow \neg Sx)$

6. $\exists x \exists y((Px \wedge Sx) \wedge (Oy \wedge Sy) \wedge x > y)$

7. $\exists x(Rx \wedge Sx) \wedge \exists y(Ry \wedge Ty) \wedge \exists z(Rz \wedge Cz)$

8. $\neg\exists x(Px \wedge Cx)$

9. $\forall x(Cx \to Rx)$

10. $\forall x(Bx \to Sx)$

11. $\exists x\exists y(Tx \wedge Cy \wedge ((Rx \wedge Ry) \vee (Ox \wedge Oy) \vee (Bx \wedge By) \vee (Px \wedge Py))$.

12. $\exists x\exists y(Tx \wedge Ty \wedge x \neq y)$.

13. $\exists x\exists y(Fx \wedge Fy \wedge x > y)$

Given what we said in the previous section, we could also render 3 and 8 as $\forall x(Fx \to \neg Gx)$ and $\forall x(Px \to \neg Cx)$, respectively.

Example 8.2.26. Suppose there are three objects, a red circle, a four-legged cat, and a brown chair. If we are interested in chairs, four-legged things, and red things, we can define a language containing three unary predicates, Cx 'is a chair', Rx 'is red', and Fx 'is four-legged'. Then, there are three different ways we can say "There is a red thing and a chair and a four-legged thing":

1. $\exists xRx \wedge \exists xCx \wedge \exists xFx$

2. $\exists xRx \wedge \exists yCy \wedge \exists zFz$

3. $\exists x\exists y\exists z(Rx \wedge Cy \wedge Fz)$

(We show that these are all equivalent in the next section, after we've introduced the semantics, cf. Example **??**). But none of these three are equivalent to

4. $\exists x(Rx \wedge Cx \wedge Fx)$

because this would require there to be *one* object that is a red, four-legged chair.

8.3 Semantics

In this section, we explain how to give meaning to a quantified formula, and hence evaluate whether or not it is true.

Definition 8.3.1 (Quantified models)**.** A quantified model is a pair $\mathfrak{M} = \langle D, I\rangle$ where:

- D is a non-empty set of objects (called the 'domain', the 'ontology', or the 'universe').[2]
- I is an interpretation function, such that
  - For every constant c, $I(\mathsf{c}) \in D$.
  - For every n-place function f^n, $I(f^n)$ is an n-ary function on D.
  - For every n-place predicate R^n, $I(R^n) \subseteq D^n$; that is, $I(R^n)$ is a set of n-tuples of D that R applies to.

Every model comes equipped with a set of variable assignments.

Definition 8.3.2 (Variable assignment)**.** v is a *variable assignment* for a model $\mathfrak{M} = \langle D, I\rangle$ if it is a function from the set of variables in the language to D.

If τ is a variable, then we take $I(\tau)$ to be $v(\tau)$.

Definition 8.3.3 (x-variant)**.** A variable assignment v' is an *x-variant* of a variable assignment v if v' and v perhaps assign x to a different object, but otherwise agree.

Definition 8.3.4 (Truth conditions for quantified formulas)**.** Fix a model $\mathfrak{M} = \langle D, I\rangle$ and a variable assignment v. The notion of truth in a model is defined recursively.

Atomic formulas

$$\begin{array}{lll} \mathfrak{M}, v \vDash R\tau_1, \ldots, \tau_n & \text{iff} & \langle I(\tau_1), \ldots, I(\tau_n)\rangle \in I(R) \\ \mathfrak{M}, v \vDash \tau_1 = \tau_2 & \text{iff} & I(\tau_1) = I(\tau_2) \end{array}$$

Propositional connectives

$$\begin{array}{lll} \mathfrak{M}, v \vDash \neg\varphi & \text{iff} & \mathfrak{M}, v \nvDash \varphi \\ \mathfrak{M}, v \vDash \varphi \wedge \psi & \text{iff} & \mathfrak{M}, v \vDash \varphi \text{ and } \mathfrak{M}, v \vDash \psi \\ \mathfrak{M}, v \vDash \varphi \vee \psi & \text{iff} & \mathfrak{M}, v \vDash \varphi \text{ or } \mathfrak{M}, v \vDash \psi \\ \mathfrak{M}, v \vDash \varphi \to \psi & \text{iff} & \mathfrak{M}, v \nvDash \varphi \text{ or } \mathfrak{M}, v \vDash \psi \end{array}$$

Quantified formulas

$$\begin{array}{lll} \mathfrak{M}, v \vDash \forall x\varphi & \text{iff} & \text{for every } x\text{-variant } v', \mathfrak{M}, v' \vDash \varphi \\ \mathfrak{M}, v \vDash \exists x\varphi & \text{iff} & \text{there is an } x\text{-variant } v', \text{ such that } \mathfrak{M}, v' \vDash \varphi \end{array}$$

[2] As a rule, logic is silent on ontological matters: Logic alone doesn't tell us what exists. The requirement that domains be non-empty comes with ontological commitment: With this requirement, $\exists x(x = x)$ is a tautology. It is important to note that this is an external requirement that is added to domains because empty domains do not behave in a nice fashion, and lead to a number of pathological properties. We adopt this requirement because it is convenient, not because it is required.

If φ is true in $\mathfrak{M}$ under any variable assignment, we write $\mathfrak{M} \vDash \varphi$, read '$\mathfrak{M}$ models φ' or '$\mathfrak{M}$ is a model of φ'. If *any* model is a model of φ, then we write $\vDash \varphi$, and call φ a tautology or say that it is logically valid.

Corollary 8.3.5. *First-order logic as defined here satisfies the Principle of Non-Contradiction and the Law of Excluded Middle. That is, for every model $\mathfrak{M}$, variable assigment v, and formula φ, either $\mathfrak{M}, v \vDash \varphi$ or $\mathfrak{M}, v \vDash \neg\varphi$, and for every model $\mathfrak{M}$, variable assigment v, and formula φ, not both $\mathfrak{M}, v \vDash \varphi$ and $\mathfrak{M}, v \vDash \neg\varphi$.*

Proof. Exercise for the reader. □

Lemma 8.3.6. *For every model $\mathfrak{M}$ and variable assignment v:*

$$\begin{array}{lll} (i) & \mathfrak{M}, v \vDash \forall x\varphi & \textit{iff} \quad \mathfrak{M}, v \vDash \neg\exists x\neg\varphi \\ (ii) & \mathfrak{M}, v \vDash \exists y\varphi & \textit{iff} \quad \mathfrak{M}, v \vDash \neg\forall y\neg\varphi \end{array}$$

Proof. (i) Fix a model $\mathfrak{M}$ and variable assignment v. Then:

$$\begin{array}{lll} \mathfrak{M}, v \vDash \forall x\varphi & \text{iff} & \text{for every } x\text{-variant } v', \mathfrak{M}, v' \vDash \varphi \\ & \text{iff} & \text{there is no } x\text{-variant } v', \text{ such that } \mathfrak{M}, v' \nvDash \varphi \\ & \text{iff} & \text{there is no } x\text{-variant } v', \text{ such that } \mathfrak{M}, v' \vDash \neg\varphi \\ & \text{iff} & \mathfrak{M}, v \nvDash \exists x\neg\varphi \\ & \text{iff} & \mathfrak{M}, v \vDash \neg\exists x\neg\varphi \end{array}$$

(ii) Exercise for the reader. □

Lemma 8.3.7. *Fix a first-order language containing two unary predicates U and H. Let $\mathfrak{M} = \langle D, I\rangle$ where $D = \{Smaug, Norbert, Puff\}$ and $I(U) = \varnothing$, $I(H) = \{Smaug\}$, and v be an arbitrary variable assignment. Letting U be the predicate 'is a unicorn' and H the predicate 'is a hoarder', on this model, both "Every unicorn is a hoarder" is true and "Every unicorn is not a hoarder". That is, $\mathfrak{M}, v \vDash \forall x(Ux \to Hx)$ and $\mathfrak{M}, v \vDash \forall x(Ux \to \neg Hx)$.*

Proof. We first show that $\mathfrak{M}, v \vDash \forall x(Ux \to Hx)$.

$$\begin{array}{lll} \mathfrak{M}, v \vDash \forall x(Ux \to Hx) & \text{iff} & \text{for every } x\text{-variant } v', \mathfrak{M}, v' \vDash Ux \to Hx \\ & \text{iff} & \text{for every } x\text{-variant } v', \mathfrak{M}, v' \nvDash Ux \text{ or } \mathfrak{M}, v' \vDash Hx \end{array}$$

Because $I(U) = \varnothing$, whatever $v'(x)$ is, $v'(x) \notin I(U)$, and hence $\mathfrak{M}, v' \nvDash Ux$, making the first disjunct true.

The case for $\mathfrak{M}, v \vDash \forall x(Ux \to \neg Hx)$ is analogous. □

Thus, whenever there are no unicorns, all unicorns are both hoarders and not hoarders; in fact, any pair of contradictory properties can be ascribed to things which do not exist. This property of predicate logic is one way in which it is distinguished from the syllogistic (where "Every unicorn is a hoarder" implies that there is at least one unicorn).

Lemma 8.3.8. *Fix a first-order language containing one constant* c *and one binary relation* R*. Let* $\mathfrak{M} = \langle D, I\rangle$ *where* $D = \{1, 2\}$*,* $I(\mathsf{c}) = 2$*, and* $I(R) = \{\langle 1, 1\rangle, \langle 1, 2\rangle, \langle 2, 1\rangle\}$*. Let* $v(x) = 1$*,* $v(y) = 2$*, and for any variable* z *distinct from* x *and* y*, let* $v(z) = 1$*. Then,* $\mathfrak{M}, v \nvDash \forall x \exists y Rxy \to \forall x Rx\mathsf{c}$*.*

Proof.

$$\mathfrak{M}, v \vDash \forall x \exists y Rxy \to \forall x Rx\mathsf{c} \quad \text{iff} \quad \begin{array}{l}\text{either } \mathfrak{M}, v \nvDash \forall x \exists y Rxy \\ \text{or } \mathfrak{M}, v \vDash \forall x Rx\mathsf{c}\end{array}$$

The second case is simpler, so we consider it first:

$$\begin{array}{lll}\mathfrak{M}, v \vDash \forall x Rx\mathsf{c} & \text{iff} & \text{for every } x\text{-variant } v', \mathfrak{M}, v' \vDash Rx\mathsf{c} \\ & \text{iff} & \langle 1, I(\mathsf{c})\rangle \in I(R) \text{ and } \langle 2, I(\mathsf{c})\rangle \in I(R) \\ & \text{iff} & \langle 1, 2\rangle \in I(R) \text{ and } \langle 2, 2\rangle \in I(R)\end{array}$$

But $\langle 2, 2, \rangle \notin I(R)$, so the second case is false.

We now consider the first case.

$$\begin{array}{lll}\mathfrak{M}, v \nvDash \forall x \exists y Rxy & \text{iff} & \mathfrak{M}, v \vDash \neg \forall x \exists y Rxy \\ & \text{iff} & \mathfrak{M}, v \vDash \exists x \forall y \neg Rxy \\ & \text{iff} & \text{there is an } x\text{-variant } v' \text{s.t. } \mathfrak{M}, v' \vDash \forall y \neg Rxy \\ & \text{iff} & \text{for every } y\text{-variant } v'', \mathfrak{M}, v'' \vDash \neg Rxy\end{array}$$

That is, is there some object in the domain that is not related to anything in the domain by R? If $v'(x) = 1$, then every y-variant v'' makes Rxy true. If $v'(x) = 2$, then the y-variant that assigns y to 1 makes Rxy true.

Thus, this formula is false on this model. □

Corollary 8.3.9. *If the model is modified so that* $I(\mathsf{c}) = 1$*, then the formula is true on the model.*

Proof. Exercise for the reader. □

Lemma 8.3.10. *Fix a first-order language containing a unary relation* P*, a binary relation* R*, and a unary function* f*. Let* $\mathfrak{M} = \langle D, I\rangle$ *where*

- $D = \{\square, \square, \circ, \boxslash, \boxslash\}$.

- *$I(f)$ is such that $f(\boxslash) = \square$, $f(\boxslash) = \square$, and for all other inputs it equals $\circ$.*
- *$I(P) = \{\square, \square, \boxslash, \boxslash\}$.*
- *$I(R) = \{\langle\square, \square\rangle, \langle\square, \circ\rangle, \langle\circ, \boxslash\rangle, \langle\boxslash, \boxslash\rangle\}$.*

Let $v(x) = \square$, $v(y) = \circ$, $v(z) = \circ$, and for all other w, $v(w) = \square$.
Then

1. $\mathfrak{M}, v \nvDash \forall x \exists y (Rxy)$.
2. $\mathfrak{M}, v \vDash \neg \forall x (Px)$
3. $\mathfrak{M}, v \vDash \forall x (\neg Px \to \exists y (Rxy))$

Proof.

1.

1	$\mathfrak{M}, v \vDash \forall x \exists y (Rxy)$	iff	for every x-variant $v', \mathfrak{M}, v' \vDash \exists y (Rxy)$

Consider the x-variant v' where $v'(x) = \boxslash$. Then

2	$\mathfrak{M}, v' \vDash \exists y (Rxy)$	iff	there is a y-variant v'' s.t. $\mathfrak{M}, v'' \vDash Rxy$
3		iff	$\langle v''(x), v''(y)\rangle \in I(R)$
4		iff	$\langle\boxslash, v''(y)\rangle \in I(R)$

But inspection of $I(R)$ shows us that there is no pair in the set with $\boxslash$ as the first element. Thus, since the right-hand side of (4) is false, the right-hand side of (3) is also false, and so on all the way up, until we can conclude that the left-hand side of (1) is false. If $\mathfrak{M}, v \vDash \forall x \exists y (Rxy)$ is false, then $\mathfrak{M}, v \nvDash \forall x \exists y (Rxy)$, as required.

2.

1	$\mathfrak{M}, v \vDash \neg \forall x (Px)$	iff	$\mathfrak{M}, v \nvDash \forall x (Px)$
2		iff	it is false that every x-variant v' is s.t. $\mathfrak{M}, v' \vDash Px$
3		iff	there is at least one x-variant v' s.t. $\mathfrak{M}, v' \nvDash Px$

Let v' be an x-variant of v where $v'(x) = \circ$. Then $v'(x) \notin I(P)$, and hence $\mathfrak{M}, v' \nvDash Px$, as required.

3.

1	$\mathfrak{M}, v \vDash \forall x (\neg Px \to \exists y (Rxy))$	iff	for every x-variant $v', \mathfrak{M}, v' \vDash (\neg Px \to \exists y (Rxy))$
2		iff	either (a) $\mathfrak{M}, v' \nvDash \neg Px$
			or (b) $\mathfrak{M}, v' \vDash \exists y (Rxy)$

We have five cases to consider: $v_1(x) = \square$, $v_2(x) = \square$, $v_3(x) = \circ$, $v_4(x) = \boxslash$, and $v_5(x) = \boxslash$. (It is clear from inspection of D that these are the only possible x-variants of v.) Let us consider v_i where $i \in \{1, 2, 4, 5\}$. In each case, $v_i(x) \in I(P)$, and hence $\mathfrak{M}, v_i \vDash Px$. By Corollary 8.3.5, $\mathfrak{M}, v_i \nvDash \neg Px$, and hence (2a) is satisfied. This leaves us with v_3:

2b $\mathfrak{M}, v_3 \vDash \exists y(Rxy)$ iff there is a y-variant v_3' s.t. $\mathfrak{M}, v_3' \vDash Rxy$

To determine whether there is such a y-variant, we simply need to look at $I(R)$, and in particular at any pairs with $\circ$ as the first element of the pair. And indeed, there is one, namely, the pair $\langle \circ, \boxslash \rangle$. Let us then set v_3' to be such that $v_3'(y) = \boxslash$. Then, $\langle v_3'(x), v_3'(y) \rangle \in I(R)$, and thus $\mathfrak{M}, v_3' \vDash Rxy$, as required.

□

Lemma 8.3.11. *Fix a first-order language containing a constant* $\mathtt{0}$*, a unary function* $\mathtt{S}$*, a binary function* $\mathtt{Add}$*, and a binary relation* $<$*. Let* $\mathfrak{M} = \langle D, I \rangle$ *where:*

- D = *the set of natural numbers (e.g.,* $\{1, 2, 3, 4, \dots\}$*).*
- $I(\mathtt{0}) = 0$.
- $\langle n, m \rangle \in I(<)$ *if* n *is a natural number that is strictly smaller than* m.
- $I(\mathtt{S}) = n \mapsto n + 1$
- $I(\mathtt{Add}) = \{n, m\} \mapsto n + m$

(Compare Definition 10.2.2.) Let v *be a variable assignment assigning every variable to* 3*. The following are true on this model:*

1. $\forall x(x < \mathtt{S}(x))$
2. $\forall x \exists y(\mathtt{Add}(x, \mathtt{0}) = y)$
3. $\exists x \exists y(\mathtt{Add}(x, y) \neq \mathtt{0})$
4. $\exists y(y < x)$.

Proof. 1.

1	$\mathfrak{M}, v \vDash \forall x(x < \mathtt{S}(x))$	iff	for every x-variant $v', \mathfrak{M}, v' \vDash x < \mathtt{S}(x)$
2		iff	$\langle I(x), I(\mathtt{S}(x)) \rangle \in I(<)$
3		iff	$\langle I(x), I(x) + 1 \rangle \in I(<)$

Suppose that there were an x-variant v' such that $\langle I(x), I(x)+1\rangle \notin I(<)$. This would imply that there was some number $n \in D$ such that $v'(x) = n$ and $\langle n, n+1\rangle \notin I(<)$. But $\langle n, n+1\rangle \notin I(<)$ only if $n \not< n+1$; but n is strictly less than $n+1$, so $\langle n, n+1\rangle \in I(<)$, which contradicts our assumption. Thus, no matter what x is assigned to by v', $\langle I(x), I(x)+1\rangle \in I(<)$. Thus, $\mathfrak{M}, v \vDash \forall x(x < \mathtt{S}(x))$.

2.

1	$\mathfrak{M}, v \vDash \forall x \exists y(\mathtt{Add}(x, \mathtt{0}) = y)$	iff	for every x-variant v', $\mathfrak{M}, v' \vDash \exists y(\mathtt{Add}(x, \mathtt{0}) = y)$
2		iff	for every x-variant v', there is a y-variant v'' such that $\mathfrak{M}, v'' \vDash (\mathtt{Add}(x, \mathtt{0}) = y)$

Let us pick an arbitrary x variant v' where $v'(x) = n$.

3	$\mathfrak{M}, v'' \vDash (\mathtt{Add}(x, \mathtt{0}) = y)$	iff	$I(\mathtt{Add}(I(x), I(\mathtt{0})) = I(y)$
4		iff	$I(x) + I(\mathtt{0}) = I(y)$
5		iff	$n + 0 = y$

What must we assign to y to $n + 0 = y$? Why, n itself: That is, the same thing that v' assigned to x. Thus, for any x-variant v', we can identify a y-variant v'' that makes the sentence true, by letting $v''(y) = v'(x)$.

3.

1	$\mathfrak{M}, v \vDash \exists x \exists y(\mathtt{Add}(x, y) \neq 0)$	iff	there is an x-variant v' such that $\mathfrak{M}, v' \vDash \exists y(\mathtt{Add}(x, y) \neq 0)$
2		iff	there is an x-variant v' and a y-variant v'' such that $\mathfrak{M}, v'' \vDash \mathtt{Add}(x, y) \neq 0$

That is, we need to find two objects in our domain that we can assign x and y to such that $I(\mathtt{Add}(I(x), I(y)) \neq I(0$, that is, $I(x) + I(y) \neq I(0)$. But this is true in many cases, e.g., $3 + 5 \neq 0$. Thus, let $v''(y) = 5$ and $v''(x) = 3$, and the sentence is satisified.

4.

1	$\mathfrak{M}, v \vDash \exists y(y < x)$	iff	there is a y-variant v' such that $\mathfrak{M}, v' \vDash y < x$
2		iff	$\langle I(y), I(x)\rangle \in I(<)$
3		iff	$\langle I(y), 3\rangle \in I(<)$

Thus, all we need to do is find a number n such that n is strictly less than 3; 2 will do. Let $v'(y) = 2$, and we are done.

□

Example 8.3.12. Here we will prove what we said we'd prove in Example **??**.

Once we have a definition of model and variable assignment, other semantic notions such as consistency and validity can be defined in a straightforward fashion, analogous to the ways they are defined for the syllogistic and propositional logic.

Definition 8.3.13 (Consistency)**.** A wff φ is *consistent* if there is a model $\mathfrak{M}$ and variable assignment v such that $\mathfrak{M}, v \vDash \varphi$.

A set of wffs Γ is *consistent* if there is a model $\mathfrak{M}$ and variable assignment v such that $\mathfrak{M}, v \vDash \varphi$ for every $\varphi \in \Gamma$.

Definition 8.3.14 (Validity)**.** A wff φ is *valid* if for every model $\mathfrak{M}$ and variable assignment v, $\mathfrak{M}, v \vDash \varphi$. If φ is valid, we sometimes write $\vDash \varphi$, dropping reference to model and variable assignment.

An argument whose premises are the set Γ and whose conclusion is a wff φ is *valid* if there is no model $\mathfrak{M}$ and variable assignment v such $\mathfrak{M}, v \vDash \psi$ for every $\psi \in \Gamma$ but $\mathfrak{M}, v \nvDash \varphi$. If the argument is valid, we sometimes write $\Gamma \vDash \varphi$.

8.4 Proof theory: General remarks

There is no effective test for determining whether an arbitrary quantified formula is a tautology; this is because the semantic models we define in §8.3 can have infinitely many objects in them, and thus to determine if a formula is true in a given model we may have to check infinitely many instances. For that reason, we introduce both proof-theory and the semantics, and then show in §8.6.6 that the properties of "being a theorem" and "being a tautology" coincide. Thus, the proof theory provides us with an alternate way to determine whether a quantified formula is a tautology, one which does not involve surveying potentially infinite models.

Definition 8.4.1 (Theory)**.** A *theory* $\mathcal{T}$ in a language $\mathcal{L}_q$ has two components:

- A set of $\mathcal{L}_q$ wffs, called 'axioms'.
- A finite set of relations among wffs, called 'rules of inference'. For each rule of inference, there is some positive number n such that for every set of n wffs and a further wff φ, one can effectively determine whether the rule of inference relates the n wffs to φ. If it does, then φ is a direct consequence of those wffs, according to the rule of inference.

Definition 8.4.2 (Proof)**.** A *proof* in a theory $\mathcal{T}$ is a finite sequence $\varphi_1, \ldots, \varphi_n$ of wffs such that every φ_i is either an axiom or is annotated with one of the rules of inference (that is, it is a direct consequence of some subset of the φ_j, $j < i$, according to the rules of inference of $\mathcal{T}$). If there is a proof of φ from assumptions $\psi_1, \psi_2, \ldots, \psi_n$, we write $\psi_1, \psi_2, \ldots, \psi_n \vdash_{\mathcal{T}} \varphi$. If there is a proof of φ from no assumptions, we write $\vdash_{\mathcal{T}} \varphi$.

When no ambiguity will result, we omit the subscript $\mathcal{T}$.

Definition 8.4.3. If $\vdash_{\mathcal{T}} \varphi$, then we call φ a *theorem of* $\mathcal{T}$.

Note that if $\Gamma \vdash_{\mathcal{T}} \varphi$, φ is *not*, in general, a theorem of $\mathcal{T}$, unless all members of Γ are axioms of $\mathcal{T}$.

8.4.1 Decidability of first-order theories

It is not always possible, given an arbitrary formula φ and a theory $\mathcal{T}$, to determine if there is a $\mathcal{T}$-proof of φ.

Definition 8.4.4 (Decidability)**.** If there is an effective procedure (i.e., an algorithm) for determining theoremhood, then $\mathcal{T}$ is called *decidable*. If there isn't then it is *undecidable*.

A set of axioms is called *decidable* if membership in the set can be determined algorithmically.

8.5 Proof theory: An axiomatic approach

In this section, we introduce the first of two different proof systems for predicate logic. It is an axiomatic approach in which we have many axioms, and few proof rules. The second, introduced in §8.6, is an extension of the propositional proof-system in §7.5, involving no axioms and a host of rules of inference.

The axioms of first-order theories are generally split into the 'logical axioms' and the 'non-logical axioms'. The logical axioms will be shared by any (useful) first-order theory; the non-logical axioms are the ones specific to the language of the theory in question.

Definition 8.5.1 (Logical axioms)**.** If φ, ψ, and χ are wffs of $\mathcal{L}_q$, x is a variable of $\mathcal{L}_q$, and τ is a term of $\mathcal{L}_q$, then the following are the logical axioms of any $\mathcal{L}_q$-theory $\mathcal{T}$:

(A1) $\varphi \to (\psi \to \varphi)$

(A2) $(\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi))$

(A3) $(\neg\psi \to \neg\varphi) \to ((\neg\psi \to \varphi) \to \psi)$

(A4) $\forall x(\varphi \to \psi) \to (\varphi \to \forall x\psi)$ if φ contains no free occurrences of x.

(Dist) $\forall x(\varphi \to \psi) \to (\forall x\varphi \to \forall x\psi)$.

(Sub) $\forall x\varphi \to \varphi(\tau/x)$, if τ may be substituted in for x in φ.

(Gen) $\varphi \to \forall x\varphi$, if x is not free in φ.

Definition 8.5.2 (Equality axioms)**.** If x, y, z are variables (not necessarily distinct) of $\mathcal{L}$, then the equality axioms are:

Reflexivity $x = x$

Symmetry $x = y \to y = x$

Transitivity $(x = y \wedge y = z) \to x = z$

The non-logical axioms will, of course, differ from theory to theory, and cannot be specified generically.

Definition 8.5.3 (Consistency)**.** Given a set S of axioms (logical and non-logical), a set of formulas Λ is S-*inconsistent* iff there is a finite $\Gamma = \{\varphi_0, \ldots, \varphi_n\} \subseteq \Lambda$ such that $\vdash_S \neg(\varphi_0 \wedge \cdots \wedge \varphi_n)$, that is, in the system S is it possible to prove the negation of the conjunction of (a subset of) Λ. If this is not possible, then Λ is S-*consistent*.

Every axiomatic system has as a rule of inference the following:

Rule 8.5.4 (Modus ponens (MP))**.** ψ is a direct consequence of φ and $\varphi \to \psi$.

We also have the following rules for quantifiers:

Rule 8.5.5 ($\forall$ Generalization)**.** Let Γ be a set of formulas and x a variable which is not free in any $\psi \in \Gamma$. If $\Gamma \vdash \varphi$, then $\Gamma \vdash \forall x\varphi$.

Rule 8.5.6 ($\exists$ introduction)**.** Let φ and ψ be formulas, x a variable, and τ a term. Then:

- If τ can be substituted in for x in φ and if

$$\Gamma \vdash \varphi \to \psi(\tau/x)$$

then

$$\Gamma \vdash \varphi \to \exists x\psi$$

- If x is not free in ψ or in any formula in Γ, and if

$$\Gamma \vdash \varphi \to \psi$$

  then

$$\Gamma \vdash \exists x \varphi \to \psi$$

From this is it straightforward to prove:

Theorem 8.5.7 (Deduction theorem)**.** *If* $\Gamma, \{\varphi\} \vdash \psi$*, then* $\Gamma \vdash \varphi \to \psi$*.*

Proof. Analogous to the proof of the propositional Deduction Theorem, Theorem 7.6.1. □

8.6 Proof theory: Natural deduction for first-order languages

In this section we extend the propositional proof rules of §7.5 with introduction and elimination rules for the quantifiers, quantifier interchange rules, and rules for =.

Two of the quantifier rules, $\forall\mathsf{E}$ and $\exists\mathsf{I}$, are easy; the rules work as you would immediately intuit that they should—a universally quantified statement is true of any individual named object, and if we can show that a given, named object has a property, we can conclude that there is *something* that hsa that property.

The other two, $\exists\mathsf{E}$ and $\forall\mathsf{I}$, are more difficult, because they require a notion of an arbitrary or generic instance. The two questions that underly these two rules are:

1. When can we eliminate an existential?

2. When can we introduce a universal, that is, generalize from a single instance?

The key to both of these is making inferences about arbitrary objects, as we explain below. First, we give the rules for =, which do not require the use of quantifiers.

8.6.1 Rules for =

Identity is an equivalence relation, which is to say it is *reflexive* (every object is identical with itself), *symmetric* (if x is identical to y, then y is identical to x), and *transitive* (if x is identical to y and y is identical to z, then x is identical to z as well), and we will have a proof rule corresponding to each of these facts. Unlike the rules for the other logical connectives, we do not have an introduction rule of identity: Logic will not tell us when two items are identical. It can, however, tell us what happens when we already know that two

items are identical (and thus in particular it can tell us when two things are *not* identical), and we have a substitution rule covering this case.

Rule 8.6.1 (Refl)**.** If τ is a term, then $\tau = \tau$ can be written on line n of a proof. The annotation for line n is Refl.

$$\begin{array}{|l} \\ \hline n\ \tau = \tau \end{array} \qquad \mathsf{Refl}$$

That is, every object is self-identical. Refl is often invoked in *reductio* arguments (that is, in the context of ¬I and ¬E, when one is attempting to reduce some assumption to a contradiction). If for some term τ it is possible to prove that $\tau \neq \tau$, Refl allows one to immediately write down a contradicting statement.

Rule 8.6.2 (Sym)**.** If $\tau_1 = \tau_2$ occurs on line i of a proof and $i < n$, then $\tau_2 = \tau_1$ can be written on line n. The annotation for line n is Sym, i.

$$\begin{array}{|l} \\ \hline i\ \tau_1 = \tau_2 \\ \vdots \\ n\ \tau_2 = \tau_1 \end{array} \qquad \mathsf{Sym}, i$$

That is, we are allowed to change the order of the terms in an identity statement.

Rule 8.6.3 (Trans)**.** If $\tau_1 = \tau_2$ occurs on line i of a proof, $\tau_2 = \tau_3$ occurs on line j of a proof, and $i < n$ and $j < n$, then $\tau_1 = \tau_3$ can be written on line n. The annotation for line n is Trans, i, j.

$$\begin{array}{|l} \\ \hline i\ \tau_1 = \tau_2 \\ \vdots \\ j\ \tau_2 = \tau_3 \\ \vdots \\ n\ \tau_1 = \tau_3 \end{array} \qquad \mathsf{Trans}, i, j$$

The similarity of this rule to the syllogism Barbara should be clear. Note that the order of the terms is important: If what we have at line i is $\tau_1 = \tau_2$ and at line j is $\tau_3 = \tau_2$, then Trans cannot (yet) be applied. First, Sym must be applied to line j, to obtain $\tau_2 = \tau_3$, as in the following example:

Example 8.6.4.

The final rule captures the principle known as the 'Indiscernibility of Identicals', that if two objects are the same, then they will have all the same properties. This principle is not to be confused with its converse, Leibniz's Law, or the Identity of Indiscernibles, which states that if two objects satisfy all the same properties, then they are identical. The Indiscernibility of Identicals is a law of logic; Leibniz's Law, however, is not, as it is easy to create models where it is false, if the language in question used to describe the models is impoverished enough.

Rule 8.6.5 (Substitution of Identicals)**.** If $\tau_1 = \tau_2$ occurs on line i of a proof, $\varphi(\tau_1)$ occurs on line j of a proof, and $i < n$ and $j < n$, then $\varphi(\tau_2/\tau_1)$ can be written on line n. The annotation for line n is SubsId, i, j.

8.6.2 Rules for $\forall x$

Rule 8.6.6 ($\forall$I)**.** If $\varphi(\tau)$ occurs on line i of a proof where $i < n$, and τ is a constant, then provided that τ does not occur in any assumption upon which φ depends and x is a variable

not occurring in φ, $\forall x \varphi(x/\tau)$ can be written on line n. The annotation for line n is $\forall \mathsf{I}, i$.

$$\begin{array}{|l|l} \hline & \\ \vdash & \\ \quad i \ \varphi(\tau) & \\ \quad \vdots & \\ \quad n \ \forall x \varphi(x/\tau) & \forall \mathsf{I}, i \\ \hline \end{array}$$

That is, if we are able to prove that a property holds of some arbitrary object, where nothing in our proof turned on the specific object that we chose, then we can prove that the property holds of all objects. Suppose we have managed to in fact prove $\varphi(\tau)$ on line i, where τ does not occur in any assumption that line i depends on. For example, the rule Refl (Definition 8.6.1, below) allows us to assert $\mathsf{a} = \mathsf{a}$ at any time (provided that a has already appeared somewhere in the proof), without any assumptions. Nothing in how we got line i depends on specific features of a, other than that it is in our language; we could easily have proven $\mathsf{b} = \mathsf{b}$, or $\mathsf{c} = \mathsf{c}$, or $x = x$. Thus, we can say that our choice of a was arbitrary; it is a generic object. Since we could have picked any object and still have proven the same (type of) formula, we can conclude that it's true of any object.

Rule 8.6.7 ($\forall \mathsf{E}$)**.** If $\forall x \varphi$ occurs on line i of a proof where $i < n$, and τ is a constant, then $\varphi(\tau/x)$ can be written on line n. The annotation for line n is $\forall \mathsf{E}, i$.

$$\begin{array}{|l|l} \hline & \\ \vdash & \\ \quad i \ \forall x \varphi & \\ \quad \vdots & \\ \quad n \ \varphi(\tau/x) & \forall \mathsf{E}, i \\ \hline \end{array}$$

That is, if we have proven that some property holds generally of any object, then we can prove that it holds of some particular, specific object, that is, it is true of every individual object. This is because if we have a universally quantified formula that is true, it will be true of every object in the domain. If we have appropriate terms to refer to those objects, then the result of substituting those terms in to the formula, with the universal quantifier removed, will also be true. For instance, suppose that we have three objects in our domain, a, b, c, and our language contains constants $\mathsf{a}, \mathsf{b}, \mathsf{c}$, interpreted in the sensible way. Then if we have proven $\forall x P x$, we can prove each of $P\mathsf{a}$, $P\mathsf{b}$, and $P\mathsf{c}$.

Rule 8.6.8 (QI)**.** If $\forall x \neg \varphi$ occurs on line i of a proof where $i < n$, then $\neg \exists x \varphi$ can be written on line n. The notation for line n is QI, i.

$$\begin{array}{|l} \\ \vdash \\ \quad i\ \forall x \neg \varphi \\ \quad \vdots \\ \quad n\ \neg \exists x \varphi \end{array} \qquad\qquad \mathsf{QI}, i$$

Similarly, if $\neg \forall x \varphi$ occurs on line i of a proof where $i < n$, then $\exists x \neg \varphi$ can be written on line n. The notation for line n is QI, i.

$$\begin{array}{|l} \\ \vdash \\ \quad i\ \neg \forall x \varphi \\ \quad \vdots \\ \quad n\ \exists x \neg \varphi \end{array} \qquad\qquad \mathsf{QI}, i$$

This rule is justified by Lemma 8.3.6. If it is not the case that everything does *not* have a particular property, φ, then there is at least one object that *does* have property φ. Similarly, if it is not the case that everything *does* have a particular property, φ, then there is at least one object that does not have property φ.

8.6.3 Rules for $\exists x$

Rule 8.6.9 ($\exists$I)**.** If $\varphi(\tau)$ occurs on line i of a proof where $i < n$ and τ is a constant, then $\exists x \varphi(\tau / x)$ can be written on line n. The annotation for line n is $\exists \mathsf{I}, i$.

$$\begin{array}{|l} \\ \vdash \\ \quad i\ \varphi(\tau) \\ \quad \vdots \\ \quad n\ \exists x \varphi(x/\tau) \end{array} \qquad\qquad \exists \mathsf{I}, i$$

That is, if you can prove that a certain property holds of a particular object, then you can prove the weaker claim, namely, that there is *some* object which has the property. For instance, if you have proven $P\mathsf{c}$, then you can prove the weaker claim $\exists x Px$.

Rule 8.6.10 ($\exists$E)**.** If $\exists x \varphi$ occurs on line i of a proof where $i < n$, and if ψ can be derived from the assumption of $\varphi(\tau / x)$ on line j $(i < j < n)$, where τ is a constant not occurring in

ψ or in any assumption on which ψ depends other than the assumption at line j, then ψ can be written on line n. The annotation for line n is $\exists\mathsf{E}, i, j\text{–}k$.

That is, suppose that we know that there is *something* that is φ. The way to eliminate this existential claim is by picking a generic object arbitrarily and assuming that it is the one that is φ. We need to pick a name for this object, and it must be a name that is not used elsewhere in the proof (for if it already occurred in the proof, then we would not have a guarantee that the object we picked was arbitrary or generic). If we can prove some independent formula ψ that doesn't contain this name, then we know that *whatever* object is the one that is φ, we would be able to make the same proof of ψ. Hence, we can prove ψ on its own.

Rule 8.6.11 (QI)**.** If $\exists x\neg\varphi$ occurs on line i of a proof where $i < n$, then $\neg\forall x\varphi$ can be written on line n. The notation for line n is QI, i.

Similarly, if $\neg\exists x\varphi$ occurs on line i of a proof where $i < n$, then $\forall x\neg\varphi$ can be written on line n. The notation for line n is QI, i.

This rule is also justified by Lemma 8.3.6. Similar to the version of this rule for the universal quantifier, if it's not the case that there is something that doesn't have some property φ, then it is the case that *everything* does have property φ. Likewise, if it is not the case that there is something that *does* have some property φ, then it is the case that *everything* doesn't have property φ.

8.6.4 Proof strategies

Because predicate logic involves propositional logic, all of the proof strategies outlined in §7.5.1 are still relevant here. However, we need to extend them with steps for the quantifiers. The extensions are marked in blue:

1. Do you have any assumptions? If yes, go to (1a); if no, go to (1c).
   a) Use as many propositional elimination rules as you can. Then go to (1b).
   b) Do you have any assumptions with $\forall$ as the main connective? If yes, go to (1(b)i); if no, go to (1c).
      i. Do any constants already appear in the proof? If no, go to (1c). If yes, apply $\forall$E (Rule 8.6.7) once for each constant that already appears in the proof.
   c) What is the main connective of what you are trying to prove?
      i. It is a conjunction. Go to (2).
      ii. It is a disjunction. Go to (3).
      iii. It is a negation or an atom.[3] Go to (4).
      iv. It is a conditional. Go to (5).
      v. It is a universal. Go to (6).
      vi. It is an existential. Go to (7).
      vii. I am trying to prove a contradiction, but don't know *which* contradiction. Go to (8).
2. If the wff you are trying to prove is a conjunction, then you can obtain it via $\wedge$I (Rule 7.5.6) if you can obtain both conjuncts. Consider each conjunct individually (return to (1)).
3. If the wff you are trying to prove is a disjunction, then you can obtain it via $\vee$I (Rule 7.5.8) if you can obtain a disjunct. Does either disjunct contain any of the atoms that appear in any of your premises? If yes, then try to prove that disjunct (return to 1c)).

[3] Remember that identity statements are atomic.

4. If the wff you are trying to prove is an atom or a negation, then you can obtain it by ¬E (Rule 7.5.13) or ¬I (Rule 7.5.12), respectively. That is, if you are trying to prove an atom, make its negation a new assumption; if you are trying to prove a negated atom or complex formula, make the atom or complex formula (without the negation) an assumption. Then return to (1).

5. If the wff you are trying to prove is a conditional, then you can obtain it by → I (Rule 7.5.10). That is, make the antecedent a new assumption, and try to prove the consequent (return to (1)).

6. If the wff you are trying to prove is a universal, then you can obtain it by ∀I (Rule 8.6.6). In order to apply ∀I to prove a universal $\forall x\varphi$, one should first prove $\varphi(\tau)$ where τ is an entirely new constant, one that does not occur previously in the proof. Pick appropriate τ, and then go to (1c).

7. If the wff you are trying to prove is an existential, then you can obtain it by ∃I (Rule 8.6.9). In order to apply ∃I to prove an existential $\exists x\varphi$, it suffices to prove an instance of the existential, e.g., $\varphi(\tau/x)$, for some arbitrary τ (your choice). Pick τ, and then go to (1c).

8. Good luck, this is one context in which no algorithm can reliably tell you what to do. Rely on the intuitions that you develop through practice, but also take a look at the next hints:

What to do when you get stuck:

- Look at any assumptions you haven't yet used, and see if you can now use them.

- Apply derived rules (see §7.5.1), identity rules (Rules 8.6.1, 8.6.2, 8.6.3), and QI (Rules 8.6.8 and 8.6.11) wherever you can.

8.6.5 Examples

Our first two examples are straightforward:

Example 8.6.12. $\mathsf{a} = \mathsf{b}, \mathsf{b} = \mathsf{c} \vdash \mathsf{a} = \mathsf{a}$ (proven without Refl).

1 $\mathsf{a} = \mathsf{b}$	Assumption
2 $\mathsf{b} = \mathsf{c}$	Assumption
3 $\mathsf{a} = \mathsf{c}$	Trans,1,2
4 $\mathsf{c} = \mathsf{a}$	Sym,3
5 $\mathsf{a} = \mathsf{a}$	Trans,3,4

Example 8.6.13. $\forall x(Rx \to Cx), R\mathsf{a} \vdash C\mathsf{a}$.

1 $\forall x(Rx \to Ca)$	Assumption
2 $R\mathsf{a}$	Assumption
3 $R\mathsf{a} \to C\mathsf{a}$	$\forall$E,1
4 $C\mathsf{a}$	$\to$E,2,3

Next, we give some examples with commentary at each stage:

Example 8.6.14. $\forall x(Ax \to Bx), \forall x(Bx \to Cx) \vdash \forall x(Ax \to Cx)$.

First, write down the premises as assumptions:

1 $\forall x(Ax \to Bx)$	Assumption
2 $\forall x(Bx \to Cx)$	Assumption

Next, look at the conclusion: The main connective is a universal quantifier. We know that if we get something of the form $A\mathsf{a} \to C\mathsf{a}$, then we can apply $\forall$I, since a doesn't occur in either lines (1) or (2). $A\mathsf{a} \to C\mathsf{a}$ is a conditional, so we start by assuming the antecedent, $A\mathsf{a}$, with a goal of obtaining the consequent, $C\mathsf{a}$:

1 $\forall x(Ax \to Bx)$	Assumption
2 $\forall x(Bx \to Cx)$	Assumption
3 $A\mathsf{a}$	Assumption

Now that we've got a constant, a, in our proof, we can use that constant when eliminating universals. (It is a good rule of thumb to eliminate universals every time a new constant is introduced).

1 $\forall x(Ax \to Bx)$	Assumption
2 $\forall x(Bx \to Cx)$	Assumption
⊢	
3 $A\mathsf{a}$	Assumption
⊢	
4 $A\mathsf{a} \to B\mathsf{a}$	∀E, 1
5 $B\mathsf{a} \to C\mathsf{a}$	∀E, 2

Now we don't have any quantifiers in the subproof, so we can simply apply propositional elimination and introduction rules:

1 $\forall x(Ax \to Bx)$	Assumption
2 $\forall x(Bx \to Cx)$	Assumption
⊢	
3 $A\mathsf{a}$	Assumption
⊢	
4 $A\mathsf{a} \to B\mathsf{a}$	∀E, 1
5 $B\mathsf{a} \to C\mathsf{a}$	∀E, 2
6 $B\mathsf{a}$	→E, 3, 4
7 $C\mathsf{a}$	→E, 5, 6

We've now obtained the consequent we were looking for, so we introduce the target conditional:

1 $\forall x(Ax \to Bx)$	Assumption
2 $\forall x(Bx \to Cx)$	Assumption
⊢	
3 $A\mathsf{a}$	Assumption
⊢	
4 $A\mathsf{a} \to B\mathsf{a}$	∀E, 1
5 $B\mathsf{a} \to C\mathsf{a}$	∀E, 2
6 $B\mathsf{a}$	→E, 3, 4
7 $C\mathsf{a}$	→E, 5, 6
8 $A\mathsf{a} \to C\mathsf{a}$	→I, 3–7

Now all that's left is the final step of introducing the universal quantifier. We must double check that a does not occur in any assumption that line (9) depends on. The only assumptions (9) depends on are those in lines (1) and (2); and neither of them contain any constants. Thus, we are allowed to generalise on a:

1 $\forall x(Ax \to Bx)$	Assumption
2 $\forall x(Bx \to Cx)$	Assumption
⊢	
3 $A\mathsf{a}$	Assumption
⊢	
4 $A\mathsf{a} \to B\mathsf{a}$	$\forall$E, 1
5 $B\mathsf{a} \to C\mathsf{a}$	$\forall$E, 2
6 $B\mathsf{a}$	$\to$E, 3, 4
7 $C\mathsf{a}$	$\to$E, 5, 6
8 $A\mathsf{a} \to C\mathsf{a}$	$\to$I, 3–7
9 $\forall x(Ax \to Cx)$	$\forall$I, 8

Example 8.6.15. $\forall x(Ax \to Bx), \neg B\mathsf{m} \vdash \exists x \neg Ax$. Again, we start by writing down the premises as assumptions:

1 $\forall x(Ax \to Bx)$	Assumption
2 $\neg B\mathsf{m}$	Assumption
⊢	

Our goal is an existential claim, which we can obtain via $\exists$I if we are able to first get $\neg A\tau$ for some term τ. The question is, what term? Well, the only non-variable term in the proof so far is m, so let's attempt to get $\neg A\mathsf{m}$. This is a negation, so we can assume the opposite and try to get a contradiction:

1 $\forall x(Ax \to Bx)$	Assumption
2 $\neg B\mathsf{m}$	Assumption
⊢	
3 $A\mathsf{m}$	Assumption
⊢	

As noted in the previous proof, whenever you have a universally quantified formula and at least one constant, always eliminate the universal using that constant:

1 $\forall x(Ax \to Bx)$	Assumption
2 $\neg B\mathsf{m}$	Assumption
3 $A\mathsf{m}$	Assumption
4 $A\mathsf{m} \to B\mathsf{m}$	$\forall$E, 1

Now the subproof doesn't involve any quantified formulas, so we can reason as we would in a propositional proof until we get our contradiction:

1 $\forall x(Ax \to Bx)$	Assumption
2 $\neg B\mathsf{m}$	Assumption
3 $A\mathsf{m}$	Assumption
4 $A\mathsf{m} \to B\mathsf{m}$	$\forall$E, 1
5 $B\mathsf{m}$	$\to$E, 3, 4
6 $\neg B\mathsf{m}$	Reit, 2
7 $\neg A\mathsf{m}$	$\neg$I, 3–6

Since we have proven that some particular object, m, is not A, we are allowed to write down the weaker claim, namely that there is *something* that is not A:

1 $\forall x(Ax \to Bx)$	Assumption
2 $\neg B\mathsf{m}$	Assumption
3 $A\mathsf{m}$	Assumption
4 $A\mathsf{m} \to B\mathsf{m}$	$\forall$E, 1
5 $B\mathsf{m}$	$\to$E, 3, 4
6 $\neg B\mathsf{m}$	Reit, 2
7 $\neg A\mathsf{m}$	$\neg$I, 3–6
8 $\exists x \neg Ax$	$\exists$I, 7

which is what we wanted to show in the first place.

Example 8.6.16. $\forall x((Ax \vee Bx) \to Cx), \exists y(Ay \wedge Dy) \vdash \exists y Cy$. As usual, we start by writing down the premises:

1 $\forall x((Ax \vee Bx) \to Cx)$	Assumption
2 $\exists y(Ay \wedge Dy)$	Assumption
⊢	

This is the first time we've had an existential premise, which means we're going to need to use the $\exists$E rule. The first step of this rule is to pick a constant to name the object that has the relevant property:

1 $\forall x((Ax \vee Bx) \to Cx)$	Assumption
2 $\exists y(Ay \wedge Dy)$	Assumption
⊢	
3 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
⊢	

Our goal is to get some formula which does *not* contain m. A good target formula is the ultimate conclusion, $\exists y Cy$, so keep that in the back of your mind. For now, though, we'll apply our standard trick of eliminating universals as soon as a new constant has been introduced:

1 $\forall x((Ax \vee Bx) \to Cx)$	Assumption
2 $\exists y(Ay \wedge Dy)$	Assumption
⊢	
3 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
⊢	
4 $(A\mathsf{m} \vee B\mathsf{m}) \to C\mathsf{m}$	$\forall$E, 1

Now we've got plenty that we can do at the propositional level:

1 $\forall x((Ax \vee Bx) \rightarrow Cx)$	Assumption
2 $\exists y(Ay \wedge Dy)$	Assumption
3 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
4 $(A\mathsf{m} \vee B\mathsf{m}) \rightarrow C\mathsf{m}$	$\forall$E, 1
5 $A\mathsf{m}$	$\wedge$E, 3
6 $A\mathsf{m} \vee B\mathsf{m}$	$\vee$I, 5
7 $C\mathsf{m}$	$\rightarrow$E, 4, 6

Given that we now know that m is C, it follows that there is *something* that is C:

1 $\forall x((Ax \vee Bx) \rightarrow Cx)$	Assumption
2 $\exists y(Ay \wedge Dy)$	Assumption
3 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
4 $(A\mathsf{m} \vee B\mathsf{m}) \rightarrow C\mathsf{m}$	$\forall$E, 1
5 $A\mathsf{m}$	$\wedge$E, 3
6 $A\mathsf{m} \vee B\mathsf{m}$	$\vee$I, 5
7 $C\mathsf{m}$	$\rightarrow$E, 4, 6
8 $\exists yCy$	$\exists$I, 7

Our final step is to finish eliminating the existential in line (2); for that, we need to check that (i) m does not occur in $\exists xCx$ AND (ii) m does not occur in any assumption that line

(9) depends on *except* line (3). Both of these conditions are met:

1 $\forall x((Ax \vee Bx) \to Cx)$	Assumption
2 $\exists y(Ay \wedge Dy)$	Assumption
3 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
4 $(A\mathsf{m} \vee B\mathsf{m}) \to C\mathsf{m}$	$\forall$E, 1
5 $A\mathsf{m}$	$\wedge$E, 3
6 $A\mathsf{m} \vee B\mathsf{m}$	$\vee$I, 5
7 $C\mathsf{m}$	$\to$E, 4, 6
8 $\exists y Cy$	$\exists$I, 7
9 $\exists y Cy$	$\exists$E, 2, 3–8

Example 8.6.17. $\exists x Ax \to \forall x(Bx \to Cx), \exists x Dx \to \exists x \neg Cx, \exists x(Ax \wedge Dx) \vdash \exists x \neg Bx$. Ooh, in this one we've got *three* premises instead of our usual two!

1 $\exists x Ax \to \forall x(Bx \to Cx)$	Assumption
2 $\exists x Dx \to \exists x \neg Cx$	Assumption
3 $\exists x(Ax \wedge Dx)$	Assumption

This is going to be a complicated one, because we again have an existential premise. Because the main connectives of lines (1) and (2) are $\to$, and we don't have the antecedent of either, so we cannot eliminate them, we *have* to engage with line (3). We start by picking an arbitrary name for the element which is both A and D:

1 $\exists x Ax \to \forall x(Bx \to Cx)$	Assumption
2 $\exists x Dx \to \exists x \neg Cx$	Assumption
3 $\exists x(Ax \wedge Dx)$	Assumption
4 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption

When in doubt, eliminate conjunctions:

1 $\exists x Ax \rightarrow \forall x(Bx \rightarrow Cx)$	Assumption
2 $\exists x Dx \rightarrow \exists x \neg Cx$	Assumption
3 $\exists x(Ax \wedge Dx)$	Assumption
4 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
5 $A\mathsf{m}$	$\wedge$E, 4
6 $D\mathsf{m}$	$\wedge$E, 4

Now we need to start looking at how we can eliminate the conditionals in lines (1) and (2). To do that, we need the antecedents, $\exists x Ax$ and $\exists x Dx$. Conveniently, we can generate them from lines (5) and (6):

1 $\exists x Ax \rightarrow \forall x(Bx \rightarrow Cx)$	Assumption
2 $\exists x Dx \rightarrow \exists x \neg Cx$	Assumption
3 $\exists x(Ax \wedge Dx)$	Assumption
4 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
5 $A\mathsf{m}$	$\wedge$E, 4
6 $D\mathsf{m}$	$\wedge$E, 4
7 $\exists x Ax$	$\exists$I, 5
8 $\exists x Dx$	$\exists$I, 6

This allows us to eliminate the conditionals:

1 $\exists x Ax \to \forall x(Bx \to Cx)$	Assumption
2 $\exists x Dx \to \exists x \neg Cx$	Assumption
3 $\exists x(Ax \wedge Dx)$	Assumption
4 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
5 $A\mathsf{m}$	$\wedge$E, 4
6 $D\mathsf{m}$	$\wedge$E, 4
7 $\exists x Ax$	$\exists$I, 5
8 $\exists x Dx$	$\exists$I, 6
9 $\forall x(Bx \to Cx)$	$\to$E, 1, 7
10 $\exists x \neg Cx$	$\to$E, 2, 8

Hurrah, this gives us a universal we can eliminate!

1 $\exists x Ax \to \forall x(Bx \to Cx)$	Assumption
2 $\exists x Dx \to \exists x \neg Cx$	Assumption
3 $\exists x(Ax \wedge Dx)$	Assumption
4 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
5 $A\mathsf{m}$	$\wedge$E, 4
6 $D\mathsf{m}$	$\wedge$E, 4
7 $\exists x Ax$	$\exists$I, 5
8 $\exists x Dx$	$\exists$I, 6
9 $\forall x(Bx \to Cx)$	$\to$E, 1, 7
10 $\exists x \neg Cx$	$\to$E, 2, 8
11 $B\mathsf{m} \to C\mathsf{m}$	$\forall$E, 9

Now we're in a position where we've used every line in the proof except (10), which is another existential. To eliminate that, we need to introduce another constant — one that hasn't been used before:

1	$\exists x Ax \to \forall x(Bx \to Cx)$	Assumption
2	$\exists x Dx \to \exists x \neg Cx$	Assumption
3	$\exists x(Ax \wedge Dx)$	Assumption
4	$A\mathsf{m} \wedge D\mathsf{m}$	Assumption
5	$A\mathsf{m}$	$\wedge$E, 4
6	$D\mathsf{m}$	$\wedge$E, 4
7	$\exists x Ax$	$\exists$I, 5
8	$\exists x Dx$	$\exists$I, 6
9	$\forall x(Bx \to Cx)$	$\to$E, 1, 7
10	$\exists x \neg Cx$	$\to$E, 2, 8
11	$B\mathsf{m} \to C\mathsf{m}$	$\forall$E, 9
12	$\neg C\mathsf{n}$	Assumption

The goal here is to get a formula that doesn't contain *either* m *or* n. One possible target formula is the overall conclusion, $\exists x \neg Bx$. But in the meantime, we've introduced a new

constant, so let's eliminate our universal again:

1 $\exists x Ax \to \forall x(Bx \to Cx)$	Assumption
2 $\exists x Dx \to \exists x \neg Cx$	Assumption
3 $\exists x(Ax \wedge Dx)$	Assumption
4 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
5 $A\mathsf{m}$	$\wedge$E, 4
6 $D\mathsf{m}$	$\wedge$E, 4
7 $\exists x Ax$	$\exists$I, 5
8 $\exists x Dx$	$\exists$I, 6
9 $\forall x(Bx \to Cx)$	$\to$E, 1, 7
10 $\exists x \neg Cx$	$\to$E, 2, 8
11 $B\mathsf{m} \to C\mathsf{m}$	$\forall$E, 9
12 $\neg C\mathsf{n}$	Assumption
13 $B\mathsf{n} \to C\mathsf{n}$	$\forall$E, 9
14 $\neg B\mathsf{n}$	MT, 12, 13
15 $\exists x \neg Bx$	$\exists$I, 14

We've reached our target formula; in order to bring it without the scope lines, we need to confirm that (i) n doesn't occur in $\exists x \neg Bx$ AND (ii) n does not occur in any assumption

that line (16) depends on *except* line (12).

1 $\exists x Ax \to \forall x(Bx \to Cx)$	Assumption
2 $\exists x Dx \to \exists x \neg Cx$	Assumption
3 $\exists x(Ax \wedge Dx)$	Assumption
4 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
5 $A\mathsf{m}$	$\wedge$E, 4
6 $D\mathsf{m}$	$\wedge$E, 4
7 $\exists x Ax$	$\exists$I, 5
8 $\exists x Dx$	$\exists$I, 6
9 $\forall x(Bx \to Cx)$	$\to$E, 1, 7
10 $\exists x \neg Cx$	$\to$E, 2, 8
11 $B\mathsf{m} \to C\mathsf{m}$	$\forall$E, 9
12 $\neg C\mathsf{n}$	Assumption
13 $B\mathsf{n} \to C\mathsf{n}$	$\forall$E, 9
14 $\neg B\mathsf{n}$	MT, 12, 13
15 $\exists x \neg Bx$	$\exists$I, 14
16 $\exists x \neg Bx$	$\exists$E, 10, 12–15

Next, we need to confirm that (i) m doesn't occur in $\exists x \neg Bx$ AND (ii) m does not occur in

any assumption that line (17) depends on *except* line (2):

1 $\exists x Ax \to \forall x(Bx \to Cx)$	Assumption
2 $\exists x Dx \to \exists x \neg Cx$	Assumption
3 $\exists x(Ax \wedge Dx)$	Assumption
4 $A\mathsf{m} \wedge D\mathsf{m}$	Assumption
5 $A\mathsf{m}$	$\wedge$E, 4
6 $D\mathsf{m}$	$\wedge$E, 4
7 $\exists x Ax$	$\exists$I, 5
8 $\exists x Dx$	$\exists$I, 6
9 $\forall x(Bx \to Cx)$	$\to$E, 1, 7
10 $\exists x \neg Cx$	$\to$E, 2, 8
11 $B\mathsf{m} \to C\mathsf{m}$	$\forall$E, 9
12 $\neg C\mathsf{n}$	Assumption
13 $B\mathsf{n} \to C\mathsf{n}$	$\forall$E, 9
14 $\neg B\mathsf{n}$	MT, 12, 13
15 $\exists x \neg Bx$	$\exists$I, 14
16 $\exists x \neg Bx$	$\exists$E, 10, 12–15
17 $\exists x \neg Bx$	$\exists$E, 3, 4–16

Example 8.6.18. $\forall x(Ax \to Cx), \exists x(Ax \wedge Gx) \vdash \exists x(Gx \wedge Cx)$.

1 $\forall x(Ax \to Cx)$	Assumption
2 $\exists x(Ax \wedge Gx)$	Assumption
3 $A\mathsf{a} \wedge G\mathsf{a}$	Assumption
4 $A\mathsf{a} \to C\mathsf{a}$	$\forall$E,1
5 $A\mathsf{a}$	$\wedge$E,3
6 $G\mathsf{a}$	$\wedge$E,3
7 $C\mathsf{a}$	$\to$E,4,5
8 $G\mathsf{a} \wedge C\mathsf{a}$	$\wedge$I,6,7
9 $\exists x(Gx \wedge Cx)$	$\exists$I,8
10 $\exists x(Gx \wedge Cx)$	$\exists$E,2,3–9

Example 8.6.19. $\forall x Px \vdash \forall x(x = x)$.

1 $\forall x Px$	Assumption
2 $P\mathsf{a}$	$\forall$E,1
3 $\mathsf{a} = \mathsf{a}$	Refl,2
4 $\forall x(x = x)$	$\forall$I,3

Example 8.6.20. $(P\mathsf{s} \wedge I\mathsf{s}) \wedge \forall x((Px \wedge Ix) \rightarrow x = \mathsf{s}), P\mathsf{c} \wedge L\mathsf{c}, \exists x((Px \wedge Ix) \wedge \neg Lx) \vdash \mathsf{c} \neq \mathsf{s}$.

1 $(P\mathsf{s} \wedge I\mathsf{s}) \wedge \forall x((Px \wedge Ix) \rightarrow x = \mathsf{s})$	Assumption
2 $P\mathsf{c} \wedge L\mathsf{c}$	Assumption
3 $\exists x((Px \wedge Ix) \wedge \neg Lx)$	Assumption
4 $\mathsf{c} = \mathsf{s}$	Assumption
5 $L\mathsf{c}$	$\wedge$E,2
6 $L\mathsf{s}$	SubsId,4,5
7 $(P\mathsf{a} \wedge I\mathsf{a}) \wedge \neg L\mathsf{a}$	Assumption
8 $P\mathsf{a} \wedge I\mathsf{a}$	$\wedge$E,7
9 $\forall x((Px \wedge Ix) \rightarrow x = \mathsf{s})$	$\wedge$E,1
10 $(P\mathsf{a} \wedge I\mathsf{a}) \rightarrow \mathsf{a} = \mathsf{s}$	$\forall$E,9
11 $\mathsf{a} = \mathsf{s}$	$\rightarrow$E,8,10
12 $\mathsf{s} = \mathsf{a}$	Sym,11
13 $\mathsf{c} = \mathsf{a}$	Trans,4,12
14 $\neg L\mathsf{a}$	$\wedge$E,7
15 $\neg L\mathsf{c}$	SubsId,13,14
16 $\neg L\mathsf{c}$	$\exists$E,3,7–15
17 $\mathsf{c} \neq \mathsf{s}$	$\neg$I,4–16

Example 8.6.21. $F\mathsf{e} \wedge \forall x(Fx \to x = \mathsf{e}), \mathsf{e} \neq \mathsf{n}, C\mathsf{n} \vdash \exists x(Cx \wedge \neg Fx)$.

1 $F\mathsf{e} \wedge \forall x(Fx \to x = \mathsf{e})$	Assumption
2 $\mathsf{e} \neq \mathsf{n}$	Assumption
3 $C\mathsf{n}$	Assumption
4 $\forall x(Fx \to x = \mathsf{e})$	$\wedge$E,1
5 $F\mathsf{n} \to \mathsf{n} = \mathsf{e}$	$\forall$E,4
6 $\mathsf{n} \neq \mathsf{e}$	Sym,2
7 $\neg F\mathsf{n}$	MT,5,6
8 $C\mathsf{n} \wedge \neg F\mathsf{n}$	$\wedge$I,3,7
9 $\exists x(Cx \wedge \neg Fx)$	$\exists$I,8

Example 8.6.22. $\forall x \forall y(Axy \to (Bx \wedge Cy)), \forall x \forall y((Bx \vee Dy) \to \neg Axy) \vdash \neg\exists x \exists y Axy$.

1 $\forall x \forall y(Axy \to (Bx \wedge Cy))$	Assumption
2 $\forall x \forall y((Bx \vee Dy) \to \neg Axy)$	Assumption
3 $\forall y(A\mathsf{a}y \to (B\mathsf{a} \wedge Cy))$	$\forall$E,1
4 $A\mathsf{ab} \to (B\mathsf{a} \wedge C\mathsf{b})$	$\forall$E,3
5 $\forall y((B\mathsf{a} \vee Dy) \to \neg A\mathsf{a}y)$	$\forall$E,2
6 $(B\mathsf{a} \vee D\mathsf{b}) \to \neg A\mathsf{ab}$	$\forall$E,5
7 $A\mathsf{ab}$	Assumption
8 $B\mathsf{a} \wedge C\mathsf{b}$	$\to$E,4,7
9 $B\mathsf{a}$	$\wedge$E, 8
10 $B\mathsf{a} \vee D\mathsf{b}$	$\vee$I, 9
11 $\neg A\mathsf{ab}$	$\to$E, 6, 10
12 $\neg A\mathsf{ab}$	$\neg$I,7-11
13 $\forall y \neg A\mathsf{a}y$	$\forall$I,12
14 $\neg\exists y A\mathsf{a}y$	QI,13
15 $\forall x \neg\exists y Axy$	$\forall$I,14
16 $\neg\exists x \exists y Axy$	QI,15

Example 8.6.23. $\forall x \neg \varphi(x) \vdash \neg \exists x \varphi(x)$

1 $\forall x \neg \varphi(x)$	Assumption
2 $\neg\neg \exists x \varphi(x)$	Assumption
3 $\exists x \varphi(x)$	DN, 2
4 $\varphi(\mathsf{a}/x)$	Assumption
5 $\neg \varphi(\mathsf{a}/x)$	$\forall$E, 1
6 $\neg(\psi \wedge \neg\psi)$	Assumption
7 $\varphi(\mathsf{a}/x)$	Reit, 4
8 $\neg \varphi(\mathsf{a}/x)$	Reit, 5
9 $\psi \wedge \neg\psi$	$\neg$E, 6–8
10 $\psi \wedge \neg\psi$	$\exists$E, 3, 4–9
11 ψ	$\wedge$E, 10
12 $\neg\psi$	$\wedge$E, 10
13 $\neg \exists x \varphi(x)$	$\neg$E, 2–12

Example 8.6.24. $\exists x(Fx \wedge \neg Gx), \forall x Gx \vdash \forall x \forall y(x \neq y)$

1 $\exists x(Fx \wedge \neg Gx)$	Assumption
2 $\forall x Gx$	Assumption
3 $F\mathsf{a} \wedge \neg G\mathsf{a}$	Assumption
4 $\mathsf{b} = \mathsf{c}$	Assumption
5 $\neg G\mathsf{a}$	$\wedge$E, 3
6 $G\mathsf{a}$	$\forall$E, 2
7 $\mathsf{b} \neq \mathsf{c}$	$\neg$I, 4–6
8 $\forall y(\mathsf{b} \neq y)$	$\forall$I, 7
9 $\forall x \forall y(x \neq y)$	$\forall$I, 8
10 $\forall x \forall y(x \neq y)$	$\exists$E, 1, 3–9

8.6.6 Soundness and completeness

This system of proof rules introduced in this section is sound and complete with respect to the semantics introduced in §8.3:

Theorem 8.6.25 (Soundness)**.**

$$\textit{If } \vdash \varphi \textit{ then } \vDash \varphi$$

Soundness

Just as we did in the propositional case, we prove the soundness of the entire proof system by proving the soundness of each rule individually. While the semantic definitions of truth, interpretation, model, etc., differ, the essential proof strategy is the same.

First, note that the Deduction Theorem (cf. Theorem 7.6.1) holds for predicate logic:

Theorem 8.6.26 (Deduction theorem)**.** *Let* Γ *be a set of* $\mathcal{L}_q$ *formulas. Then*

$$\Gamma \vdash \varphi \to \psi \textit{ iff } \Gamma, \varphi \vdash \psi$$

Proof. Analogous to the proof of Theorem 7.6.1. □

This means that Facts 7.6.4 and Corollary 7.6.5 also hold, and that means if we are able to prove the following Lemma, we have effectively established Theorem 8.6.25:

Lemma 8.6.27.

$$\textit{If } \Gamma \vdash \varphi \textit{ then } \Gamma \vDash \varphi$$

Proof. Suppose that $\Gamma \vdash \varphi$. Then the annotation of φ is one of the basic rules other than Assumption, and φ occurs on some line $n \geq 2$.

First, note that in the case of the propositional operators, the truth conditions in Definition 8.3.4 correspond directly to the truth tables given in Definitions 7.4.4, **??**, 7.4.6, and 7.4.7.

This means if we take as our truth value assignments the truth values provided by an arbitrary quantified model $\mathfrak{M}$ and variable assignment v, we do not need to prove the soundness of any of the propositional proof rules, since they have already been proven sound with respect to the classical truth tables.

This means we only need to prove the soundness of the new rules, for identity and the quantifiers. We consider the cases of $\forall\mathsf{I}$ and $\forall\mathsf{E}$ and one pair of QI rules, leaving the other quantifier rules and the identity rules as exercises for the reader.

$\forall\mathsf{I}$ (to be written)

$\forall$**E** Let n be a line of a proof annotated with $\forall\mathsf{E}, i$ for $n < i$. That is, the formula at line n is of the form $\varphi(\tau/x)$, where $\forall x\varphi$ occurs on line i and τ is a constant that has been substituted in for every free occurence of x in φ. Assume that the formula on line n is true. This means that for every x-variant v' of v, and any model $\mathfrak{M}$, $\mathfrak{M}, v' \vDash \varphi$ is true. This includes the x-variant where x is assigned the same objection as $I(\tau)$, in which case $\mathfrak{M}, v' \vDash \varphi(\tau/x)$.

QI (to be written)

(to be completed) □

Completeness

to be written

Part III

Extensions of Classical Logics

Chapter 9

Introduction to Part III [last modified 4 Oct 18]

In this part of the book, we explore extensions of classical logic: New logics which go beyond classical logic in what they are able to prove, whether through the addition of new axioms or new logical vocabulary.

Each of these new logics are nevertheless still *classical* in the sense that they retain the basic principles of classical logic

Chapter 10

Peano Arithmetic [last modified 5 Mar 26]

10.1 Introduction

In this chapter, we address the following question:

Is every truth of arithmetic on the standard model* provable**?

To give a complete answer to this question we must specify what is meant both by "standard model" and "provable". The standard model of arithmetic is the model that everyone is implicitly using when they are learning basic arithmetic in primary school—the elements of the model function as numbers ought to, and the functions in the model are the usual functions. The standard model is introdued in §10.2.

When we say "provable", we have a very specific notion of provability. We are interested in whether there is a proof system that is:

- Consistent (cf. Definition 8.5.3.)
- Decidable. (cf. Definition 8.4.4.)

The main result of this chapter is that the answer to the question is 'No': There is no consistent and decidable proof system that proves all the truths of arithmetic. We will construct a specific sentence, the Gödel sentence, which is true on the standard model of arithemtic if and only if it is not provable.

The basic idea is this: The language we define contains a finite number of non-logical symbols and only countably many variables. Thus, it is possible to give an enumeration of

all the symbols. Because each wff is of finite length and constructed out of finitely many symbols, we can also give an enumeration of all the wffs of the language. More than that: We can define our enumeration to be such that all (and only) those strings which are, e.g., atomic wffs are enumerated with numbers that bear a specific property, and the same with, e.g., conjunctions, implications, etc. This allows us to equate sets of formulas with sets of numbers, and it will turn out that these sets are *recursive* (cf. Definition 10.3.3). This method of numbering generalises, so that we can also define a notion of 'proof' and give an enumeration of proofs. However, the set of numbers that correspond to proofs will turn out to be only *recursively enumerable* (cf. Definition 10.3.2), and not recursive. It is this fact that gives rise to the phenomenon of incompleteness that underpins the negative answer to the question we kicked this chapter off with.

10.2 Language and semantics

Definition 10.2.1. $\mathcal{L}_{\text{PA}}$, the language of PA, is a first-order language containing one constant symbol, $\mathtt{0}$; one unary function, $\mathtt{S}$; and three binary functions, $\mathtt{Add}$, $\mathtt{Mult}$, and $\mathtt{Exp}$.

Terms and formulas in this language are defined recursively in the standard way (cf. Definitions 8.2.2 and 8.2.6). Recall that every formula involving $\exists x$, $\vee$, or $\rightarrow$ can be rewritten as a formula involving only $\forall x$, $\neg$, and $\wedge$.[1] In what follows, we will often exploit these facts by proving results for the fragment of $\mathcal{L}_{\text{PA}}$ involving $\forall x$, $\neg$, and $\wedge$, leaving it to the reader to reduce any formulas involving the extra connectives and operators to ones that don't.

This language is interpreted on the *standard model* of arithmetic.

Definition 10.2.2. The *standard model* of arithmetic takes the set of natural numbers, $\mathbb{N}$, as its domain, interprets $\mathtt{0}$ as 0, and interprets the functions with the standard successor, addition, multiplication, and exponentiation functions. That is: $\mathfrak{M} = \langle \mathbb{N}, I \rangle$ where:

- $I(\mathtt{0}) = 0$
- $I(\mathtt{S}) = n \mapsto n + 1$
- $I(\mathtt{Add}) = \{n, m\} \mapsto n + m$
- $I(\mathtt{Mult}) = \{n, m\} \mapsto n \times m$
- $I(\mathtt{Exp}) = \langle n, m \rangle \mapsto n^m$

[1] The propositional result is Corollary 7.4.28. The quantificational result is a consequence of the soundness of Rules 8.6.8 and 8.6.11.

(Here it is important to recognize that n and m are being used as meta-variables ranging over natural numbers; they are not variables inside the language.)

In what follows, where no confusion results, we will use $+$, $\times$, and $\uparrow$ instead of $\mathtt{Add}$, $\mathtt{Mult}$, and $\mathtt{Exp}$. In many cases, we will also refer to the standard model of arithmetic simply by the set of natural numbers, $\mathbb{N}$. Thus, instead of saying, e.g.,

$$\forall x \forall y \forall z (\mathtt{Add}(\mathtt{Add}(x, y), z) = \mathtt{Add}(x, \mathtt{Add}(y, z)) \tag{10.1}$$

we will write the more readable

$$\forall x \forall y \forall z ((x + y) + z = x + (y + z)) \tag{10.1$'$}$$

Next, we introduce a notion of a "building sequence" for a term (or a formula) which we will make use of in Section 10.6.

Definition 10.2.3 (Term-building sequences)**.** An ordered sequence of terms $\langle \tau_1, \ldots, \tau_n \rangle$ *builds a term* τ if:

- τ occurs in the sequence.
- For every $k \leq n$, τ_k is either
  1. the constant $\mathtt{0}$ or a variable x or
  2. it is of the form $\mathtt{S}\tau_i$ for some $i < k$.
  3. it is of the form $\tau_i + \tau_j$ or $\tau_i \times \tau_j$ or $\tau_i^{\tau_j}$ for some $i, j < k$ or

Definition 10.2.4 (Wff-building sequences)**.** A sequence of formulas $\langle \varphi_1, \ldots, \varphi_n \rangle$ *builds a formula* φ if:

- φ occurs in the sequence.
- For every $k \leq n$, φ_k is either
  1. of the form $\tau_1 = \tau_2$ for two terms τ_1, τ_2 or
  2. it is of the form $\neg\varphi_i$ or $\forall x \varphi_i$ for some variable x and $i < k$.
  3. it is of the form $\varphi_i \wedge \varphi_j$ for some $i, j < k$ or

Note that a term- or wff-building sequence builds *every* term or wff in the sequence.

Example 10.2.5. The following sequence of terms is a term-building sequence:

$$\langle x, y, z, x + y, \mathtt{S}x, \mathtt{S}x \times z, \mathtt{S}(\mathtt{S}x \times z), (x + y) \times \mathtt{S}(\mathtt{S}x \times z) \rangle$$

Example 10.2.6. The following sequence of terms is a not term-building sequence:

$$\langle x, y, y^x, \mathtt{S}(y^x), \mathtt{S}z, \mathtt{S}z + \mathtt{S}(y^x), x + \mathtt{S}(y^x)\rangle$$

Example 10.2.7. The following sequence of wffs is a wff-building sequence:

$$\langle \mathtt{S}(x + y) = x + \mathtt{S}y, \forall y(\mathtt{S}(x + y) = x + \mathtt{S}y), \forall x \forall y(\mathtt{S}(x + y) = x + \mathtt{S}y)\rangle$$

We use term- and wff-building sequences when we develop our method of encoding terms and formulas by numbers in §10.6.

The language we've introduced provides us with exactly as much expressive power as we need to express any arithmetic statement we want, without proliferating the number of different symbols (which would require us to multiply the number of necessary axioms). However, while this makes for a neat logic, it doesn't necessarily allow us to express arithmetic statements in the most straightforward of fashions. To that end, we introduce a number of short-hand definitions that allow us to write down formulas which when read off in English sound like ordinary arithmetic statements.

First, note that we have no constant in our language other than $\mathtt{0}$, which means that the only way we have to refer to other natural numbers within PA is via $\mathtt{S}$, e.g., $\mathtt{S0}$ picks out 1, $\mathtt{SS0}$ is 2, and in general for any n. Because writing out n instances of $\mathtt{S}$ followed by '0' is unnecessarily tedious, we introduce the following short-hand:

Definition 10.2.8. If n is a natural number, then $\underline{n}$ is the term $\mathtt{S}\ldots\mathtt{S0}$ that contains exactly n occurrences of $\mathtt{S}$.

Note also that the language contains no primitive relations. We introduce three defined relations, $\leq$, $<$, and $|$:

Definition 10.2.9. We write $x \leq y$ 'x is less than or equal to y' for $\exists r(x + r = y)$ and $x < y$ 'x is (strictly) less than y' for $x \leq y \wedge x \neq y$.

Definition 10.2.10. We write $x|y$ 'x divides y' for $\exists r(x \times r = y)$.

Using $\leq$ (or $<$), we also introduce the following short-hand notation:

Definition 10.2.11 (Bounded quantifiers)**.**

$$\begin{aligned} \forall x \leq \tau \varphi &:= \forall x(x \leq \tau \rightarrow \varphi) \\ \exists x \leq \tau \varphi &:= \exists x(x \leq \tau \wedge \varphi) \end{aligned}$$

where τ is any term (whether constant, variable, or function). These are read "for every x less than or equal to τ, φ" and "there is an x less tha nor equal to τ and φ", respectively.

The next definitions crucially underpin the Incompleteness Theorem, as we will see later in this chapter:

Definition 10.2.12 (Bounded formula)**.** The set of *bounded formulas* is defined recursively as follows:

1. Every atomic formula is a bounded formula.
2. If ψ is a bounded formula, then $\neg\psi$ is a bounded formula.
3. If ψ_1 and ψ_2 are bounded formulas, then $\psi_1 \wedge \psi_2$ is a bounded formula.
4. If ψ is a bounded formula and x a variable not occurring in term τ, then $\forall x \leq \tau\psi$ is a bounded formula.[2]

That is, bounded formulas are propositional combinations of atomic formulas and formulas where every quantifier has an explicit upper bound.

Definition 10.2.13 (Σ^0_1 formula)**.** A Σ^0_1 *formula* is a formula of the form $\exists x\psi$ where ψ is a bounded formula.

Definition 10.2.14 (Π^0_1 formula)**.** A Π^0_1 *formula* is a formula of the form $\forall x\psi$ where ψ is a bounded formula.

Note that these are *syntactic* definitions, concerning the shape or form of the wff. Many formulas will not have the right shape and thus will not be Σ^0_1 formulas. However, in general we are not interested in classifying formulas on the basis of their forms, but rather on the basis of their truth conditions. That is, instead of only being interested in Σ^0_1 formulas, we will be interested in anything that is equivalent to a Σ^0_1 formula according to the following definition of equivalent:

Definition 10.2.15 (Equivalence)**.** We say that two formulas φ and ψ are *equivalent* if

$$\mathbb{N} \vDash \varphi \quad \text{iff} \quad \mathbb{N} \vDash \psi$$

Note that this is a *semantic* notion, involving truth in a model. It is weaker than another notion of equivalence that we will define in §10.5, of provable equivalence (cf. Definition 10.5.1).

Fact 10.2.16. $x|y$ is equivalent to a bounded formula.

[2]The reason why x must not occur in τ is to prevent formulas with circular bounds such as $\forall x \leq (x+y)\varphi$.

Proof. Let $\mathbb{N}$ be the standard model of arithemtic. Then:

$$\mathbb{N} \vDash \exists r(x \times r = y) \quad \text{iff} \quad \mathbb{N} \vDash \exists r \leq y(x \times r = y)$$

□

Note that while no bounded formula is, strictly speaking, a Σ_1^0 formula, every bounded formula is equivalent to a Σ_1^0 formula, because a redundant 'dummy' existential quantifier that scopes over no variable in the bounded formula can always be appended to the front of a bounded formula. In fact, we can prove a more general result:

Theorem 10.2.17 (Closure properties of Σ_1^0 formulas)**.** *If φ and ψ are Σ_1^0 formulas, and x is a variable not free in term τ, then the following are all equivalent to Σ_1^0 formulas:*

1. $\exists x\varphi$
2. $\varphi \wedge \psi$
3. $\varphi \vee \psi$
4. $\exists x \leq \tau\varphi$
5. $\forall x \leq \tau\varphi$

Note the conspicuous absence of negated formulas, including conditionals and formulas with unbounded universal quantification.

Proof.

1. Since φ is a Σ_0^1 formula, it is of the form $\exists y\psi$, where ψ is a bounded formula. We need to show that

   $$\exists x\exists y\psi$$

   is equivalent to a Σ_0^1 formula. Let z be a variable not occuring in ψ. Then, $\exists x\exists y\psi$ is equivalent to $\exists z\exists x \leq z\exists y \leq z\psi$, which is a Σ_0^1 formula as $\exists x \leq z\exists y \leq z\psi$ is a bounded formula (bounded existential quantification of a bounded formula remains bounded).

2. Since φ and ψ are Σ_0^1 formulas, they are of the form $\exists x_1\varphi'$ and $\exists x_2\psi'$, where φ' and ψ' are bounded formulas. We need to show that

   $$\exists x_1\varphi' \wedge \exists x_2\psi'$$

is equivalent to a Σ_0^1 formula. Note that we can assume that x_1 does not appear in ψ' and x_2 does not appear in φ'; if either does, then we rename variables appropriately so that there is no clash. It then follows that $\varphi \wedge \psi$ is equivalent to

$$\exists x_1 \exists x_2 (\varphi' \wedge \psi')$$

By (1), this is equivalent to a Σ_0^1 formula, because conjunctions of bounded formulas are themselves bounded.

3. This case is analogous to the previous.

4. Recall (Definition 10.2.11) that $\exists x \leq \tau \varphi$ is short-hand for $\exists x (x \leq \tau \wedge \varphi)$, which itself is definitionally equivalent to $\exists x (\exists r (x + r = \tau) \wedge \varphi)$. It then follows from (2) and then (1) that this is equivalent to a Σ_0^1 formula.

5. Since φ is a Σ_0^1 formula, it is of the form $\exists y \psi$ where ψ is bounded. We need to show that

$$\varphi' \coloneqq \forall x \leq \tau \exists y \psi$$

is equivalent to a Σ_0^1 formula. Let z be a variable not occuring in φ'. Because

$$\mathbb{N} \vDash \forall x \leq \tau \exists y \leq z \psi \to \forall x \leq \tau \exists y \psi,$$

it follows that

$$\mathbb{N} \vDash \exists z \forall x \leq \tau \exists y \leq z \psi \to \forall x \leq \tau \exists y \psi,$$

the antecedent of which is a Σ_0^1 formula. Now we show that

$$\mathbb{N} \vDash \forall z \leq \tau \exists y \psi \to \exists z \forall x \leq \tau \exists y \leq z \psi.$$

Let $x_1, \ldots, x_k$ be the free variables in φ, let ψ be $\psi(x, y, x_1, \ldots, x_k)$, and let $n_1, \ldots, n_k$ be natural numbers. We now show that if

$$\mathbb{N} \vDash \forall x \leq \tau(n_1, \ldots, n_k) \exists y \psi(x, y, n_1, \ldots, n_k) \tag{10.2}$$

then also

$$\mathbb{N} \vDash \exists z \forall x \leq \tau(n_1, \ldots, n_k) \exists y \leq z \psi(z, y, n_1, \ldots, n_k) \tag{10.3}$$

Note that since $\tau(n_1/x_1, \ldots, n_k/x_k)$ is a closed term in $\mathbb{N}$, it has a value, call it n. Then (10.2) says that there are natural numbers $m_0, m_1, \ldots, m_n$ such that for $i = 0, 1, \ldots, n$ we have:

$$\mathbb{N} \vDash \psi(i, m_i, n_1, \ldots, n_k)$$

Let m be the maximum of $m_0, \ldots, m_i$. Then:

$$\mathbb{N} \vDash \forall x \leq n \exists y \leq m \psi(x, y, n_1, \ldots, n_k),$$

which implies (10.3).

□

Given the results in Theorem 10.2.17, we will sometimes say that "Σ_1^0 formulas are closed under conjunction, disjunction, existential quantification, and bounded quantification", taking equivalent formulas as identical.

10.2.1 Non-standard models

to be written

10.3 Recursive functions

In this section we introduce the notion of *recursive* and *primitive recursive* functions, as well as *definable sets*. The primitive recursive functions are the ones that underpin PA.

Definition 10.3.1. A set $A \subseteq \mathbb{N}$ of natural numbers is *expressed by* or *defined by* a formula $\varphi(x)$ with one free variable x if

$$A = \{n \in \mathbb{N} : \mathbb{N} \vDash \varphi(n)\}.$$

For example, the set of even numbers is defined by the formula $\exists y(y + y = x)$. say something here about set comprehension and reference back to §1.3.3.

Definition 10.3.2. A set $A \subseteq \mathbb{N}$ of natural numbers is *recursively enumerable* or *r.e.* if it is defined by a Σ_1^0 formula $\varphi(x)$ with one free variable x.

Similarly, a k-ary relation $A \subseteq \mathbb{N}^k$ is *recursively enumerable* or *r.e.* if it is defined by a Σ_1^0 formula $\varphi(x_1, \ldots, x_k)$ with k free variables.

Definition 10.3.3. A set $A \subseteq \mathbb{N}^k$ is *recursive* if both A and $\mathbb{N} - A$ are recursively enumerable.

We shall see in §10.6 that the set of natural numbers which are Gödel codes of theorems of PA is recursively enumerable but not recursive.

Functions can be thought of as sets of pairs, the input and the output of the function. Sets of pairs, in turn, can be thought of as binary relations: A function is a relation that relates an input to a unique output. It then makes sense to talk of *recursive functions*.

Definition 10.3.4. A unary function $f : \mathbb{N} \mapsto \mathbb{N}$ is *recursive* if the set

$$A = \{(n, f(n)) : n \in \mathbb{N}, f(n) \in \mathbb{N}\}$$

is recursive, and similarly for an n-ary function $g : \mathbb{N}^n \mapsto \mathbb{N}$. This set is also called the *graph* of the function.

Recursive sets are important because they are decidable: Because both the set and its complement are expressed by a bounded formula, there will only ever be finitely many values that must be checked to determine whether n is a member of the set or its complement.

What we will see in the remainder of this section is another way to characterise recursive functions: They are functions whose values are defined by using other values of the same function.[3]

If the function takes more than one input (e.g., the addition function + takes two inputs), then we look at the set which associates n inputs with one output.

Fact 10.3.5.

- Addition is a recursive function from $\mathbb{N}^2$ to $\mathbb{N}$.
- Any function defined by a term in the language of PA is recursive.
- If f is recursive and $\psi(x, y)$ is a bounded formula, then

  $$A = \{n \in \mathbb{N} : \mathbb{N} \vDash \exists y(y \leq f(n) \wedge \psi(n, y))\}$$

  is recursive.
- Every total function that is r.e. is also recursive.

(Recall that a function is total if it is defined for every input. Addition in $\mathbb{N}$ is a total function, but subtraction is not: $4 - 7$ is not defined, because $-3 \notin \mathbb{N}$.)

Why are recursive functions so important? (a) They have very useful computational properties (they can be computed in an algorithmic fashion) and (b) most of the basic number theoretic functions are recursive functions, so there is reason to study them as a

[3] Such a definition is a "definition by recursion" or a "recursive definition", and we've already seen examples of recursive definitions. The definition of well-formed formula in propositional logic is a recursive one: First the basis clause is defined (i.e., what are the atomic formulas) and then the recursive clause takes things that are already known to be formulas and converts them into more complex formulas. Sometimes you'll also hear such definitions called 'inductive', because they start with a basis case and then have an inductive clause: If something of level n falls under the definition, then things of level $n + 1$ also do. There is a close conceptual link between induction and recursion.

class and determine what their properties are. Within the class of recursive functions we can pick out an important subset, the ones which are *primitive recursive*; they are called this because they are all composed from a small set of recursive functions and a few types of function composition (we outline these below). (In a sense, you can think of the initial functions as the 'axioms' and the types of function composition as the 'rules of inference'). It will turn out that there are in fact very few total recursive functions which are *not* primitive recursive — and these recursive but not primitive recursive functions are ones which are not 'naturally occurring' but rather were defined by mathematicians to show that recursion and primitive recursion do not coincide.

10.3.1 Primitive recursive functions

There are three basic types of primtive recursive functions: The 0 function, the successor function, and the projection function.

Definition 10.3.6. The 0-ary constant function `0` is the function which takes no inputs and always returns `0`.

Definition 10.3.7. The 1-ary *successor function* `S` is the function which takes one input m and returns $m+1$.

Definition 10.3.8. For $n \geq 1$ and $1 \leq i \leq n$, the n-ary *projection function* P_i^n takes n inputs $m_1, \ldots, m_n$ and returns that ith argument.

More complex functions can be created from these via a finite number of applications of composition and primitive recursion.

Definition 10.3.9. Let f be a k-ary primitive recursive function, and $g_1, \ldots, g_k$ be m-ary primitive recursive functions. The *composition of f with $g_1, \ldots, g_k$* is the m-ary function

$$h(x_1, \ldots, x_m) = f(g_1(x_1, \ldots, x_m), \ldots, g_k(x_1, \ldots, x_m)).$$

Definition 10.3.10. Let f be a k-ary recursive function and g a $(k+2)$-ary primitive recursion. Then the $(k+1)$-ary function h is the *primitive recursion of f and g* when h is defined as follows:

$$\begin{aligned} h(0, x_1, \ldots, x_k) &= f(x_1, \ldots, x_k) \\ h(\mathtt{S}(y), x_1, \ldots, x_k) &= g(y, h(y, x_1, \ldots, x_k), x_1, \ldots, x_k) \end{aligned}$$

What you can see here is that primitive recursive functions formed by primitive recursion have an inductive flavor to them: The function is first defined by how it behaves for 0, taking

its behavior from f alone, and then it is defined by how it behaves for successors, and this is defined in terms of both g and h.

Example 10.3.11. Let $f(x) = P_1^1(x) = x$ and $g(x, y, z) = \mathtt{S}(P_2^3(x, y, z)) = \mathtt{S}(y)$. Then $h(0, x) = x$ and $h(\mathtt{S}(y), x) = g(y, h(y, x), x) = \mathtt{S}(h(y, x))$. f, g, and h are all primitive recursive.

Example 10.3.12. The function $\mathtt{Add}(x, y)$ is primitive recursive. It can be defined as:

$$\begin{aligned} \mathtt{Add}(0, x) &= x \\ \mathtt{Add}(\mathtt{S}(y), x) &= \mathtt{S}(\mathtt{Add}(y, x)) \end{aligned}$$

(Note: This should look suspiciously familiar. From this, can you show that `Mult` and `Exp` are primitive recursive?) Recall that when we first defined the notions of 'recursively enumerable' and 'recursive', it was with respect to sets, and then we showed how functions could be thought of as being recursively enumerable or recursive on the basis of the sets formed from those functions. But above we define 'primitive recursive' with respect to functions. Is there a way to generalize this back to sets? Yes, via the notion of a characteristic function for a set.

Definition 10.3.13. The characteristic function χ_A of a set $A \subseteq \mathbb{N}^k$ is a function from $\mathbb{N}^k$ to $\{0, 1\}$ defined by:

$$\chi_A(n_1, \ldots, n_k) = \begin{cases} 1 & \text{if } (n_1, \ldots, n_k) \in A \\ 0 & \text{if} (n_1, \ldots, n_k) \notin A \end{cases}$$

Definition 10.3.14. A set $A \subseteq \mathbb{N}^k$ is *primitive recursive* iff its characteristic function is primitive recursive.

10.3.2 Extensions of primitive recursion

Note that in Definition 10.3.10, when defining a function by primitive recursion, only the value of the immediately preceding argument can be used. A natural way to generalise this is by allowing the value of more than one earlier argument to be used. This is called *course-of-values* recursion and quite old examples of this can be found, the most well-known being the Fibonacci function, introduced by Leonardo da Pisa in his early 13th-century *Liber abaci* in response to the question "How many pairs of rabbits can be bred in one year from

one pair?"[4] The function that Leonardo defines is as follows:

$$\begin{aligned} f(0) &= 0 \\ f(1) &= 1 \\ f(n+2) &= f(n) + f(n+1) \end{aligned}$$

Fibonacci sequences can be found all over nature and artifice.

Another way in which primitive recursion can be generalized is by allowing recursion on more than one variable instead of just one; examples of this type of generalization can be found at least as far back as Archimedes.

In general (but not always), both course-of-values recursion and double recursion can be reduced to ordinary primitive recursion: So we don't actually gain much by these generalizations beyond an easier method of definition. The exceptions are functions defined by double recursion which are recursive but not primitive recursive; we give examples of such functions in the next section.

10.3.3 Non-primitive recursive, but recursive, functions

Not all total recursive functions are primitive recursive, as noted before; these functions tend to be highly specialized and artificial in nature. Three examples include (1) the Ackermann function, (2) the Sudan function, and (3) the Goodstein function. The details of how these functions are defined and proven to be non-primitive-recursive are not covered here, but we sketch the functions because they are interesting.

Ackermann function

The Ackermann function is one of the earliest-discovered examples of a function which is total and recursive but not primitive recursive. Since Ackermann introduced the function in 1928, various versions have been given. Here is a straightforward one due to Ackermann and Péter. For $m, n \in \mathbb{N}$:

$$A(m,n) = \begin{cases} n+1 & \text{if } m = 0 \\ A(m-1,1) & \text{if } m > 0 \text{ and } n = 0 \\ A(m-1, A(m,n-1)) & \text{if } m > 0 \text{ and } n > 0 \end{cases}$$

Note that this is an example of double recursion.

[4] Answer: Lots.

This is the version of the function most commonly known simply as "the Ackermann" function, but the original function took three values as inputs, and was defined to incorporate all of the arithmetic operations of addition, multiplication, and exponentiation as well as hyper-arithmetic ones (I can't really tell you much about these, sadly, but the Wikipedia article on them at `https://en.wikipedia.org/wiki/Hyperoperation` is quite good).

How does one prove that a function is *not* primitive recursive? It would be extremely difficult to prove that there is no way to construct a given function from the initial ones via composition and primitive recursion — how do you know that you merely haven't figured out the right combinations? The way it is proved that this is recursive without being primitive recursive is to find a property that all primitive recursive functions have which the Ackermann function doesn't have. A sketch of such a proof can be seen at `http://planetmath.org/ackermannfunctionisnotprimitiverecursive`. Essentially, it is possible to provide a bound on how fast a primitive recursive function grows (that is, as the inputs get bigger, how much bigger do the outputs get), and the Ackermann function violates this bound. (In fact, if you're ever bored, compute $A(4,4)$. When you're done, let me know the answer.)

Sudan function

The Sudan function is the earliest published example of a recursive-but-not-primitive-recursive function, published in 1927. It is defined as follows:

$$\begin{aligned} F_0(x,y) &= x+y \\ F_{n+1}(x,0) &= x \\ F_{n+1}(x,y+1) &= F_n(F_{n+1}(x,y), F_{n+1}(x,y)+y+1) \end{aligned}$$

Again, if you're bored, try calculating $F_2(2,2)$.

Goodstein function

The Goodstein function g is defined by $g(n)$ being the length of the Goodstein sequence beginning with n. The Goodstein sequence $G(n)$ is a sequence of natural numbers such that (a) the first element of the sequence is n and (b) the $n+1$th entry in the sequence is obtained by (i) writing n in hereditary base-$n+1$ notation[5], (ii) change all $n+1$s to $n+2$s, and (iii) substract 1 from the result. At some point, the result will be 0, after which the sequence terminates. $G(3)$ is 333210, and hence $g(3) = 6$.

[5]Hereditary base-m notation is a derivative of base-m notation: Once you've written a number in base m notation, then write all of the exponents in base m notation, and so on until you can't any more.

Goodstein proved that every Goodstein sequence terminates (i.e., eventually reach 0). Interestingly, this theorem is *not* provable in PA (this is equivalent to saying that PA is unable to prove that $g(n)$ is total).

10.4 Proof theory

Having introduced the basic language and defined some important syntactic and semantic notions concerning the language, we can now move on to the next component of our theory: axioms!

Peano Arithmetic has eight axioms, two for each of the functions, one that says what to do in the 0 case and one that says what to do in the successor case, and one axiom schema.

The axioms for $\mathtt{S}$ are as follows:

Axiom 10.4.1 (PA1)**.** $\forall x(\mathtt{S}x \neq 0)$.

Axiom 10.4.2 (PA2)**.** $\forall x \forall y(\mathtt{S}x = \mathtt{S}y \to x = y)$.

One might ask why PA2 is a simple conditional and not a biconditional. We do not need to have the right-to-left version because it can already be proven from the laws of identity, without any appeal to any specific properties about $\mathtt{S}$ (other than that it is a function):

Fact 10.4.3. $\vdash_{\mathrm{PA}} \forall x \forall y(x = y \to \mathtt{S}x = \mathtt{S}y)$

Proof.

1	$\mathtt{S}x = \mathtt{S}x$	Refl
2	$x = y$	Assumption
3	$\mathtt{S}x = \mathtt{S}y$	SubsId, 1, 2
4	$x = y \to \mathtt{S}x = \mathtt{S}y$	→I, 2–3
5	$\forall y(x = y \to \mathtt{S}x = \mathtt{S}y)$	∀I, 4
6	$\forall x \forall y(x = y \to \mathtt{S}x = \mathtt{S}y)$	∀I, 5

□

From these two axioms, PA1 and PA2, alone, we can already prove a property of $\mathbb{N}$:

Lemma 10.4.4. *Any model satisfying these axioms must be infinite.*

Proof. See Exercise 96. □

Next, we have two axioms governing how addition works:

Axiom 10.4.5 (PA3)**.** $\forall x(x + 0 = x)$.

Axiom 10.4.6 (PA4)**.** $\forall x \forall y(x + \mathsf{S}y = \mathsf{S}(x + y))$.

These two axioms alone are sufficient for us to prove basic arithmetic facts, such as $\underline{2} + \underline{2} = \underline{4}$:

Fact 10.4.7. $\vdash_{\mathrm{PA}} \underline{2} + \underline{2} = \underline{4}$

Proof.

1	$\mathsf{SS0} + \mathsf{SS0} = \mathsf{S}(\mathsf{SS0} + \mathsf{S0})$	PA4
2	$\mathsf{SS0} + \mathsf{S0} = \mathsf{S}(\mathsf{SS0} + 0)$	PA4
3	$\mathsf{SS0} + 0 = \mathsf{SS0}$	PA3
4	$\mathsf{SS0} + \mathsf{S0} = \mathsf{SSS0}$	SubsId, $2, 3$
5	$\mathsf{SS0} + \mathsf{SS0} = \mathsf{SSSS0}$	SubsId, $1, 4$

□

It is clear that this method can be generalised: Through successive applications of PA4 we can strip the occurrences of S from the right-hand number and pull them out in front of the parentheses, until we are in a position to apply PA3. Then it is merely a matter of successively substituting identities until we reach our desired goal. We can thus prove generally that:

Lemma 10.4.8. *For any natural numbers n and m, $\vdash_{\mathrm{PA}} \underline{m} + \underline{n} = \underline{m + n}$.*

Proof. The proof is by induction.

Basis case: Let $n = 0$. Then we need to show: $\vdash_{\mathrm{PA}} \underline{m} = \underline{0} = \underline{m + 0}$. First, note that $0 = \underline{0}$, by how we defined the underline notation, and that $m = m + 0$ is a meta-language truth. Then:

1	$\underline{m} + 0 = \underline{m}$	PA3
2	$\underline{m} + 0 = \underline{m + 0}$	SubsId, 1

Inductive step: Note that $\mathsf{S}\underline{n} = \underline{n + 1}$, by how we've defined the underline notation. We assume that $\vdash_{\mathrm{PA}} \underline{m} + \underline{n} = \underline{m + n}$, and show that $\vdash_{\mathrm{PA}} \underline{m} + \mathsf{S}\underline{n} = \underline{m + n + 1}$.

1	$\underline{m} + \underline{n} = \underline{m + n}$	Inductive Hypothesis
2	$\underline{m} + \underline{n + 1} = \mathsf{S}\underline{m + n}$	PA4
3	$\mathsf{S}(\underline{m} + \underline{n}) = \mathsf{S}(\underline{m + n})$	SubsId, 1
4	$\underline{m} + \underline{n + 1} = \mathsf{S}(\underline{m + n})$	SubsId, $2, 3$

□

Next we have two axioms governing how multiplication works:

Axiom 10.4.9 (PA5)**.** $\forall x(x \times \mathsf{0} = \mathsf{0})$.

Axiom 10.4.10 (PA6)**.** $\forall x(x \times \mathsf{S}y = (x \times y) + x)$.

A similar result to Lemma 10.4.8 can be proven for multiplication:

Lemma 10.4.11. $\vdash_{\mathrm{PA}} \underline{m} \times \underline{n} = \underline{m \times n}$.

Proof. The proof of this is analogous to the proof of Lemma 10.4.8, and also appeals to it. □

Note that the axiom governing how multiplication interacts with S involves addition. A similar thing happens with exponentiation and multiplication:

Axiom 10.4.12 (PA7)**.** $\forall x(x^{\mathsf{0}} = \mathsf{S0})$.

Axiom 10.4.13 (PA8)**.** $\forall x(x^{\mathsf{S}y} = x^y \times x)$.

Similarly to Lemmas 10.4.8 and 10.4.11, we can also prove the following:

Lemma 10.4.14. $\vdash_{\mathrm{PA}} \underline{m^n} = \underline{m}^{\underline{n}}$.

Proof. Exercise for the reader. □

These results generalise:

Lemma 10.4.15. *If τ is a closed term, then $I(\tau) = n$ for some $n \in \mathbb{N}$, and $\vdash_{\mathrm{PA}} \tau = \underline{n}$.*

Proof. Note: If τ is a closed term, then it contains no variables; that is, it is either a constant or the application of S to a closed term. Our language has only one constant, $\mathsf{0}$.

Basis case: We must prove that this fact holds τ which is atomic and closed. If τ is atomic and closed, then it must be a constant. The only constant in the language is $\mathsf{0}$, and $I(\mathsf{0}) = 0$, and $0 \in \mathbb{N}$, so we must now show $\vdash_{\mathrm{PA}} \mathsf{0} = \underline{0}$. Per Definition 10.2.8, $\underline{0} = \mathsf{0}$, so all we need to show is that $\vdash_{\mathrm{PA}} \mathsf{0} = \mathsf{0}$, which follows directly from the reflexivity of $=$, and hence $\vdash_{\mathrm{PA}} \mathsf{0} = \underline{0}$, as required.

Inductive step:

- Let $\tau = \mathsf{S}\tau_1$ for some τ_1 such that $I(\tau_1) = n$:

1 $\tau_1 = \underline{n}$	Assumption
2 $\mathsf{S}\tau_1 = \mathsf{S}\underline{n}$	functionality of §, 1
3 $\mathsf{S}\underline{n} = \underline{n+1}$	Definition 10.2.8
4 $\tau = \underline{n+1}$	SubsId, 2, 3

  And $I(\tau) = I(\mathsf{S}\tau_1) = n + 1$, as desired.

- Let $\tau = \tau_1 \circ \tau_2$ where $\circ \in \{+, \times, \uparrow\}$ and where τ_1 and τ_2 are closed terms (and hence the inductive hypothesis applies to them). Then $\vdash_{\mathrm{PA}} \tau = \underline{\tau_1 \circ \tau_2}$ by Lemmas 10.4.8, 10.4.11, and 10.4.14.

Since it holds for every atomic closed terms, and assuming it holds for some closed term we can show that it holds for any term of one step greater complexity, we can conclude that this holds for all terms. □

PA1–PA8 are sufficient to determine the arithmetic operations, but they are not alone sufficient for determining the structure of $\mathbb{N}$, since there are models of PA1–PA8 which are not isomorphic to $\mathbb{N}$. We now give an example of such a model.

Example 10.4.16. Let $\mathbb{N}$ be the set of natural numbers, Γ a set with at least two distinct elements, and $a_0 \in \Gamma$. We now define a model $\mathfrak{M}$ where

- The domain of $\mathfrak{M}$ is $\mathbb{N} \times \Gamma$ (that is, it is the set of ordered pairs where the first element of the pair is a natural number and the second element of the pair is a member of Γ).
- $\mathsf{0}^{\mathfrak{M}} = \langle 0, a_0 \rangle$.
- $\mathsf{S}^{\mathfrak{M}} = \langle n, a \rangle \mapsto \langle n + 1, a \rangle$.
- $\langle n, a \rangle +^{\mathfrak{M}} \langle n', a' \rangle = \langle n + n', a \rangle$
- $\langle n, a \rangle \times^{\mathfrak{M}} \langle n', a' \rangle = \langle n \times n', a \rangle$
- $\langle n, a \rangle \uparrow^{\mathfrak{M}} \langle n', a' \rangle = \langle n^{n'}, a \rangle$

Fact 10.4.17. In this model, PA1–PA8 are all true, but + is not commutative.

Proof. When a and a' are distinct elements of A, $\langle n, a \rangle +^{\mathfrak{M}} \langle n', a' \rangle = \langle n + n', a \rangle \neq \langle n', a' \rangle +^{\mathfrak{M}} \langle n, a \rangle = \langle n + n', a' \rangle$. □

What we need is a further axiom that we can add to PA that excludes strange elements like these, essentially saying that every element of our model is a natural number and that there is nothing illegitimate sneaking in.

One option would be to state explicitly that every element in the domain is either 0, or 1, or 2, etc.:

$$\forall x(x = 0 \vee x = \underline{1} \vee x = \underline{2} \vee x = \underline{3} \dots)$$

However, the very presence of the 'etc.' indicates the problem with this approach: Such a sentence would be infinitary, and violate the rules for well-formed formulas given in Definition 8.2.6.

Another option would be allow quantification over not only individual numbers, but also sets of numbers, and to add some basic set theoretic notation (cf. §1.3.3) to our language. Then, we could add the single axiom:

$$\forall \Gamma((0 \in \Gamma \wedge \forall x(x \in \Gamma \to \mathtt{S}x \in \Gamma)) \to \forall x(x \in \Gamma))$$

which says that of any subset of our domain, if 0 is in the subset, and any number's being in the set implies its successor is also in the set, then every number is in the set.

However, neither the first option (introducing infinitary wffs) nor the second option (introducing second-order quantification) meets our desiderata for a *finitistic* and *effective* proof system: For neither infinitary logic nor second-order logic admit of a well-defined, satisfactory proof system. Thus, either of these options brings in far more problems than it solves.

What we want to do we cannot do via a single axiom; instead, we introduce two *axiom schemata*. The first is for formulas with one free variable:

Definition 10.4.18. For any formula $\varphi(x)$ with free variable x, let IND_φ be the formula:

$$\varphi(\mathsf{0}) \wedge \forall x(\varphi(x) \to \varphi(\mathsf{S}(x)) \to \forall x \varphi(x)$$

The second is a generalization of this:

Definition 10.4.19. If $\varphi(x_0, \dots, x_n)$ is a formula with free variables $x_0, \dots, x_n$, let IND_φ be the formula:

$$\forall x_1, \dots, x_n(\varphi(\mathsf{0}, x_1, \dots, x_n) \wedge \forall x_0(\varphi(x_0, x_1, \dots, x_n) \to \varphi(\mathsf{S}x_0, x_1, \dots, x_n)) \to \forall x_0 \varphi(x_0, x_1, \dots, x_n))$$

These collectively give us our 9th 'axiom', which we will refer to as the 'induction axiom':

Axiom 10.4.20 (PA9)**.** PA9 *is the list of all axioms of the form* IND_φ*.*

The induction axiom allows us to exploit a very powerful proof method, namely, proof by (mathematical) induction. If we can prove that some particular formula $\varphi(x)$ holds of 0 (the basis case), and we can prove that if $\varphi(x)$ holds then $\varphi(\mathtt{S}x)$ also holds (the inductive step), then we can conclude, by invoking one of the induction axioms, that $\forall x \varphi(x)$. We will see examples of such proofs in the next section.

10.4.1 Some historical remarks: An aside

Though we call these *Peano*'s axioms for arithmetic, in truth, he was not the first person to articulate an axiomatic approach to basic number theory, and his developments owe a great deal to earlier developments by Dedekind and Grassmann. [Wang, 1957] discusses this debt of Peano's, as he considers the question of "Where do axioms come from?" This isn't just an abstract question, but a historical/conceptual one. There are a variety routes you might adopt:

- Start from typical proofs and results, and work backwards to determine what the underlying assumptions are.
- Pick some reasonable assumptions, and adopt them until they are shown to be inconsistent.
- Prove what you can, and when you get stuck, add what you need as an axiom.

The focus of Wang's paper is Peano's axiomatization borrowed from Dedekind and Grassmann [Wang, 1957, p. 145]. All three approaches were rooted in a desire to make

> an explicit statement of some adequate group of natural rules and conventions which enables us to justify all the true numerical formulae containing 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, +, ×, =, (,) [Wang, 1957, p. 146][6]

Wang describes Grassmann's *Lehrbuch der Arithmetik* (1861) as "the first serious and rather successful attempt to put numbers on a more or less axiomatic basis" [Wang, 1957, p. 147]; his scope covered not only the non-negative integers but also the negative ones. Wang shows how the contemporary (i.e., in the 1950s) characterisation of the integers in abstract algebra, a system he calls L_1, can be derived from Grassmann's system (which he calls L_2). L_1 is characters by the commutativity and associativity of + and ×, the distribution of × over +, the fact that 0 is the identity for + and 1 for ×, that $a + (-a) = 0$, that if $c \neq 0$ and $ca = cb$, then $a = b$, that sums and products of positive numbers are positive numbers, that

[6] Note the lack of exponentiation.

every number is either positive, not positive, or 0, and a version of mathematical induction. Wang points out a drawback of Grassmann's L_2, which is that it does not require distinct integers to have distinct successors, and hence L_2 has models consisting in only a single object [Wang, 1957, p. 149]. This requirement was made explicit in Peano, whose system contained the basic concepts of 1, number, and successor, and the following five axioms governing these concepts:

1. 1 is a number.
2. The successor of any number is a number.
3. No two numbers have the same successor.
4. 1 is not the successor of any number.
5. Any property which belongs to 1, and also to the successor of every number which has the property, belongs to all numbers.

(Nowadays, presentations of PA start from 0, as opposed to 1; in this, Frege's account of numbers differs from Dedekind's in that Frege did begin with 0.) These axioms were taken from Dedekind's essay *Was sind und was sollen die Zahlen?* (1888) [Wang, 1957, p. 149], and Dedekind's inspiration for these axioms is preserved in a two-page letter that Wang quotes (in translation) [Wang, 1957, pp. 150–151]. What's important is that these are an axiomatisation *of the concept of number only*—there's nothing here to cover the arithmetic operations. These (addition, multiplication, and exponentiation) Dedekind defines later in the essay.

10.5 Number theory in Peano Arithmetic

In this section we show how a number of ordinary number theoretic results can be proven in PA. Along the way, we will introduce a variety of new relations and predicates. These will not be primitive elements of our language, but defined using symbols that are primitive. This is merely a matter of convenience of notation, and in our proofs we will routinely switch between the introduced abbreviations and the complex formulas that they abbreviate. First, we introduce a new notion of equivalence, namely that of *provable* equivalence:

Definition 10.5.1 (Provable equivalence)**.** Two formulas φ and ψ are *provably equivalent* if

$$\vdash_{\mathrm{PA}} \varphi \quad \text{iff} \quad \vdash_{\mathrm{PA}} \psi$$

This is a stronger notion of equivalence than the semantic equivalence defined in Definition 10.2.15, because two things can be semantically equivalent without it being provable that they are.

Next, we discuss briefly some of the general proof strategies used to prove the results in this section. Almost all of the proofs will fall into one of four kinds: (a) proof by induction; (b) proof by contradiction; (c) proof by generalization; (d) proof by conditionalization. Of course, many of the proofs will involve proof steps of all four types, but each proof will generally invoke one of these methods as the main method for getting started with the proof. Recognizing which proof method is the right main method is one of the most important parts of mastering formal proofs. A simple proof by induction is given in the proof of Lemma 10.4.15. Another example is the following, invoking our new definition of provable equivalence:

Lemma 10.5.2. *Every closed bounded formula φ is provably equivalent to a quantifier-free formula $\hat{\varphi}$.*

Proof. Proof by induction.

Basis case: Let φ be a closed atomic formula. Then φ contains one relation symbol and one or more constants, and no variables. Per Definition 10.2.12, φ is a bounded formula. Because it is atomic, it is also quantifier free. Thus $\hat{\varphi} = \varphi$ and $\vdash_{\mathrm{PA}} \varphi \leftrightarrow \varphi$, so $\vdash_{\mathrm{PA}} \hat{\varphi} \leftrightarrow \varphi$.

Inductive step: Left as an exercise for the reader. □

The following lemma is also proven by induction, but it underpins another method of proof, namely proof by cases:

Lemma 10.5.3. *Every number is either 0 or a successor:* $\vdash_{\mathrm{PA}} \forall x(x = 0 \vee \exists y x = \mathsf{S}y)$.

Proof. By induction.

Basis case:

1	$0 = 0$	Refl
2	$0 = 0 \vee \exists y(x = \mathsf{S}y)$	$\vee$I, 1

Inductive step:

$3\ \mathsf{S}x = \mathsf{S}x$	Refl
$4\ \exists y(\mathsf{S}x = \mathsf{S}y)$	$\exists$I, 3
$5\ \mathsf{S}x = 0 \vee \exists y(\mathsf{S}x = \mathsf{S}y)$	$\vee$I, 4
$\quad 6\ x = 0 \vee \exists y(x = \mathsf{S}y)$	Assumption
$\quad 7\ \mathsf{S}x = 0 \vee \exists y(\mathsf{S}x = \mathsf{S}y)$	Reit, 5
$8\ (x = 0 \vee \exists y(x = \mathsf{S}y)) \rightarrow (\mathsf{S}x = 0 = \vee\exists y(\mathsf{S}x = \mathsf{S}y))$	$\rightarrow$I, 6–7
$9\ x = 0 = \vee\exists y(x = \mathsf{S}y)$	$\rightarrow$E, PA9, 2, 6
$10\ \forall x(x = 0 = \vee\exists y(x = \mathsf{S}y))$	$\forall$E, 7

□

One thing to notice about this proof is that it begins to get clunky being typeset/formatted according to the proof rules given in Chapter 8. From here on out, we are going to slightly modify the natural-deduction proofs defined in that chapter by reformatting them. Instead of nesting our assumptions within subproofs, we will write any assumptions that we have to the left of $\vdash_{\mathrm{PA}}$, with what we can derive from those assumptions written on the right. We will also allow ourselves to write down predicate logic tautologies at any time, as we do in Chapter 11. That is, we can rewrite the above proof as follows:

3	$\mathsf{S}x = \mathsf{S}x$	Refl
4	$\exists y(\mathsf{S}x = \mathsf{S}y)$	$\exists$I, 3
5	$\mathsf{S}x = 0 \vee \exists y(\mathsf{S}x = \mathsf{S}y)$	$\vee$I, 4
6	$(x = 0 \vee \exists y(x = \mathsf{S}y)) \rightarrow (\mathsf{S}x = 0 = \vee\exists y(\mathsf{S}x = \mathsf{S}y))$	tautology, 5
7	$x = 0 = \vee\exists y(x = \mathsf{S}y)$	$\rightarrow$ E, PA9, 2, 6
8	$\forall x(x = 0 = \vee\exists y(x = \mathsf{S}y))$	$\forall$I, 7

Note that this way of writing it also requires fewer lines.

Lemma 10.5.3 allows us to prove certain results below *by cases*: We know that every number is either 0 or a successor, so if we can prove that something holds of 0, and that it holds for an arbitrarily chosen successor, then we can conclude that the property holds of all numbers, by appeal to this Lemma. We give an example of such a proof by cases below:

Lemma 10.5.4. $\vdash_{\mathrm{PA}} \forall xy(x + y = 0 \rightarrow y = 0)$.

Proof. Proof by cases.

Zero case: $y = 0$.

1	$y = 0$	$\vdash_{\mathrm{PA}}$	$x + y = 0 \to y = 0$	tautology
2		$\vdash_{\mathrm{PA}}$	$y = 0 \to (x + y = 0 \to y = 0)$	DedThm, 1

Successor case: $y = \mathsf{S}t$ for some t.

3	$y = \mathsf{S}t$	$\vdash_{\mathrm{PA}}$	$x + y = x + \mathsf{S}t$	SubsId
4	$y = \mathsf{S}t$	$\vdash_{\mathrm{PA}}$	$x + y = \mathsf{S}(x + t)$	PA4
5	$y = \mathsf{S}t$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}(x + t) \neq 0$	PA1
6	$y = \mathsf{S}t$	$\vdash_{\mathrm{PA}}$	$x + y \neq 0$	SubsId, $4, 5$
7	$y = \mathsf{S}t$	$\vdash_{\mathrm{PA}}$	$x + y = 0 \to y = 0$	tautology
8		$\vdash_{\mathrm{PA}}$	$y = \mathsf{S}t \to (x + y = 0 \to y = 0)$	DedThm, 7
9		$\vdash_{\mathrm{PA}}$	$\exists t(y = \mathsf{S}t \to (x + y = 0 \to y = 0))$	$\exists$I, 8
10		$\vdash_{\mathrm{PA}}$	$y = 0 \vee \exists t(y = \mathsf{S}t)$	Lemma 10.5.3
11		$\vdash_{\mathrm{PA}}$	$x + y = 0 \to y = 0$	tautology, $2, 9, 10$
13		$\vdash_{\mathrm{PA}}$	$\forall y(x + y = 0 \to y = 0)$	$\forall$I, 12
14		$\vdash_{\mathrm{PA}}$	$\forall xy(x + y = 0 \to y = 0)$	$\forall$I, 13

□

A very similar result can be proven:

Lemma 10.5.5. $\vdash_{\mathrm{PA}} \forall xy(x + y = x \to y = 0)$

Lemma 10.5.6. *Every non-zero number is the sum of two numbers of which at least one is non-zero:* $\vdash_{\mathrm{PA}} \forall x(0 < x \to \exists yz(y + z = x \wedge (y \neq 0 \vee z \neq 0)))$

Definition 10.5.7.

$$\mathrm{prime}(x) \quad := \quad \underline{1} < x \wedge \forall y \leq x(y | x \to y = \underline{1} \vee y = x)$$

Fact 10.5.8. Let *Prime* be the set defined by the formula $\mathrm{prime}(x)$. Then $\mathbb{N} \vDash \mathrm{prime}(n)$ iff $n \in \mathit{Prime}$.

Using this definition of 'prime', we can introduce a binary relation that adjacent primes stand in:

Definition 10.5.9.

$$\mathrm{nextprime}(x, y) \quad := \quad \mathrm{prime}(x) \wedge \mathrm{prime}(y) \wedge x < y \wedge \forall z(\mathrm{prime}(z) \to z \leq x \vee z \geq y)$$

That is, y is the next prime after y iff x and y are both primes, x is smaller than y, and every other prime is either less than or equal to x or greater than or equal to y (that is, there is no prime between the two).

The following theorem is a consequence of the Fundamental Theorem of Arithmetic, which states that every number has a unique prime factorisation.

Lemma 10.5.10. *Every number greater than* 1 *has at least one prime divisor:* $\vdash_{\mathrm{PA}} \forall x(x > \mathsf{S0} \to \exists y(\mathit{prime}(y) \wedge y|x))$

Lemma 10.5.11. *No number is identical to its successor:* $\vdash_{\mathrm{PA}} \forall x(x \neq \mathsf{S}x)$

Lemma 10.5.12. *No number is greater than its successor:* $\vdash_{\mathrm{PA}} \forall x\neg(x > \mathsf{S}x)$

Corollary 10.5.13. *Every number is strictly less than its successor:* $\vdash_{\mathrm{PA}} \forall x(x < \mathsf{S}x)$.

Proof. An immediate consequence of Lemmas 10.5.11 and 10.5.12. □

Lemma 10.5.14. *At least one divisor of any prime number is* 1*:* $\vdash_{\mathrm{PA}} \forall x(\mathit{prime}(x) \to \exists yz(y \times z = x \to (y = \mathsf{S0} \vee z = \mathsf{S0})))$

Lemma 10.5.15. $\vdash_{\mathrm{PA}} \forall x(\mathsf{0} + x = x)$.

Proof. By induction on x.

Basis case:

1		$\vdash_{\mathrm{PA}}$	$\mathsf{0} + \mathsf{0} = \mathsf{0}$	PA3

Inductive step:

2	$\mathsf{0} + x = x$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}(\mathsf{0} + x) = \mathsf{S}x$	Fact 10.4.3
3	$\mathsf{0} + x = x$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}(\mathsf{0} + x) = \mathsf{0} + \mathsf{S}x$	PA4
4	$\mathsf{0} + x = x$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}x = \mathsf{0} + \mathsf{S}x$	Transitivity, 2, 3
5		$\vdash_{\mathrm{PA}}$	$\mathsf{0} + x = x \to \mathsf{0} + \mathsf{S}x = \mathsf{S}x$	DedThm 4

□

Lemma 10.5.16. $\vdash_{\mathrm{PA}} \forall x \forall y \mathsf{S}x + y = \mathsf{S}(x + y)$.

Proof. By induction on y.

Basis case:

1	$\vdash_{\mathrm{PA}}$	$x + \mathsf{0} = x$	PA3
2	$\vdash_{\mathrm{PA}}$	$\mathsf{S}(x + \mathsf{0}) = \mathsf{S}x$	Fact 10.4.3, 1
3	$\vdash_{\mathrm{PA}}$	$\mathsf{S}x + \mathsf{0} = \mathsf{S}x$	PA3
4	$\vdash_{\mathrm{PA}}$	$\mathsf{S}x + \mathsf{0} = \mathsf{S}(x + \mathsf{0})$	Transitivity, 2, 3

Inductive step:

5	$\mathsf{S}x + y = \mathsf{S}(x + y)$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}(\mathsf{S}x + y) = \mathsf{S}(\mathsf{S}x + y)$	PA4
6	$\mathsf{S}x + y = \mathsf{S}(x + y)$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}x + \mathsf{S}y = \mathsf{SS}(x + y)$	Transitivity, 5
7	$\mathsf{S}x + y = \mathsf{S}(x + y)$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}x + \mathsf{S}y = \mathsf{S}(x + \mathsf{S}y)$	PA4, 6
8		$\vdash_{\mathrm{PA}}$	$\mathsf{S}x + y = \mathsf{S}(x + y) \to \mathsf{S}x + \mathsf{S}y = \mathsf{S}(x + \mathsf{S}y)$	DedThm, 7

From here, the proof is completed via universal generalisation, PA9, and another application of universal generalisation. □

Lemma 10.5.17 (Commutativity of +)**.** $\vdash_{\mathrm{PA}} \forall x \forall y (x + y = y + x)$.

Proof. By induction on y.

Basis case:

1	$\vdash_{\mathrm{PA}}$	$x + 0 = x$	PA3
2	$\vdash_{\mathrm{PA}}$	$0 + x = x$	Lemma 10.5.15
3	$\vdash_{\mathrm{PA}}$	$x + 0 = 0 + x$	Transitivity, 1, 2

Inductive step:

4	$x + y = y + x$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}(x + y) = \mathsf{S}(y + x)$	Fact 10.4.3
5	$x + y = y + x$	$\vdash_{\mathrm{PA}}$	$x + \mathsf{S}y = \mathsf{S}(y + x)$	PA4, 4
6	$x + y = y + x$	$\vdash_{\mathrm{PA}}$	$x + \mathsf{S}y = \mathsf{S}y + x$	Lemma 10.5.16, 5

The proof is finished in the same manner as the proof of Lemma 10.5.16. □

Lemma 10.5.18 (Commutativity of ×)**.** $\vdash_{\mathrm{PA}} \forall xy (x \times y = y \times x)$

Lemma 10.5.19. 1 *is the multiplicative identity:* $\vdash_{\mathrm{PA}} \forall x (x \times \mathsf{S0} = x)$

Proof.

1	$\vdash_{\mathrm{PA}}$	$x \times \mathsf{S0} = (x \times 0) + x$	PA6
2	$\vdash_{\mathrm{PA}}$	$x \times \mathsf{S0} = 0 + x$	PA5, 1
3	$\vdash_{\mathrm{PA}}$	$x \times \mathsf{S0} = x + 0$	Lemma 10.5.17
4	$\vdash_{\mathrm{PA}}$	$x \times \mathsf{S0} = x$	PA3
5	$\vdash_{\mathrm{PA}}$	$\forall x (x \times \mathsf{S0} = x)$	∀I, 4

□

Lemma 10.5.20. $\vdash_{\mathrm{PA}} \forall xy (x \times y = x \to y = \mathsf{S0})$

Lemma 10.5.21 (Associativity of +)**.** $\vdash_{\mathrm{PA}} \forall xyz ((x + y) + z = x + (y + z))$

Proof. By induction on z.

Basis case:

1	$\vdash_{\mathrm{PA}}$	$(x + y) + 0 = x + y$	PA3
2	$\vdash_{\mathrm{PA}}$	$y + 0 = y$	PA3
3	$\vdash_{\mathrm{PA}}$	$(x + y) + 0 = x + (y + 0)$	SubsId, $1, 2$

Inductive step:

4	$(x + y) + z = x + (y + z)$	$\vdash_{\mathrm{PA}}$	$(x + y) + \mathsf{S}z = \mathsf{S}((x + y) + z)$	PA4
5	$(x + y) + z = x + (y + z)$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}((x + y) + z) = \mathsf{S}(x + (y + z))$	SubsId
6	$(x + y) + z = x + (y + z)$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}(x + (y + z)) = x + \mathsf{S}(y + z)$	PA4
7	$(x + y) + z = x + (y + z)$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}(y + z) = y + \mathsf{S}z$	PA4
8	$(x + y) + z = x + (y + z)$	$\vdash_{\mathrm{PA}}$	$\mathsf{S}(x + (y + z)) = x + (y + \mathsf{S}z)$	Transitivity, $6, 7$
9	$(x + y) + z = x + (y + z)$	$\vdash_{\mathrm{PA}}$	$(x + y) + \mathsf{S}z = x + (y + \mathsf{S}z)$	Transitivity, $4, 5, 8$

The proof is finished in the same manner as the proof of Lemma 10.5.16. □

Lemma 10.5.22 (Associativity of $\times$)**.** $\vdash_{\mathrm{PA}} \forall xyz((x \times y) \times z = x \times (y \times z))$

Lemma 10.5.23 (Distribution of $\times$ over $+$)**.** $\vdash_{\mathrm{PA}} \forall xyz(x \times (y + z) = (x \times y) + (x \times z))$

Proof. By induction on z.

Basis case:

1	$\vdash_{\mathrm{PA}}$	$x \times (y + 0) = x \times y$	PA3
2	$\vdash_{\mathrm{PA}}$	$x \times y = (x \times y) + 0$	PA3
3	$\vdash_{\mathrm{PA}}$	$(x \times y) + 0 = (x \times y) + (x \times 0)$	PA5
4	$\vdash_{\mathrm{PA}}$	$x \times (y + 0) = (x \times y) + (x \times 0)$	Transitivity, $1, 2, 3$

Inductive step:

5	$x \times (y + z) = (x \times y) + (x \times z)$	$\vdash_{\mathrm{PA}}$	$x \times (y + \mathsf{S}z) = x \times \mathsf{S}(y + z)$	PA4
6	$x \times (y + z) = (x \times y) + (x \times z)$	$\vdash_{\mathrm{PA}}$	$x \times \mathsf{S}(y + z) = (x \times (y + z)) + x$	PA6
7	$x \times (y + z) = (x \times y) + (x \times z)$	$\vdash_{\mathrm{PA}}$	$x \times \mathsf{S}(y + z) = ((x \times y) + (x \times z)) + x$	SubsId, 6
8	$x \times (y + z) = (x \times y) + (x \times z)$	$\vdash_{\mathrm{PA}}$	$x \times \mathsf{S}(y + z) = (x \times y) + ((x \times z) + x)$	Lemma10.5.21, 7
9	$x \times (y + z) = (x \times y) + (x \times z)$	$\vdash_{\mathrm{PA}}$	$x \times \mathsf{S}(y + z) = (x \times y) + (x \times \mathsf{S}z)$	PA6, 8
10	$x \times (y + z) = (x \times y) + (x \times z)$	$\vdash_{\mathrm{PA}}$	$x \times (y + \mathsf{S}z) = (x \times y) + (x \times \mathsf{S}z)$	Transitivity, $5, 9$

The proof is finished in the same manner as the proof of Lemma 10.5.16.

□

Lemma 10.5.24 (Transitivity of $\leq$)**.** $\vdash_{\mathrm{PA}} \forall xyz(x \leq y \wedge y \leq z \rightarrow x \leq z)$

Proof. By conditionalization.

1	$x + r = y, y + t = z$	$\vdash_{\mathrm{PA}}$	$(x + r) + t = z$	Identity
2	$x + r = y, y + t = z$	$\vdash_{\mathrm{PA}}$	$x + (r + t) = z$	Lemma 10.5.21
3	$x + r = y, y + t = z$	$\vdash_{\mathrm{PA}}$	$\exists s x + s = z$	$\exists \mathsf{I}, 2$
4	$x + r = y, y + t = z$	$\vdash_{\mathrm{PA}}$	$x \leq z$	Def. of $\leq$
5		$\vdash_{\mathrm{PA}}$	$x \leq y \wedge y \leq z \rightarrow x \leq z$	DedThm, 4
6		$\vdash_{\mathrm{PA}}$	$\forall x y z (x \leq y \wedge y \leq z \rightarrow x \leq z)$	$\forall \mathsf{I}, 5$

□

A similar result holds of the strict version of the relation:

Lemma 10.5.25 (Transitivity of $<$)**.** $\vdash_{\mathrm{PA}} \forall x y z (x < y \wedge y < z \rightarrow x < z)$.

The following lemma says that there is no number preceding 0:

Lemma 10.5.26. $\vdash_{\mathrm{PA}} \forall x (\neg x < \mathsf{0}))$

Lemma 10.5.27. $\vdash_{\mathrm{PA}} \forall x y (x < \mathsf{S} y \leftrightarrow (x < y \vee x = y)$

Lemma 10.5.28. $\vdash_{\mathrm{PA}} \forall x y z (x \leq y \leq \mathsf{S} x \rightarrow (x = y \vee y = \mathsf{S} x))$

Lemma 10.5.29 (Transitivity of $|$)**.** $\vdash_{\mathrm{PA}} \forall x y z ((x|y \wedge y|z) \rightarrow x|z)$

Proof. By conditionalisation.

1	$x \times r = y, y \times t = z$	$\vdash_{\mathrm{PA}}$	$(x \times r) \times t = z$	SubsId
2	$x \times r = y, y \times t = z$	$\vdash_{\mathrm{PA}}$	$x \times (r \times t) = z$	Lemma 10.5.22, 1
3	$x \times r = y, y \times t = z$	$\vdash_{\mathrm{PA}}$	$\exists s x \times s = z$	$\exists \mathsf{I}$, 2
4	$x \times r = y$	$\vdash_{\mathrm{PA}}$	$y \times t = z \rightarrow \exists s x \times s = z$	DedThm, 3
5		$\vdash_{\mathrm{PA}}$	$x \times r = y \rightarrow (y \times t = z \rightarrow \exists s x \times s = z)$	DedThm, 4
6		$\vdash_{\mathrm{PA}}$	$(x \times r = y \wedge y \times t = z) \rightarrow \exists s x \times s = z$	Exportation, 5
7		$\vdash_{\mathrm{PA}}$	$(\exists r x \times r = y \wedge \exists y y \times t = z) \rightarrow \exists s x \times s = z$	$\exists \mathsf{I}, 6$
8		$\vdash_{\mathrm{PA}}$	$(x|y \wedge y|z) \rightarrow x|z$	notation, 7
9		$\vdash_{\mathrm{PA}}$	$\forall x y z ((x|y \wedge y|z) \rightarrow x|z)$	$\forall \mathsf{I}, 8$

□

Lemma 10.5.30. $\vdash_{\mathrm{PA}} \forall x \exists y (y > \mathsf{S0} \wedge \forall z ((z \leq x \wedge z > \mathsf{S0}) \rightarrow z|y)$

Lemma 10.5.31. $\vdash_{\mathrm{PA}} \forall x \exists y (y > \mathsf{S0} \wedge \forall z ((z \leq x \wedge z > \mathsf{S0}) \rightarrow \neg z|y)$

Lemma 10.5.32. $\vdash_{\mathrm{PA}} \forall x \exists y(prime(y) \wedge x \leq y)$

An important principle is one known as the 'minimum principle' or the 'least number principle':

$$\exists x \psi(x) \rightarrow \exists x(\psi(x) \wedge \forall y < x \neg \psi(y)),$$

which states that if there is a number of with a certain property ψ, there is a smallest such number. This principle is provable in PA, as we will show by proving an equivalent version:

Theorem 10.5.33. *Let φ be a formula with only one free variable x. Then*

$$\vdash_{\mathrm{PA}} \forall x(\forall y < x \varphi(y) \rightarrow \varphi(x)) \rightarrow \forall x \varphi(x)$$

Instead of proving this theorem directly, we will prove it in two steps. The proof depends on the following two facts:

Lemma 10.5.34.
(1) $\vdash_{\mathrm{PA}} \forall x < 0 \varphi(x)$.
(2) $\vdash_{\mathrm{PA}} \forall y < \mathsf{S}x \varphi(y) \leftrightarrow (\forall y < x \varphi(y) \wedge \varphi(x))$.

Proof. (1) can be rewritten first as $\vdash_{\mathrm{PA}} \forall x(x < 0 \rightarrow \varphi(x))$ and then as $\vdash_{\mathrm{PA}} \forall x(\exists r(x + r = 0 \wedge x \neq 0) \rightarrow \varphi(x))$. Then:

1	$x + r = 0, x \neq 0, \neg\varphi(x)$	$\vdash_{\mathrm{PA}}$	$r + x = 0$	Lem.10.5.17
2	$x + r = 0, x \neq 0, \neg\varphi(x)$	$\vdash_{\mathrm{PA}}$	$x = 0$	Lem.10.5.4
3	$x + r = 0, x \neq 0$	$\vdash_{\mathrm{PA}}$	$\varphi(x)$	RAA
4	$x < 0$	$\vdash_{\mathrm{PA}}$	$\varphi(x)$	Def. of $<$
5		$\vdash_{\mathrm{PA}}$	$x < 0 \rightarrow \varphi(x)$	Ded.Thm., 4
6		$\vdash_{\mathrm{PA}}$	$\forall x(x < 0 \rightarrow \varphi(x))$	$\forall \mathsf{I}, 5$
7		$\vdash_{\mathrm{PA}}$	$\forall x < 0 \varphi(x)$	Notation, 6

(2) appeals to Lemma 10.5.27:

1	$\vdash_{\mathrm{PA}}$	$y < \mathsf{S}x \leftrightarrow (y < x \vee y = x)$	Lemma 10.5.27
2	$\vdash_{\mathrm{PA}}$	$(y < \mathsf{S}x \rightarrow \varphi(y)) \leftrightarrow ((y < x \vee y = x) \rightarrow \varphi(y))$	tautology, 1
3	$\vdash_{\mathrm{PA}}$	$(y < \mathsf{S}x \rightarrow \varphi(y)) \leftrightarrow (y < x \rightarrow \varphi(y) \wedge (y = x \rightarrow \varphi(y)))$	tautology, 2
4	$\vdash_{\mathrm{PA}}$	$\forall y < \mathsf{S}x \varphi(y) \leftrightarrow (\forall y < x \varphi(y) \wedge \varphi(x))$	tautology, Notation, 3

□

We are now in a position to prove Theorem 10.5.33:

Proof of Theorem 10.5.33. The two intermediary results that we prove are the following:

1. $\vdash_{\mathrm{PA}} \forall x(\forall y < x\varphi(y) \to \varphi(x)) \to \forall x \forall y < x\varphi(y)$.

2. $\vdash_{\mathrm{PA}} \forall x \forall y < x\varphi(y) \to \forall x \varphi(x)$.

Proof of (1): By induction on x.

1	$\forall x(\forall y < x\varphi(y) \to \varphi(x))$	$\vdash_{\mathrm{PA}}$	$\forall y < \mathsf{0}\varphi(y)$	Lem. 10.5.34
2	$\forall x(\forall y < x\varphi(y) \to \varphi(x)), \forall y < x\varphi(y)$	$\vdash_{\mathrm{PA}}$	$\varphi(x)$	MP
3	$\forall x(\forall y < x\varphi(y) \to \varphi(x)), \forall y < x\varphi(y)$	$\vdash_{\mathrm{PA}}$	$\forall y < \mathsf{S}x\varphi(y)$	MP, 2, Lem. 10.5.34
4	$\forall x(\forall y < x\varphi(y) \to \varphi(x))$	$\vdash_{\mathrm{PA}}$	$\forall y < x\varphi(y) \to \forall y < \mathsf{S}x\varphi(y)$	Ded.Thm., 3
5	$\forall x(\forall y < x\varphi(y) \to \varphi(x))$	$\vdash_{\mathrm{PA}}$	$\forall y < x\varphi(y)$	MP, 1, 4, PA9
6	$\forall x(\forall y < x\varphi(y) \to \varphi(x))$	$\vdash_{\mathrm{PA}}$	$\forall x \forall y < x\varphi(y)$	Gen., 5
7		$\vdash_{\mathrm{PA}}$	$\forall x(\forall y < x\varphi(y) \to \varphi(x)) \to \forall x \forall y < x\varphi(y)$	Ded.Thm., 6

Proof of (2). By induction on x, and making use of Lemmas 10.5.13 and 10.5.25:

1		$\vdash_{\mathrm{PA}}$	$\forall y < \mathsf{0}\varphi(y)$	Lem. 10.5.34(1)
2		$\vdash_{\mathrm{PA}}$	$x < \mathsf{S}x$	Lem. 10.5.13
3	$\forall y < x\varphi(y)$	$\vdash_{\mathrm{PA}}$	$\forall y < \mathsf{S}x\varphi(y)$	Lem. 10.5.25, 2
4		$\vdash_{\mathrm{PA}}$	$\forall y < x\varphi(y) \to \forall y < \mathsf{S}x\varphi(y)$	Ded.Thm., 3

These two results together are sufficient to prove the theorem. □

In this section we have proven many basic facts about ordinary arithmetic, and stated many more without proof (always encouraging the reader to work out the details for themself.) Had we started with this, the complacent reader might have begun to suspect that *any* truth about arithmetic that we can formulate in the language of PA will be provable from the axioms we've identified.

But we started this chapter with spoilers: If PA is sound, then there is a sentence of PA which is true, but not provable. This sentence we call the *Gödel sentence*, and informally, it is the following:

This sentence is true iff it is not provable.

What we need next is a way to turn this informal, meta-language sentence into something that is a sentence *in the language of* PA—that is, something that is written involving only the

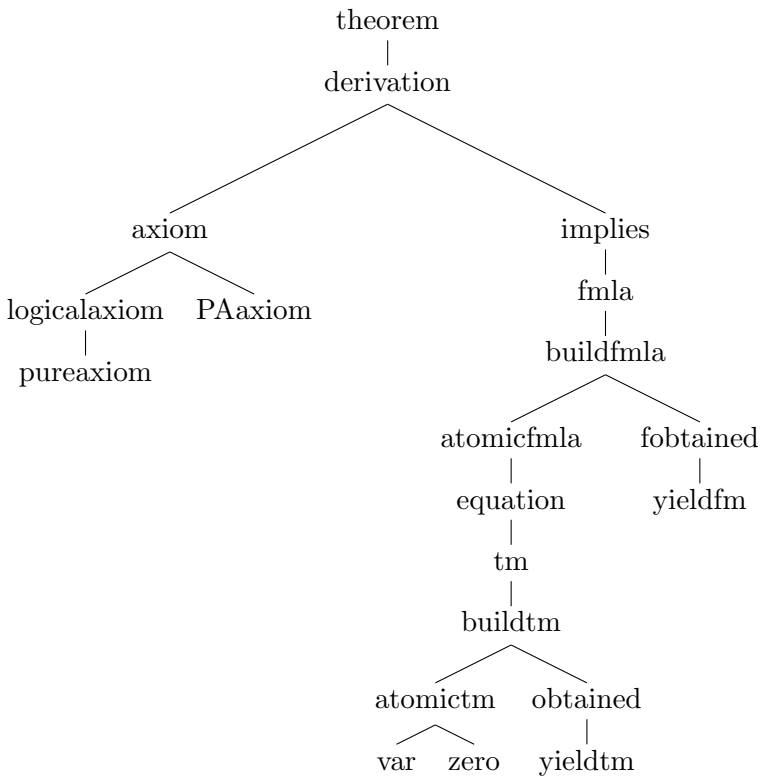


Figure 10.1: A road map to incompleteness.

logical connectives and quantifiers, identity, the constant $\mathsf{0}$, and the arithmetic operations $(\mathsf{S}, +, \times, \uparrow)$.

Another way to put it is that we have to find a way to translate talk about formulas, sentences, provabiliy, etc., into talk about *numbers*, and properties of numbers. That is the aim of the next section.

10.6 Gödel numbers

In this section, we introduce the method of 'Gödel numbering' and show how a variety of meta-level concepts, such as being a term, being a formula, being a theorem, being a derivation, etc., can be encoded by numbers and hence represented in the language of PA. For a roadmap through this section, see Figure 10.1.

The basic strategy is this: for every term τ (or formula φ) of PA, we associate a natural number $\ulcorner\tau\urcorner$ ($\ulcorner\varphi\urcorner$), called the 'Gödel number' of the term τ (the formula φ). This number

will be unique and reflect the inductive structure of τ (φ), and will be such that it is possible to determine algorithmically of any given natural number n whether it is the Gödel number of a term or a formula of PA.

Definition 10.6.1 (Gödel numbers)**.** If x is a primitive symbol in the language of PA, then $\ulcorner x \urcorner$, the *Gödel number of* x, is defined by Table 10.2:

x	$\ulcorner x \urcorner$
0	2
$\mathtt{S}$	4
$=$	6
$+$	8
$\times$	10
$\uparrow$	12
$\forall$	14
$\wedge$	16
$\neg$	18
x_n	$2n+1$

Table 10.2: Gödel numbers for PA symbols

Note that $\rightarrow$, $\vee$, and $\exists$ are omitted from the table; this is because every formula written with these connectives and operators can be rewritten in terms of $\neg$, $\wedge$, and $\forall$. Thus, we will sometimes rewrite formulas into logically equivalent forms and *then* encode them.

Note also that because our method of encoding will use the logical and arithmetic operators as prefixes, we do not need to provide a code for parentheses. This is because prefix notation is never ambiguous, so parentheses are not needed to disambiguate when a formula is represented in prefix notation.

Definition 10.6.2 (Gödel numbers of terms)**.** If τ is a term of PA, then $\ulcorner \tau \urcorner$, the *Gödel number of* τ, is defined recursively as follows:

- If τ is a variable or the constant 0, then $\ulcorner \tau \urcorner$ is defined in Table 10.2.
- If $\tau = \mathtt{S}\tau_1$, then $\ulcorner \tau \urcorner := 2^{\ulcorner \mathtt{S} \urcorner} \times 3^{\ulcorner \tau_1 \urcorner}$.
- If $\tau = \tau_1 + \tau_2$ then $\ulcorner \tau \urcorner := 2^{\ulcorner + \urcorner} \times 3^{\ulcorner \tau_1 \urcorner} \times 5^{\ulcorner \tau_2 \urcorner}$.
- If $\tau = \tau_1 \times \tau_2$ then $\ulcorner \tau \urcorner := 2^{\ulcorner \times \urcorner} \times 3^{\ulcorner \tau_1 \urcorner} \times 5^{\ulcorner \tau_2 \urcorner}$.
- If $\tau = \tau_1^{\tau_2}$ then $\ulcorner \tau \urcorner := 2^{\ulcorner \uparrow \urcorner} \times 3^{\ulcorner \tau_1 \urcorner} \times 5^{\ulcorner \tau_2 \urcorner}$.

For example, $\ulcorner 0 \urcorner = 2$, so $\ulcorner \mathsf{S}0 \urcorner = 2^4 \times 3^2 = 144$ and $\ulcorner 0 + \mathsf{S}0 \urcorner = 2^8 \times 3^2 \times 5^{144}$.

Definition 10.6.3 (Gödel numbers of formulas)**.** If φ is a formula of PA, then $\ulcorner \varphi \urcorner$, the *Gödel number of* φ, is defined recursively as follows:

- If φ is an atomic formula, then it is of the form $\tau_1 = \tau_2$ for some terms τ_1, τ_2, and $\ulcorner \varphi \urcorner := 2^{\ulcorner = \urcorner} \times 3^{\ulcorner \tau_1 \urcorner} \times 5^{\ulcorner \tau_2 \urcorner}$.
- If $\varphi = \neg\varphi_1$, then $\ulcorner \varphi \urcorner := 2^{\ulcorner \neg \urcorner} \times 3^{\ulcorner \varphi_1 \urcorner}$.
- If $\varphi = \varphi_1 \wedge \varphi_2$ then $\ulcorner \varphi \urcorner := 2^{\ulcorner \wedge \urcorner} \times 3^{\ulcorner \varphi_1 \urcorner} \times 5^{\ulcorner \varphi_2 \urcorner}$.
- If $\varphi = \forall x_n \varphi_1$ then $\ulcorner \varphi \urcorner := 2^{\ulcorner \forall \urcorner} \times 3^{\ulcorner x_n \urcorner} \times 5^{\ulcorner \varphi_1 \urcorner}$.

For example, since $\ulcorner x_1 \urcorner = 3$, $\ulcorner x_1 = 0 \urcorner = 2^6 \times 3^3 \times 5^2 = 43200$, and $\ulcorner \forall x_1 x_1 = 0 \urcorner = 2^{14} \times 3^3 \times 5^{43200}$. From this it is clear that the Gödel numbers of even very simple terms and formulas are going to be incredibly large. Because of this and the fact that the structure of a term or formula can be read off from the structure of its Gödel number, we will in general *not* multiply out the Gödel numbers. For example, instead of $2^{14} \times 3^3 \times 5^{43200}$, it will be more perspicuous to render $\ulcorner \forall x_1 x_1 = 0 \urcorner$ as:

$$2^{14} \times 3^3 \times 5^{2^6 \times 3^3 \times 5^2}$$

A final thing to note about this method of encoding is that it is sensitive to the syntactic structure of a term or formula, not its semantic meaning. While $0 + x_1 = x_1$ and $x_1 = 0 + x_1$ are semantically equivalent, they do not have the same Gödel numbers:

$$\begin{aligned} \ulcorner 0 + x_1 = x_1 \urcorner &:= 2^{\ulcorner = \urcorner} \times 3^{2^{\ulcorner + \urcorner} \times 3^{\ulcorner 0 \urcorner} \times 5^{\ulcorner x_1 \urcorner}} \times 5^{\ulcorner x_1 \urcorner} \\ \ulcorner x_1 = 0 + x_1 \urcorner &:= 2^{\ulcorner = \urcorner} \times 3^{\ulcorner x_1 \urcorner} \times 5^{2^{\ulcorner + \urcorner} \times 3^{\ulcorner 0 \urcorner} \times 5^{\ulcorner x_1 \urcorner}} \end{aligned}$$

(Both of these numbers are too large to be multiplied out on a cheap computer calculator; but it is clear from their structure that they will not be identical.)

Similarly, because Gödel numbers are constructed recursively according to the internal structure of the term or formula, we do not need to include punctuation (parentheses) in our table of codes for basic symbols, because there will be no ambiguity in the codes for terms and formulas. For instance, $(x_1 + x_2) + x_3$ is not the same term as $x_1 + (x_2 + x_3)$ (even if we can prove in PA that these two terms will always have the same value). We distinguish the terms on the basis of where the parentheses occur. But these parentheses occur in different places in the two terms because of the way the terms were constructed, and this different construction is represented in their respective Gödel numbers:

$$\begin{aligned} \ulcorner (x_1 + x_2) + x_3 \urcorner &= 2^{\ulcorner + \urcorner} \times 3^{2^{\ulcorner + \urcorner} \times 3^{\ulcorner x_1 \urcorner} \times 5^{\ulcorner x_2 \urcorner}} \times 5^{\ulcorner x_3 \urcorner} \\ \ulcorner x_1 + (x_2 + x_3) \urcorner &= 2^{\ulcorner + \urcorner} \times 3^{\ulcorner x_1 \urcorner} \times 5^{2^{\ulcorner + \urcorner} \times 3^{\ulcorner x_2 \urcorner} \times 5^{\ulcorner x_3 \urcorner}} \end{aligned}$$

And these are clearly distinct numbers.

Our aim is to define appropriate formulas φ_1 and φ_2 such that $\mathbb{N} \vDash \varphi_1(n)$ iff $n = \ulcorner \tau \urcorner$ for some term τ and $\mathbb{N} \vDash \varphi_2(n)$ iff $n = \ulcorner \varphi \urcorner$ for some formula φ. These formulas will exploit the fact that Gödel numbers of both terms and formulas are products of powers of consecutive primes, where the powers form a sequence from which the structure of the term or the formula can be read off. We therefore begin by defining how arbitrary sequences of numbers can be encoded by products of powers of primes. That is, we show how arbitrary finite sequences of numbers can be encoded by a single number, and our method depends on the fact that every number can be uniquely written as the product of prime powers.

Definition 10.6.4. Let $p_1 := 2$, $p_2 := 3$, $p_3 := 5$, $p_4 := 7$, $p_5 := 11$, $\ldots$, $p_k :=$ the kth prime number.

Our goal now is to define a formula that is true of only those numbers which are products of powers of consecutive primes. In order to reach this goal, we must introduce a number of subsidiary definitions.

Definition 10.6.5 (Codes of sequences)**.** Let $\langle a_1, \ldots, a_n \rangle$ be a sequence of n positive natural numbers. The *code* of this sequence is defined to be:

$$2^{a_1} \times 3^{a_2} \times \cdots \times p_n^{a_n}$$

Some examples of sequences and their codes are given in Table 10.3. One thing that is clear from these examples, and from the definition, is that not every number is the code of a sequence. For example, 3 is not the code of a sequence, because it is not divisible by two, and since every sequence is a sequence of positive numbers, the first multiplicand in the code must be at least 2^1. We can now also see that every Gödel number is the code of a sequence (and that some elements of this sequence can themselves be the code of a sequence), and each sequence corresponds to one of the formation rules for Gödel numbers.

Sequence	**Code**		**Value**
$\langle 1 \rangle$	2^1	=	2
$\langle 2 \rangle$	2^2	=	4
$\langle 1, 1 \rangle$	$2^1 \times 3^1$	=	6
$\langle 3, 5, 2 \rangle$	$2^3 \times 3^5 \times 5^2$	=	$48,600$
$\langle 3, 4, 3, 1 \rangle$	$2^3 \times 3^4 \times 5^3 \times 7^1$	=	$567,000$

Table 10.3: Some sequences and their codes

We can define a formula that picks out all and only those numbers which are the codes of sequences.

Definition 10.6.6.

$$\text{seq}(x) \quad := \quad \forall y \forall z (((\text{prime}(y) \wedge \text{prime}(z) \wedge y < z \wedge z|x) \rightarrow y|x) \wedge x > \underline{1}$$

That is, some number n greater than 1 is the code of a sequence if for every pair of primes x and y, if the larger prime divides n, then the smaller one does as well. This captures the fact that the sequences are (a) finite and (b) do not contain zero; so as soon as we reach a prime that does not divide n, no larger prime will divide it either.

Definition 10.6.7. For any number c, define c_k, for $k > 0$, to be $\max\{l : p_k^l | c\}$.

That is, c_k is the largest power that the kth prime can be raised to and still be a divisor of c. We will use c_k to pick out the exponents of each prime in the code of a sequence.

Example 10.6.8. Let $c = 23400$. Then, $c_1 = 3$, since $8|23400$ but not $16|23400$; $c_2 = 2$, since $9|23400$ but not $27|23400$; $c_3 = 2$ since $25|23400$ but not not $125|23400$, $c_6 = 1$, since $13|23400$ but not $169|23400$, and for every other i, $c_i = 0$.

Fact 10.6.9. $\mathbb{N} \nvDash \text{seq}(23400)$ (that is, 23400 does not code a sequence).

Proof. 13 and 11 are both primes, and $11 < 13$, but while $13|23400$, it is not the case that $11|23400$, so the antecedent of the definition is satisfied but not the consequent. □

In what follows, we are not interested in *any* sequence of non-zero numbers, but in particular types of sequences. The first type is sequences which are finite initial segments of $\mathbb{N}$, that is, sequences of the following type:

Example 10.6.10.

$$\langle 1 \rangle$$

$$\langle 1, 2 \rangle$$

$$\langle 1, 2, 3 \rangle$$

$$\langle 1, 2, 3, 4 \rangle$$

$$\langle 1, 2, 3, 4, 5 \rangle$$

$$\langle 1, 2, 3, 4, 5, 6 \rangle$$

$$\vdots$$

Such sequences, which we will call 'prime-sequences', will receive the following codes:

Example 10.6.11.

$$2^1$$

$$2^1 \times 3^2$$

$$2^1 \times 3^2 \times 5^3$$

$$2^1 \times 3^2 \times 5^3 \times 7^4$$

$$2^1 \times 3^2 \times 5^3 \times 7^4 \times 11^5$$

$$2^1 \times 3^2 \times 5^3 \times 7^4 \times 11^5 \times 13^6$$

$$\vdots$$

Note here the useful coincidence between the exponent a prime gets in the code and the place of that prime in the ordering of primes—the first prime (2) receives the exponent 1, the second prime (3) receives the exponent 2, the third prime (5) receives the exponent 3, and so on. This is precisely the fact that we will exploit when we identify a formula that is true of all and only codes of such sequences:

Definition 10.6.12.

$$\text{primeseq}(x) \quad := \quad \underline{2}|x \wedge \neg \underline{4}|x \wedge \text{seq}(x) \wedge \forall y \forall z \forall w((\text{nextprime}(z,y) \wedge y|x) \rightarrow (z^w|x \leftrightarrow y^{\mathtt{S}w}|x))$$

That is, some number c is a code of a finite sequence of successive primes if:

1. It is divisible by 2 (since every such code will at least have 2^1 as a actor).
2. It is *not* divisible by 4 (since the only way such a code could be divisible by 4 is if it were divisible by 2^2, but for each code c, $\max\{l : 2^l|c\} = 1$.)
3. For any two adjacent primes p and q, such that the larger one (say, q) divides c, then p raised to a power will divide c iff q raised to the successor of that power also divides c.

From this definition, we can define what it means for a number to be identical to one of the p_is, that is, to be a prime occupying a certain position in the ordering of prime numbers.

Definition 10.6.13.

$$\text{nthprime}(x,y) \quad := \quad \text{prime}(x) \wedge \exists w(\text{primeseq}(w) \wedge x^y|w \wedge \neg x^{y+\underline{1}}|w)$$

That is, some prime p is the nth prime iff p is a prime, and there is some number that is the code of a prime-sequence, where p raised to n divides that code, but p raised to no larger power does.

Fact 10.6.14. $\mathbb{N} \vDash \text{nthprime}(3, 2)$, because 3 is a prime number, and there is a code of a prime-sequence, namely $2^1 \times 3^2$, where 3 raised to 2 divides the code, but 3 raised to 3 does not. Thus, 3 is the 2nd prime.

What follows is still not completely written: See also [Goldstern and Judah, 1998, §4.4].

Definition 10.6.15.

$$\text{entry}(x, y, z) \quad := \quad \text{seq}(x) \wedge y > 0 \wedge z > 0 \wedge \forall w((\text{prime}(w) \wedge \text{nthprime}(w, y)) \to (w^z | x \wedge \neg w^{z+1} | x))$$

10.6.1 Properties of Gödel numbers of terms

In this section, we show how to identify all the natural numbers which are Gödel numbers of terms. For the atomic terms, this is straightforward: We simply have to check whether the numbers are the right codes according to our table of codes for symbols (Table 10.2). First we define when a number encodes a constant. Since we have only one constant ($\mathsf{0}$), the definition is straightforward:

Definition 10.6.16.

$$\text{zero}(x) \quad := \quad x = \underline{\ulcorner \mathsf{0} \urcorner}$$

That is, a number is a code of $\mathsf{0}$ if it is identical to the code for $\mathsf{0}$, that is, if it is 2. No other number is a code for $\mathsf{0}$, so $\text{zero}(\underline{2})$ is true while $\text{zero}(\underline{n})$ is false for any other n.

What follows is still not completely written: See also [Goldstern and Judah, 1998, §4.5] ffor definitions not yet defined here.

Definition 10.6.17.

$$\text{var}(x) \quad := \quad \exists y(x = (\underline{2} \times y) + \underline{1})$$

Definition 10.6.18.

$$\begin{aligned} \text{yieldtm}(t_1, t_2, o, t) \quad := \quad & ((o = \underline{\ulcorner + \urcorner} \vee o = \underline{\ulcorner \times \urcorner} \vee o = \underline{\ulcorner \uparrow \urcorner}) \wedge (t = \underline{2}^o \times \underline{3}^{t_1} \times \underline{5}^{t_2})) \vee \\ & (o = \underline{\ulcorner \mathsf{S} \urcorner} \wedge t = \underline{2}^o \times \underline{3}^{t_1}) \end{aligned}$$

Definition 10.6.19.

$$\text{obtained}(c, k, l, o, n) \quad := \quad \exists e \exists f \exists g(\text{entry}(c, k, e) \wedge \text{entry}(c, l, f) \wedge \text{entry}(c, n, g) \wedge \text{yieldtm}(e, f, o, g))$$

Definition 10.6.20.

$$\text{atomictm}(c,n) \quad := \quad \exists e(\text{entry}(c,n,e) \wedge (\text{zero}(e) \vee \text{var}(e)))$$

Definition 10.6.21.

$$\begin{aligned}\text{buildtm}(c,t) \quad := \quad & \text{seq}(c) \wedge \exists n(\text{entry}(c,n,t)) \wedge \\ & \forall n \forall e(\text{entry}(c,n,e) \rightarrow (\text{atomictm}(c,n) \vee \exists k < n \exists l < n \exists o(\text{obtained}(c,k,l,o,n))))\end{aligned}$$

Definition 10.6.22.

$$\text{term}(x) \quad := \quad \exists y(\text{buildtm}(y,x))$$

Fact 10.6.23. $\mathbb{N} \vDash \text{term}(n)$ iff there is a term t such that $n = \ulcorner t \urcorner$.

Proof. By induction. Note that if $\mathbb{N} \vDash \text{buildtm}(c,t)$, then for all $k > 0$, either $c_k = 0$ or c_k codes a term, and if $c_k > 0$ and $l < k$, then $c_l > 0$ as well. □

But not every natural number is the code of a term. For example, we show that:

Lemma 10.6.24. *4 is not the code of a term.*

Proof. Suppose that it is. Then $\mathbb{N} \vDash \text{term}(4)$ by Fact 10.6.23.

$$\begin{aligned}\mathbb{N} \vDash \text{term}(\underline{4}) \quad & \text{iff} \quad \mathbb{N} \vDash \exists c \text{buildtm}(c,\underline{4}) \\ & \text{iff} \quad \mathbb{N} \vDash \exists c(\text{seq}(c) \wedge \\ & \qquad \exists n \text{entry}(c,n,\underline{4}) \wedge \\ & \qquad \forall n,e(\text{entry}(c,n,e) \rightarrow (\text{atomictm}(c,n) \vee \exists k,l < n \exists o(\text{obtained}(c,k,l,o,n)))\end{aligned}$$

Let us give a name to this c, namely, c. Then we know three things about c: (1) $\mathbb{N} \vDash \text{seq}(\mathsf{c})$; (2) $\mathbb{N} \vDash \exists n \text{entry}(\mathsf{c},n,\underline{4})$; and (3) $\mathbb{N} \vDash \forall n,e(\text{entry}(\mathsf{c},n,e) \rightarrow (\text{atomictm}(\mathsf{c},n) \vee \exists k,l < n \exists o(\text{obtained}(\mathsf{c},k,l,o,n)))$. We'll treat each conjunct in turn.

Conjunct (1):

$$\mathbb{N} \vDash \text{seq}(\mathsf{c}) \quad \text{iff} \quad \mathbb{N} \vDash (\forall p,q((\text{prime}(p) \wedge \text{prime}(q) \wedge p < q \wedge q|\mathsf{c}) \rightarrow p|\mathsf{c}) \wedge \mathsf{c} > 1$$

Conjunct (2):

$$\begin{aligned}\mathbb{N} \vDash \exists n(\text{entry}(\mathsf{c},n,\underline{4}) \quad \text{iff} \quad & \mathbb{N} \vDash \exists n(\text{seq}(\mathsf{c}) \wedge (n > 0) \wedge (\underline{4} > 0) \wedge \\ & \forall p((\text{prime}(p) \wedge \text{nthprime}(p,n)) \rightarrow (p^{\underline{4}}|\mathsf{c} \wedge \neg p^{\underline{5}}|\mathsf{c})))\end{aligned}$$

Let us give a name to this n, namely, n. Then we know that $\mathbb{N} \vDash \text{seq}(\mathsf{c})$ (which we already knew by Conjunct (1) above). Finally, we have that if p is the nth prime, $\mathbb{N} \vDash p^{\underline{4}}|\mathsf{c} \wedge \neg p^{\underline{5}}|\mathsf{c}$.

Since we have that $\mathbb{N} \vDash \text{entry}(\mathsf{c}, \mathsf{n}, \underline{4})$, let us turn to Conjunct (3). Since the antecedent is satisfied, it follows that $\mathbb{N} \vDash \text{atomictm}(\mathsf{c}, \mathsf{n}) \vee \exists k, l < \mathsf{n} \exists o(\text{obtained}(\mathsf{c}, k, l, o, \mathsf{n}))$. We'll show that both disjuncts leads to a contradiction.

Disjunct (1):

$$\mathbb{N} \vDash \text{atomictm}(\mathsf{c}, \mathsf{n}) \quad \text{iff} \quad \mathbb{N} \vDash \exists e(\text{entry}(\mathsf{c}, \mathsf{n}, e) \wedge (\text{zero}(e) \vee \text{var}(e)))$$

Now, by Conjunct (2) above, we already know that $\mathbb{N} \vDash \text{entry}(\mathsf{c}, \mathsf{n}, \underline{4})$. If this disjunct is true, then:

$$\mathbb{N} \vDash \text{zero}(\underline{4}) \vee \text{var}(\underline{4})$$

But since $\mathbb{N} \vDash \underline{4} \neq \underline{2}$, it follows that $\mathbb{N} \vDash \neg\text{zero}(\underline{4})$. Further, since $\mathbb{N} \nvDash \exists y(\underline{4} = (\underline{2} \times y) + \underline{1})$, it also follows that $\mathbb{N} \vDash \neg\text{var}(\underline{4})$. Thus, the first disjunct is not true.

Disjunct (2): Let us give names to the k, l, o, namely, $\mathsf{k}, \mathsf{l}, \mathsf{o}$.

$$\mathbb{N} \vDash \text{obtained}(\mathsf{c}, \mathsf{k}, \mathsf{l}, \mathsf{o}, \mathsf{n}) \quad \text{iff} \quad \mathbb{N} \vDash \exists e, f, g((\text{entry}(\mathsf{c}, \mathsf{k}, e) \wedge \text{entry}(\mathsf{c}, \mathsf{l}, f) \wedge \text{entry}(\mathsf{c}, \mathsf{n}, g) \wedge \text{yieldtm}(e, f, \mathsf{o}, g)$$

Again, let us give names to e, f, g, namely, $\mathsf{e}, \mathsf{f}, \mathsf{g}$. Now, we already know that $\mathbb{N} \vDash \text{entry}(\mathsf{c}, \mathsf{n}, \underline{4})$, so it follows that $\mathbb{N} \vDash \text{yieldtm}(\mathsf{e}, \mathsf{f}, \mathsf{o}, \underline{4})$. We'll focus on this as this is where the problems arise:

$$\begin{aligned} \mathbb{N} \vDash \text{yieldtm}(\mathsf{e}, \mathsf{f}, \mathsf{o}, \underline{4}) \quad \text{iff} \quad & \mathbb{N} \vDash ((\mathsf{o} = \underline{8} \vee \mathsf{o} = \underline{10} \vee \mathsf{o} = \underline{12}) \wedge \underline{4} = \underline{2}^{\underline{\mathsf{o}}} \times \underline{3}^{\underline{\mathsf{e}}} \times \underline{5}^{\underline{\mathsf{f}}}) \\ & \vee (\mathsf{o} = \underline{4} \wedge \underline{4} = \underline{2}^{\underline{\mathsf{o}}} \times \underline{3}^{\underline{\mathsf{e}}}) \end{aligned}$$

But since 2^4, 2^6, 2^8, and 2^{10} are all greater than 4, it follows that none of these possibilities are true.

And that's a contradiction. □

10.6.2 Properties of Gödel numbers of formulas and proofs

Next, we focus on identifying properties of Gödel numbers of codes of formulas that are substitution instances of each other.

This section is not complete. See [Goldstern and Judah, 1998, §4.5].

10.6.3 Encoding substitution

This section is not complete. See [Goldstern and Judah, 1998, §4.6].

10.7 The incompleteness theorem

This section is not complete.

Proof. Let

$$\text{formula}(\ulcorner\underline{\varphi}\urcorner) \wedge \exists y(\text{diag}(\ulcorner\underline{\varphi}\urcorner, y) \wedge \neg\text{theorem}(y)) \tag{σ}$$

Since φ *is* a formula, we know that

$$\mathbb{N} \vDash \text{formula}(\ulcorner\underline{\varphi}\urcorner)$$

So:

$$\mathbb{N} \vDash \sigma \leftrightarrow \exists y(\text{diag}(\ulcorner\underline{\varphi}\urcorner, y) \wedge \neg\text{theorem}(y))$$

Now, recall that $\mathbb{N} \vDash \text{diag}(\ulcorner\underline{\varphi}\urcorner, y) \leftrightarrow y = \ulcorner\mathcal{D}(\varphi)\urcorner$, i.e., if $y = \ulcorner\sigma\urcorner$. So, substituting in the identity, we have:

$$\mathbb{N} \vDash \sigma \leftrightarrow \exists y(y = \ulcorner\sigma\urcorner \wedge \neg\text{theorem}(y))$$

That is,

$$\mathbb{N} \vDash \sigma \leftrightarrow \neg\text{theorem}(\ulcorner\sigma\urcorner)$$

That is (swapping negations),

$$\mathbb{N}\neg\sigma \leftrightarrow \text{theorem}(\ulcorner\sigma\urcorner)$$

There are two options, either $\vdash_{\text{PA}} \sigma$ or $\nvdash_{\text{PA}} \sigma$.

If $\vdash_{\text{PA}} \sigma$, then $\mathbb{N} \vDash \text{theorem}(\ulcorner\sigma\urcorner)$ and $\mathbb{N} \vDash \neg\sigma$, which makes PA unsound.

If $\nvdash_{\text{PA}} \sigma$, then $\mathbb{N} \vDash \neg\text{theorem}(\ulcorner\sigma\urcorner)$ and hence $\mathbb{N} \vDash \sigma$. Thus, σ is true in the standard model of arithmetic, but is unprovable in PA. So the Gödel sentence σ is the sentence which is both true, but unprovable, demonstrating that PA is incomplete; it does not derive all true arithmetical sentences. □

The question then is this: Is this a problem with our choice of axioms? Unlike some instances of incompleteness (for instance, 'mere' incompleteness, where a system is incomplete because it lacks an axiom or a rule, and can be made complete simply by adding the necessary axiom or rule), there is no reasonable way to extend the axioms given in §10.4 and obtain a complete system.

What do we mean by "reasonable" here? There are a number of properties of axiomatic systems that we might desire in this context, the central of which is *decidability*, that is, the existence of a finite decision procedure that determines whether something is an axiom of the system. PA is decidable; PA + $\{\varphi\}$ is decidable, for any individual formula φ. But $\{\varphi : \mathbb{N} \vDash \varphi\}$ is *not*.

"Decidable" is itself a rather informal concept; it can be made more precise via the notion of recursion introduced in §10.3:

- (first-order) definable/expressible

10.8 Some background material

10.8.1 Deriving Russell's paradox from Basic Law V

For further reading, see [Boolos and Clark, 1993] and the Stanford Encyclopedia of Philosophy articles at `http://plato.stanford.edu/entries/russell-paradox/` and `http://plato.stanford.edu/entries/frege/` (esp. §2.4.1)

Frege's Basic Law V says, informally: For any concepts F and G, the extension of F is identical to the extension of G iff for all a, Fa iff Ga. In formal notation:

$$\bar{\epsilon}f(\epsilon) = \bar{\alpha}g(\alpha) \leftrightarrow \forall x(f(x) = g(x))$$

$\bar{\epsilon}f(\epsilon)$ is the extension of the concept F and $\bar{\alpha}g(\alpha)$ is the extension of the concept G. The extension is those objects which F maps to true.

Extensions of concepts are concepts themselves, and so Basic Law V applies to extensions. In fact, any formula with a free variable in it can be thought of as a concept in the sense that any such formula can be turned into an extension of a concept. Let $\varphi(x)$ be a formula with x as a free variable. Then $\bar{\epsilon}\varphi(\epsilon)$ is the extension of $\varphi(x)$: That is, it is the set of all objects n such that $\varphi(n)$ is true. Now, some concepts are such that they apply to themselves, and some are not. The concept 'blue' is not blue, and the concept 'cat' is not a cat, but the concept 'concept' is a concept, and the concept 'abstract' is itself abstract. Consider then the concept 'extension which is not an element of itself'. (Remember that above we said that extensions are themselves concepts.) Let E be this concept, and let e be its extension. Now: is e an element of E? It is iff $E(e)$ is true. $E(e)$ is true iff e is an extension which is not an element of itself, that is, e is an extension such that $\neg E(e)$. So we have that $E(e)$ is true iff $\neg E(e)$ is true.

What this shows is that Frege's concept of extension is inconsistent. The ZFC Axiom 3 of Comprehension avoids the problems of Basic Law V by rejecting the possibility of collecting into a set *all* objects which satisfy a particular formula, and instead says that one must first start with some well-defined set, and then take all the elements *of that set* which satisfy a particular formula, and form a new set out of that.

10.8.2 The Axiom of Choice

Definition 10.8.1 (Well-order)**.** A binary relation R on a set A is a *well-order* if every non-empty subset of A has an R-least element.

The Axiom of Choice states that every set is well-orderable. It is non-constructive in that it simply states the existence of a well-order R, but it doesn't say anything about what R is like or how to find it.

The real interval $[0,1]$ is not well-ordered on the usual $\leq$ relation, since there is no $\leq$-least element in the subset $(0,1)$. However, under Choice, it is well-orderable. The fact that $\mathbb{N}$ (the set of natural numbers) is well-ordered on the usual $\leq$ relation is what underpins our use of both mathematical induction and the "Least Number Principle".

For more on well-ordering, chapter 7 of [Goldrei, 1996] (at least partially available on googlebooks) has a lot of discussion and examples. The Axiom of Choice is not universally accepted, particularly amongst people who worry about the foundations of mathematics. An alternative candidate is the Axiom of Determinacy (AD), a good introduction to which is found in Wikipedia: `https://en.wikipedia.org/wiki/Axiom_of_determinacy`. And if you are *really* interested in AD and higher infinite cardinals, see Daisuke Ikegami's PhD thesis *Games in Set Theory and Logic*, at `http://www.illc.uva.nl/Research/Publications/Dissertations/DS-2010-04.text.pdf`.

10.8.3 Ordinal and cardinal arithmetic, finite and transfinite

The ordinals measure *place in order*; the cardinals measure *size*. These coincide with finite numbers: The number in the nth place in the order is also going to be the nth largest number. In general, this will not hold when we move to the infinite case. This is because there are many different ways to order infinite numbers, but this does not change their size.

The first infinite ordinal is ω. The first infinite cardinal is $\aleph_0$.

In finite arithmetic, + and $\times$ are commutative. This does not hold in transfinite arithmetic. In transfinite ordinal arithmetic, $2+\omega \neq \omega+2$ and $2\times\omega \neq \omega\times 2$, and $2+\omega = 2\times\omega = \omega$. In transfinite cardinal arithmetic, $\kappa+\lambda = \kappa\times\lambda = \max\{\kappa,\lambda\}$.

In finite arithmetic, exponentiation can be defined in terms of + and $\times$, and it is straightforward to extend this to infinite ordinals:

$$a^0 = 1 \tag{10.4}$$

$$a^{(\beta+1)} = (\alpha^\beta)\times\alpha \tag{10.5}$$

$$a^{(\delta)} \text{ for } \delta \text{ limit} = \text{limit of } \alpha^\beta \text{ for all } \beta<\delta \tag{10.6}$$

What about if the exponents are cardinals? What cardinal does $2^{\aleph_0}$ equal? This is the same question as asking "How big is $2^{\aleph_0}$?" More precisely, asking what cardinal $2^{|x|}$ is is asking how big is the power set of x. We know that the following holds:

$$\forall n, |n| < 2^{|n|}$$

I.e., $4 < 2^4$; $\aleph_0 < 2^{\aleph_0}$. (Since $2^w = w$, it is clear that ordinal exponentiation and cardinal exponentiation diverge when the exponents are infinite.)

Why is it that asking "how big is $2^{|x|}$" is the same as asking "how big is the power set of x"? One answer is that it is defined this way.[7] But a better way to see it is via characteristic function. Any subset of x can be represented by its characteristic function, that is, the function that gives 0 if the element is not in the set and 1 if the element is in. The number of distinct functions on any given set x is $2^{|x|}$. Since each function gives rise to a distinct subset, asking how big is $2^{|x|}$ is equivalent to asking how big the power set of x is.

10.8.4 About the consistency of PA

The 2nd Incompleteness Theorem says that if PA is consistent, then it cannot prove its own consistency. (The consistency of PA can be written as a PA formula, e.g., $\neg\text{thm}(0 = 1)$, often abbreviated Cons(PA).) When discussing modal logic, we identified at least three different things we could mean by 'consistent', and it turned out that all of them were equivalent in the context. One is a proof-theoretic notion: "an axiomatic system is *consistent* iff not every wff is a theorem of that system" [Hughes and Cresswell, 1996, p. 46]. But another notion of 'consistent' is semantic: We call a set of formulas consistent if they are satisfiable, i.e., if there is a model where all the formulas are true. One might say that we do have a proof of the consistency of PA, namely, it has a model. Not only that, it has *lots* of models, one of which, namely $\mathbb{N}$ with the usual $<$ order, is considered the *standard* model. (Non-standard models of arithmetic are *really really* interesting, and discussed in §10.2.1.)

But what is a model? It is a set-theoretic structure: A collection of sets with various relations/predicates/functions describing them. So, when we say that "there is a model", we are making a statement about set theory, usually ZFC. But... set theory is an extension of PA, and thus it falls prey to the incompleteness theorems: If it is consistent, we cannot prove it. Thus, semantic consistency proofs via satisfiability are also all relative consistency proofs: "If ZFC is consistent, then a set of formulas Γ is satisfiable in a model $\mathfrak{M}$."

So, to the question why is it that pointing to a model of the axioms of, say, K is sufficient to prove the consistency of K but not of PA, the answer is that producing a model *is* sufficient in both cases, but only if ZFC itself is consistent. If it isn't, then even if we could prove in it that there is a model of the relevant type, the proof means nothing, because the statement could be false.

Here's a useful discussion from a textbook on mathematical logic:

[7] The definition of cardinal exponentiation given in Def. 10.24 in Kunen's *Set Theory* is: $A^B =^B A = \{f : f \text{ is a functon s.t. } dom(f) = B \wedge ran(f) \subset A\}$.

[W]e have emphasized the experience that mathematical statements can be formalized in $\mathcal{L}^{\{\epsilon\}}$[8] and that provable statements lead to formalizations which are derivable from ZFC. Taking this for granted, suppose it were possible in mathematics to prove both a statement and its negation. Let φ be a formalization of this statement. Then both ZFC$\vdash \varphi$ and ZFC$\vdash \neg\varphi$ would hold, and thus ZFC would be inconsistent. Therefore, a proof that ZFC is consistent could be regarded as strong evidence for the consistency of mathematics. In fact, the question of the consistency of ZFC is one of the key problems of foundational investigations. In an explicit formulation, it asks: Is there a derivation in the sequent calculus of a sequent of the form $\varphi_1 \dots \varphi_n(\varphi \wedge \neg\varphi)$ where $\varphi_1, \dots, \varphi_n$ are ZFC axioms? [No, by the second incompleteness theorem.]... In particular, one cannot prove the existence of a model of ZFC (since [the satisfiability of ZFC] would imply [the consistency of ZFC]). Nevertheless, the fact that ZFC has been investigated and used in mathematics for decades and no inconsistency has been discovered, attests to the consistency of ZFC [Ebbinghaus et al., 1994, p. 111].

[8] I.e., the language which includes as its sole non-logical symbol set-inclusion.

Chapter 11

Propositional modal logic: theory [last modified 02 Feb 26]

11.1 Introduction

The language of Peano Arithmetic introduced in the previous chapter extends classical predicate logic by extending the non-logical vocabulary and stipulating new axioms. In this chapter, we introduce a logic which extends classical logic with the addition of new *logical* vocabulary to the language. This new vocabulary will allow us to talk about necessity and possibility, specifically the necessity and possibility of propositions, and hence it is an extension of propositional logic (cf. Chapter §7).

The modal notions of necessity and possibility are not truth functional in the way that the other propositional connectives are. That is, we cannot tell from the truth value of a formula alone whether it is necessary, possible, or impossible. We'll give a few informal examples of this. Consider the following sentences:

1. Not all cats are black.
2. 2 is even.
3. All humans are animals.
4. It is Tuesday.
5. $3 + 3 = 7$.

The first is true, but only *merely* true; it is not necessary, as we can perfectly well imagine the possibility of all cats being black.

The second one is both true and—on pretty much all views about the nature of numbers—necessary.

The third is true, but whether it is necessary depends on the account of necessity in question. If the account is logical necessity, then it is not necessary, since there is no contradiction in humans being other than animals. But if the account is some sort of metaphysical necessity, then one might argue that since 'animal' is part of the essence of 'human', there is no possibility of there being a human which is not an animal.

The fourth one is (most likely) false, but when it is false, it is only *merely* false, not impossible; the sentence is true every Tuesday.

Finally, the fifth one is false, and—again, on pretty much all views about the nature of numbers—necessarily false (that is, impossible).

From these examples, we can see that there are a variety of ways in which sentences can be true—sometimes they are true necessarily, with no possibility of falsehood; and sometimes they are only *merely* true, with the possibility of falsehood—as well as a variety of ways in which sentences can be false—sometimes they are false necessarily, with no possibility of being true; and sometimes they are only *merely* false, with the possibility of truth. What this shows is that not only can we not tell from the truth value of these sentences in the actual world alone whether they are necessary or merely possible, but in order to make this determination we must look at situations other than the actual world. These possible situations or scenarios—or "possible worlds"—are what allow us to make the determination. Informally, possibility is truth in some (relevant) possible world while necessity is truth in all (relevant) possible worlds.

Why 'relevant' in parentheses? Because depending on the type of necessity involved, we do not always consider *all* of the possible worlds. The approach to modal logic that we take in this chapter is designed to be able to accommodate multiple types of necessity, including the following:

- Logical necessity.
- Temporal necessity (either in the past or in the future); see Chapter 13.
- Epistemic necessity; see Chapter 12.
- Metaphysical necessity.
- Physical/causal necessity.
- Deontic necessity; see Chapter 12.

When we are interested in physical necessity, only those possible situations or scenarios where the laws of physics hold will be relevant. When we are interested in deontic necessity, only those possible situations or scenarios where necessity implies possibility (another formulation of the famous dictum "Ought implies can") will be relevant.

The informal analysis of the possibility or necessity of atomic propositions just given can be extended to complex propositions as well. If φ is the sentence "It is Tuesday", then, as noted above, it is not necessary. But $\varphi \vee \neg\varphi$ will be necessary, because there is no way in which this sentence could be false. In the next section, we define a language that allows us to speak in the object language about necessity and possibility, and then in §11.3 we introduce two different approaches to the semantics for modal logic (Kripke frames in §11.3.1 and neighborhood models in §11.3.2).

11.2 Language

Modal propositional logic is built on top of non-modal proposition logic, and uses the same base language with two new unary connectives.

Definition 11.2.1. A *propositional modal language* $\mathcal{L}_{Mp}$ consists of:

- A countably infinite set of atomic propositional letters p, q, r, ...
- Six logical connectives: Three unary $\neg, \Box, \Diamond$, three binary: $\vee, \wedge, \rightarrow$.
- Punctuation: (,).

This language is identical to the language of propositional logic (cf. Definition 7.3.2) with the addition of two new unary operators: $\Box$ and $\Diamond$.

The set of wffs of $\mathcal{L}_{Mp}$ is defined recursively:

Definition 11.2.2 (Propositional modal wffs)**.**

- Every atomic proposition is an $\mathcal{L}_{Mp}$-wff.
- If φ is a $\mathcal{L}_{Mp}$-wff, then so are $\neg\varphi$, $\Box\varphi$, and $\Diamond\varphi$. We read $\Box\varphi$ as "It is necessary that φ" or "Necessarily φ" or "Necessary φ" and $\Diamond\varphi$ as "It is possible that φ" or "Possibly φ" or "Possible φ".
- If φ and ψ are $\mathcal{L}_{Mp}$-wffs, then so are $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, and $(\varphi \rightarrow \psi)$.

This set of modal wff is the same as the set of propositional wff (cf. Definition 6.1.5) with the addition of the relevant clauses for $\Box$ and $\Diamond$. As we do in ordinary propositional logic, we will drop the outermost pair of parentheses when no ambiguity will occur.

We will show below that we did not have to introduce $\Diamond$ as a distinct operator, but could have defined it as $\neg \Box \neg$. As a result, we will also be able to show that the single operator $\Box$ is interchangeable with the string of three operators $\neg \Diamond \neg$.

Recall that we introduced (pg. 121) the notation $\bot$ and $\top$ to represent arbitrary contradictions and tautologies. We will avail ourselves of this notation here.

Definition 11.2.3 (Modal constant wff)**.** A *modal constant wff* is a wff that is constructed only from $\top$, $\bot$, truth-functional operators, and modal operators. (Cf. Definition 7.4.10.)

11.2.1 The standard translation, part 1

Modal logic can be translated into predicate logic by the standard translation:

Definition 11.2.4 (Standard translation)**.** We translate the language $\mathcal{L}_{Mp}$ of modal logic containing atomic propositions p_i for every natural number i into a predicate language $\mathcal{L}_q$ containing the unary relation symbols $\dot{P}_i$ and the binary relation symbol $\dot{R}$ via the *Standard Translation* ST. For a variable x, we define ST_x recursively:

$$\begin{aligned}
\mathrm{ST}_x(p_i) &:= \dot{P}_i(x) \\
\mathrm{ST}_x(\neg\varphi) &:= \neg\mathrm{ST}_x(\varphi) \\
\mathrm{ST}_x(\varphi \vee \psi) &:= \mathrm{ST}_x(\varphi) \vee \mathrm{ST}_x(\psi) \\
\mathrm{ST}_x(\Diamond\varphi) &:= \exists y((\dot{\mathrm{R}}(x,y) \wedge \mathrm{ST}_y(\varphi)))
\end{aligned}$$

We will show in §11.3.1 that we do not need to provide translation clauses for wffs using $\wedge$, $\rightarrow$, and $\Box$, because every modal wff containing one of these symbols can be rewritten into an equivalent one containing only $\neg$, $\vee$, and $\Diamond$.

11.3 Semantics

The non-modal propositional operators are *truth-functional* in the sense that the truth value of a complex formula such as $\varphi \wedge \psi$ depends solely on the truth values of the component formulas φ and ψ. The new modal operators are not truth-functional: There is no way to determine the truth value of, e.g., $\Box\varphi$ from the truth value of φ alone, as we discussed in §11.1. Any semantics that we give for the modal operators must therefore involve more than just truth values and truth value assignments. In this section, we introduce two standard

types of semantics for modal logic, Kripke semantics (§11.3.1) and neighborhood semantics (§11.3.2).

11.3.1 Kripke semantics

In our informal introduction to modality, we made much use of talking about different ways in which sentences can be true, giving rise to different scenarios or situations, some of which are relevant to the calculation of possibility and necessity, and some of which are not. In this section, we make this sort of talk explicit by formalising a notion of "possible worlds", and of "accessibility" between possible worlds. The semantics we define in this section are due to [Kripke, 1959].

At this point, it is worth pausing to note that from the point of view of logic, what possible worlds *are* does not matter. One can adopt a robust modal realism a la [Lewis, 1979], or something more moderate a la [Stalnaker, 1976], or even remain agnostic, treating the notion of possible world as a mere pragmatic tool. Logic alone does not make any statement about the metaphysical nature of possible worlds or about which worlds exist.[1] There will come a point (in §11.5.2) where it will be useful to interpret possible worlds in a very specific way: Because we are not specifying in advance what can count as a possible world, we are free to do so in particular contexts when needed.

Definition 11.3.1 (Kripke frames)**.** A *Kripke frame* is a pair $\mathfrak{F} = \langle W, R\rangle$ where

- W is a non-empty[2] set of possible worlds.
- $R \subseteq W \times W$ is a binary relation between possible worlds (that is, it is a set of ordered pairs $\langle w, w'\rangle$, where w and w' are both elements of W).

We indicate worlds in W by lower case letters at the end of the alphabet, e.g., w, x, y, z, with or without subscripts or $'$ (ticks, or 'primes'). We call R an "accessibility" relation. If world y is accessible to world x in a frame (that is, $\langle x, y\rangle \in R$), then we write xRy (and we also say that "y is possible relative to x" or, sometimes, that "x sees y").[3] This accessibility relation is what allows us to identify which possible worlds, out of all the possible possible worlds, are "relevant" in a given context.

One of the advantages of Kripke frames as an analysis of modality is that they lend themselves naturally to diagrammatic presentation. For example, if xRy, we can depict

[1] With one exception which we will see in Definition 11.3.1; this constraint on modal models is analogous to the constraint on first-order models found in Definition 8.3.1.

[2] This is the one constraint that logic puts on the metaphysics of modality: There has to be at least one possible world, namely, the actual world.

[3] In some textbooks, xRy is written Rxy.

this as in Figure 11.1, with the worlds represented by dots and the accessibility relation between worlds by arrows (if there is an arrow pointing from x to y, then y is accessible from x):

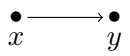


Figure 11.1: xRy

Kripke frames are turned into Kripke models by the addition of a truth value assignment, which says which basic atomic propositions are true at which possible worlds.

Definition 11.3.2 (Kripke model)**.** A Kripke model is a pair $\mathfrak{M} = \langle \mathfrak{F}, V \rangle$ where

- $\mathfrak{F} = \langle W, R \rangle$ is a Kripke frame, and
- V is a function from the atomic propositions and the possible worlds to the truth values T and F. (Formally, if P is a set of atomic propositiitions, then $V : \mathrm{P} \times W \to \{T, F\}$).

If $\mathfrak{M} = \langle \mathfrak{F}, V \rangle$, we say that "model $\mathfrak{M}$ is *based on* frame $\mathfrak{F}$", and we sometimes write $\mathfrak{M}$ as $\mathfrak{M} = \langle W, R, V \rangle$. If $V(p, w) = T$, we say that "V makes p true at w" or that "p is true at w (according to V)". V tells us which atomic propositions are true at which worlds, but from this basic information, we can compute the truth of more complex formulas—both the truth-functional (purely propositional) ones and the modal ones:

Definition 11.3.3 (Truth conditions)**.** The truth of a formula is evaluated at an individual world w of a given model $\mathfrak{M}$ recursively as follows:

$$
\begin{array}{lll}
\mathfrak{M}, w \vDash p & \text{iff} & V(p, w) = T \\
\mathfrak{M}, w \vDash \neg\varphi & \text{iff} & \mathfrak{M}, w \nvDash \varphi \\
\mathfrak{M}, w \vDash \varphi \wedge \psi & \text{iff} & \mathfrak{M}, w \vDash \varphi \text{ and } \mathfrak{M}, w \vDash \psi \\
\mathfrak{M}, w \vDash \varphi \vee \psi & \text{iff} & \mathfrak{M}, w \vDash \varphi \text{ or } \mathfrak{M}, w \vDash \psi \\
\mathfrak{M}, w \vDash \varphi \to \psi & \text{iff} & \mathfrak{M}, w \nvDash \varphi \text{ or } \mathfrak{M}, w \vDash \psi \\
\mathfrak{M}, w \vDash \Box\varphi & \text{iff} & \text{for all } w' \text{ such that } wRw', \mathfrak{M}, w' \vDash \varphi \\
\mathfrak{M}, w \vDash \Diamond\varphi & \text{iff} & \text{there is } w' \text{ such that } wRw' \text{ and } \mathfrak{M}, w' \vDash \varphi
\end{array}
$$

(We read, e.g., $\mathfrak{M}, w \vDash p$ as "w is a model of p" or "p is true at w".) The first five clauses simply reiterate that the propositional connectives have the same truth values in modal logic as they do in propositional logic. The final two clauses provide the semantics for the new operators: Necessity is defined relative to a world w, and it is defined to be

truth in all worlds that are relevant (accessible) to w. Possibility is also defined relative to a world w, but it is truth in *some* world that is relevant (accessible) to w.

This definition makes explicit how it is that the operators $\Box$ and $\Diamond$ are non-truthfunctional: It is not just the truth values of the subformulas at the current world that are required to determine the truth value of the complex formula, but we also need to look at the truths of those subformulas at *other* worlds (cf. Figure 11.4, which shows two models with the same propositional valuation but different truth values for modal formulas).

Diagrams for frames like that given in Figure 11.1 can be turned into diagrams for models by annotating each of the possible worlds with the atomic propositions that are true at that world, by listing the true atomic propositions underneath the world(s) they are true at. An example is given in Figure 11.2; if an atom is *not* listed beneath a world, then we are to assume that the atom is false at that world. In this model, $\mathfrak{M}, x \vDash \neg q$, and $\mathfrak{M}, y \vDash q$. Even though q is false at x, because y is accessible to x and it is true there, q is possible at x, e.g., $\mathfrak{M}, x \vDash \Diamond q$.

Figure 11.2: An example model

Sometimes it is useful to list not only those that are true at a world, but also those that are false, in which case, the formulas annotating a world are themselves annotated by a truth value; Figure 11.2 is adapted to be annotated this way in Figure 11.3. Note that even in Figure 11.3, we have not exhaustively listed all atomic propositions (indeed, this wouldn't be plausible, since we have infinitely many of them); we have only written down and given the truth values for the atoms that are relevant to a given proposition.

Figure 11.3: An example model with truth values explicitly annotated

Once we have defined truth at a world in a model, we can extend the propositional definitions of validity, inconsistency, and satisfiability (cf. Definitions 7.4.8, 7.4.9, 7.4.12) to

Figure 11.4: The non-truthfunctionality of modal truth

cover these modal models. We give definitions for the notions below in order of strength:

Definition 11.3.4.

1. A wff φ is *satisfiable* if there is a model $\mathfrak{M}$ and $w \in W$ such that $\mathfrak{M}, w \vDash \varphi$. A satisfiable wff is also called *consistent*.

2. A wff φ is *valid in a model* $\mathfrak{M}$ if for every $w \in W$, $\mathfrak{M}, w \vDash \varphi$. We then write $\mathfrak{M} \vDash \varphi$.

3. A wff φ is *valid on a frame* $\mathfrak{F}$ if φ is valid in every model based on $\mathfrak{F}$. We then write $\mathfrak{F} \vDash \varphi$.

4. A wff φ is *valid in a class of frames* $\mathfrak{C}$ if φ is valid on every frame $\mathfrak{F}$ in $\mathfrak{C}$. We then write $\mathfrak{C} \vDash \varphi$.

5. A wff φ is *valid* if φ is valid in every class of frames $\mathfrak{C}$. We then write $\vDash \varphi$.

Note 11.3.5. We sometimes call validity of the last, strongest type K-validity (for 'Kripke').

Corollary 11.3.6. *A wff φ is* invalid *if its negation is satisfiable.*

Proof. Suppose that $\neg\varphi$ is satisfiable. Then there is a model $\mathfrak{M} = \langle \mathfrak{F}, V \rangle$ and a world $w \in W$ such that $\mathfrak{M}, w \vDash \neg\varphi$. This means that φ is not valid on the frame $\mathfrak{F}$, and that there is at least one class of frames $\mathfrak{C}$ (namely, the class containing just $\mathfrak{F}$!) that contains a frame on which φ is not valid. Therefore, there is at least one class of frames in which φ is not valid; therefore, it is invalid. □

Given the truth conditions for $\Box$ and $\Diamond$, we can prove a number of straightforward results about these two operators and their relationship, including:

Lemma 11.3.7. *For every $\mathfrak{M}$, and every w in $\mathfrak{M}$:*

$$\mathfrak{M}, w \vDash \Box\varphi \quad \textit{iff} \quad \mathfrak{M}, w \vDash \neg \Diamond \neg\varphi$$

Proof. Because this is an 'if and only if' claim, we must prove both directions.

($\Rightarrow$) Assume that $\mathfrak{M}, w \vDash \Box\varphi$. Then for every w' such that wRw', $\mathfrak{M}, w' \vDash \varphi$. Now, suppose the opposite of what we want to prove, namely, $\mathfrak{M}, w \nvDash \neg \Diamond \neg\varphi$. By the definition of truth, it follows that $\mathfrak{M}, w \vDash \Diamond\neg\varphi$, which means that there is some w'' such that wRw'' and $\mathfrak{M}, w'' \vDash \neg\varphi$. But that means $\mathfrak{M}, w'' \nvDash \varphi$, which contradicts the fact that $\mathfrak{M}, w' \vDash \varphi$ for all w' such that wRw', including w''.

($\Leftarrow$) See Exercise 111. □

Corollary 11.3.8.

$$\vDash \Box p \leftrightarrow \neg \Diamond \neg p$$

As a consequence, we say that $\Box$ and $\Diamond$ are *duals* of each other.

Corollary 11.3.9. *For every $\mathfrak{M}$, and every w in $\mathfrak{M}$, if there is no w' such that wRw', then for any wff φ:*

$$\mathfrak{M}, w \vDash \Box\varphi$$

That is, every wff is necessary at a dead-end (and correspondingly, no formula is possible).

Proving validity and invalidity

The validity and invalidity of modal formulas is proven the same way as for propositional formulas, except that instead of looking at a single truth value assignment, we must look at multiple ones (i.e., multiple possible worlds). But just as in the purely propositional case, validity is a *universal* notion—requiring truth on *every* truth-value assignment—and invalidity is an *existential* notion—requiring truth on just *some* truth-value assignment. This is why, when trying to prove validity, it is easier to start from the assumption of invalidity, because invalidity is easier to demonstrate. If we succeed, then we have shown that the formula in question is invalid; if we reach a contradiction, then we know that it is valid.

We now go through a number of examples demonstrating how validity or invalidity can be shown.

Lemma 11.3.10. $\Box(p \to q) \to (\Box p \to \Box q)$ *is valid.*

Proof. Suppose it isn't. Then there is a class of frames $\mathfrak{C}$ and a frame $\mathfrak{F}$ in $\mathfrak{C}$ and a model $\mathfrak{M}$ on $\mathfrak{F}$ and world w in $\mathfrak{M}$[4] such that $\mathfrak{M}, w \not\vDash \Box(p \to q) \to (\Box p \to \Box q)$. That is:

$$\mathfrak{M}, w \vDash \Box(p \to q) \tag{11.1}$$

$$\mathfrak{M}, w \not\vDash \Box p \to \Box q \tag{11.2}$$

$$\mathfrak{M}, w \vDash \Box p \text{ (by 11.2)} \tag{11.3}$$

$$\mathfrak{M}, w \not\vDash \Box q \text{ (by 11.2)} \tag{11.4}$$

To summarise the above facts in a diagram, we have:

$\Box(p \overset{w}{\to} \overset{\bullet}{q})$ T
$\Box p \to \Box q$ F
$\Box p$ T
$\Box q$ F

We then continue, updating our diagram along the way:

$$\mathfrak{M}, w \vDash \neg \Box q \text{ (by 11.4)} \tag{11.5}$$

$$\mathfrak{M}, w \vDash \Diamond \neg q \text{ (by 11.5)} \tag{11.6}$$

$\Box(p \overset{w}{\to} \overset{\bullet}{q})$ T
$\Box p \to \Box q$ F
$\Box p$ T
$\Box q$ F
$\neg \Box q$ T
$\Diamond \neg q$ T

11.6 means that there is another world, accessible to w (call it w') where $\neg q$ is true. That is:

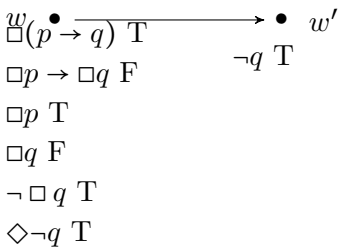
$$wRw' \text{ and } \mathfrak{M}, w' \vDash \neg q \tag{11.7}$$

and hence:

$$\mathfrak{M}, w' \not\vDash q \tag{11.8}$$

This requires us to draw a new world in the diagram, and an arrow from the first world to the new world:

[4] This is rather a mouthful. In the future, we will simply note that if a formula is invalid, then there is some model with a world where the formula is false, omitting reference to frames and classes of frames.

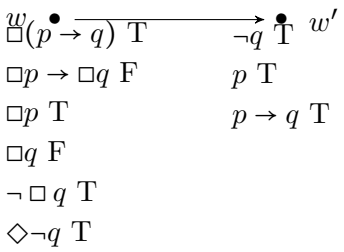


But now recall 11.1 and 11.3. Now that we have a new world to consider, w', which is accessible from w, we must take into account the truth conditions for the necessity formulas. That is:

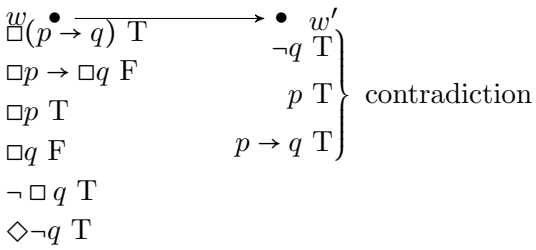
$$\mathfrak{M}, w' \vDash p \rightarrow q \text{ (by 11.1)} \tag{11.9}$$

$$\mathfrak{M}, w' \vDash p \text{ (by 11.3)} \tag{11.10}$$

This gives us the following diagram to consider:



But 11.9, 11.10, and 11.8 are inconsistent, as the diagram clearly illustrates. At no point in the construction of the model did we have any option or choice about what to do; thus, this contradiction shows that there is no way in which to create a model with a world where this formula is false:



□

When a formula is not valid, we can use this method to construct a countermodel, that is, a model with a world where the formula is false.

Lemma 11.3.11. $\Box(\Box(p \to \Box p) \to p)$ *is not valid.*

Proof. We prove that it is invalid by constructing a model $\mathfrak{M}$ and a world w in $\mathfrak{M}$ such that $\mathfrak{M}, w \not\vDash \Box(\Box(p \to \Box p) \to p)$, that is:

$$\mathfrak{M}, w \vDash \neg\, \Box\, (\Box(p \to \Box p) \to p) \tag{11.11}$$

If some formula is not necessary, then it is possibly false.

$$\mathfrak{M}, w \vDash \Diamond\neg(\Box(p \to \Box p) \to p) \tag{11.12}$$

$$\Diamond\neg(\Box(p \to \overset{w}{\bullet}\Box p) \to p)\ \mathrm{T}$$

By the truth conditions for $\Diamond$, this means there is another world x such that wRx and:

$$\mathfrak{M}, x \vDash \neg(\Box(p \to \Box p) \to p) \tag{11.13}$$

$$\Diamond\neg(\Box(\overset{w}{\bullet} p \to \Box p) \to p)\ \mathrm{T} \qquad \neg(\Box(p \to \overset{\bullet}{\longrightarrow} \Box p) \overset{x}{\bullet} \to p)\ \mathrm{T}$$

This is equivalent to:

$$\mathfrak{M}, x \vDash \Box(p \to \Box p) \tag{11.14}$$

and

$$\mathfrak{M}, x \vDash \neg p \tag{11.15}$$

$$\Diamond\neg(\Box(\overset{w}{\bullet} p \to \Box p) \to p)\ \mathrm{T} \qquad \Box(p \to \overset{\bullet}{\longrightarrow} \Box p) \overset{x}{\bullet}\ \mathrm{T}$$
$$p\ \mathrm{F}$$

Now, we must check that we've taken care of all the necessity formulas; there is one, in 11.14: If $\mathfrak{M}, x \vDash \Box(p \to \Box p)$, then every world that is accessible to x must model $\Box(p \to \Box p)$. But there are no worlds accessible to x. So $\mathfrak{M}, x \vDash \Box(p \to \Box p)$ is trivially true (cf. Corollary 11.3.9).

Thus, our countermodel is: $\langle \{w, x\}, \{\langle w, x\rangle\}\rangle$, where $V(p, x) = F$. □

Satisfiability is a much weaker notion, and the satisfiability of a wff φ can be demonstrated by constructing a model with a world where φ is true.

Lemma 11.3.12. $\Box(\Box(p \to \Box p) \to p)$ *is satisfiable.*

Proof. The main connective is $\Box$; every necessity wff is true at any world that doesn't see anything. Let $W = \{w\}$, R be empty, and V be arbitrary. Then $\mathfrak{M} = \langle W, R, V\rangle, w \vDash \Box(\Box(p \to \Box p) \to p)$. □

Frame properties

Our ultimate goal in developing modal logic is to capture different properties of different types of modality in a formal system of logic. Some informal candidate properties include the following:

1. What is true is possible.
2. What is necessary is true.
3. What is necessary is possible.
4. What is necessary is necessarily necessary.
5. What is necessary is necessarily possible.
6. What is possible is necessarily possible.

Each of these principles can be rendered in our formal language):

1′. $\varphi \to \Diamond\varphi$.

2′. $\Box\varphi \to \varphi$.

3′. $\Box\varphi \to \Diamond\varphi$.

4′. $\Box\varphi \to \Box\,\Box\,\varphi$.

5′. $\Box\varphi \to \Box\,\Diamond\,\varphi$.

6′. $\Diamond\varphi \to \Box\,\Diamond\,\varphi$.

Which of these modal principles are plausible depends on the type of modality under investigation. If we are interested in logical necessity, then the second principle, that necessity implies truth, seems plausible. If, however, we are interested in deontic necessity, that same principle does not seem reasonable; there are (sadly) many things which ought to be the case that nevertheless yet fail to be the case. (Everyone ought to hand in their homework; and yet, there will always be someone who doesn't.)

When we introduced the idea of possible worlds and accessibility earlier in this chapter, we spoke of limiting our attention to *relevant* possible worlds, with relevance depending on the particular type of necessity expressed. In the context of possible worlds semantics, the way to constrain worlds to the relevant ones is by putting constraints on the accessibility relation R.

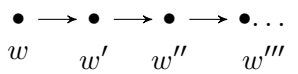
What we show next is that each of these different formulas corresponds (in a very strong notion of "correspond") with different classes for frames each picked out by a specific set of properties that the accessibility relation has. Some common properties are the following:

Definition 11.3.13 (Reflexivity)**.** A binary relation R on $W \times W$ is *reflexive* iff $\forall x(xRx)$.

Definition 11.3.14 (Irreflexivity)**.** A binary relation R on $W \times W$ is *irreflexive* iff $\forall x \neg(xRx)$.

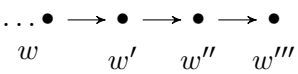
Note 11.3.15. *Irreflexivity* is not the same as *nonreflexivity*, which is the negation of *reflexivity*, i.e., $\exists x \neg(xRx)$.

Definition 11.3.16 (Seriality)**.** A binary relation R on $W \times W$ is *serial* iff $\forall x \exists y(xRy)$; this is the same as saying it has 'no last point'.



Corollary 11.3.17. *Every reflexive relation is serial.*

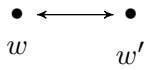
Definition 11.3.18 (No first point)**.** A binary relation R on $W \times W$ has *no first point* iff $\forall x \exists y(yRx)$.



Definition 11.3.19 (Transitivity)**.** A binary relation R on $W \times W$ is *transitive* iff $\forall x \forall y \forall z((xRy \wedge yRz) \rightarrow xRz)$.

An example of a natural ordering which is transitive is the "in the future of" relation: If t' is in the future of t and t'' is in the future of t', then t'' is in the future of t.

Definition 11.3.20 (Symmetry)**.** A binary relation R on $W \times W$ is *symmetric* iff $\forall xy(xRy \to yRx)$.



Definition 11.3.21 (Antisymmetry)**.** A binary relation R on $W \times W$ is *antisymmetric* iff $\forall xy(xRy \to \neg(yRx))$.

Note 11.3.22. *Antisymmetry* is not the same as *asymmetry*, which is the negation of *symmetry*, i.e., $\exists xy(xRy \wedge \neg(yRx))$.

An example of a natural ordering which is antisymmetric is the temporal relation "after": If t' is after t, it is never the case that t is after t'.

Note 11.3.23 (Pre-order)**.** A binary relation which is (a) reflexive, (b) transitive, and (c) antisymmetric is called a *pre-order*.

Definition 11.3.24 (Equivalence)**.** A binary relation R on $W \times W$ is an *equivalence relation* iff it is reflexive, symmetric, and transitive.

Definition 11.3.25 (Universal)**.** A binary relation R on $W \times W$ is a *universal* iff $\forall xy(xRy)$.

Corollary 11.3.26. *Every universal relation is an equivalence relation.*

Note 11.3.27. But not every equivalence relation is a universal relation, as demonstrated by the frame in Figure 11.5.

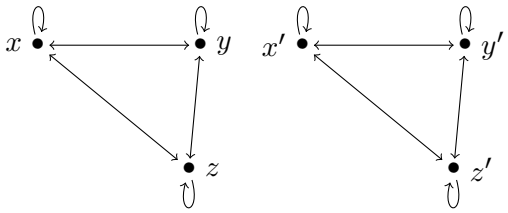


Figure 11.5: $\mathfrak{F} = \langle W, R \rangle$ where R is an equivalence relation but not universal.

Definition 11.3.28 (Density)**.** A binary relation R on $W \times W$ is *dense* iff $\forall xy(xRy \rightarrow \exists z(xRz \wedge zRy))$.

Note 11.3.29. This definition of density does not require that $x \neq z$ (following [Chagrov and Zakharyaschev pp. 65, 79]), and as a result in the presense of cycles (including reflexive points) may not correspond to our intuitions of density taken from, e.g., the realm of mathematics.

In many cases, density will only occur in conjunction with another property, such as irreflexivitiy or strict linearity.

Definition 11.3.30 (Backwards linearity)**.** A binary relation R on $W \times W$ is *backwards linear* iff $\forall x \forall y \forall z((xRz \wedge yRz) \rightarrow (xRy \vee yRx \vee x = y))$.

Definition 11.3.31 (Forwards linearity)**.** A binary relation R on $W \times W$ is *forwards linear* iff $\forall x \forall y \forall z((zRx \wedge zRy) \rightarrow (xRy \vee yRx \vee x = y))$.

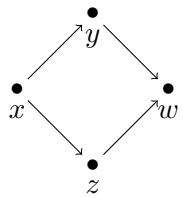
A relation that is both backwards and forwards linear is simply *linear*. It is useful to separate backwards and forwards linearity when modeling time (cf. §13.2.2), where under many natural interpretations of the structure of time, it is linear in the past but possibly not-linear (i.e., branching in the future).

Definition 11.3.32 (Trichotomy)**.** A binary relation R on $W \times W$ is *trichotomous* iff $\forall xy(xRy \vee yRx \vee x = y)$.

Note 11.3.33. In the presence of reflexivity, trichotomy is called "connectedness".

Definition 11.3.34 (Strong connectedness)**.** A binary relation R on $W \times W$ is *strongly connected* iff $\forall x \forall y \forall z((xRy \wedge xRz) \rightarrow (yRz \vee zRy))$.

Definition 11.3.35 (Convergence)**.** A binary relation R on $W \times W$ is *convergent* iff $\forall x \forall y \forall z((xRy \wedge xRz) \rightarrow \exists w(yRw \wedge zRw))$. Convergence is sometimes called the *diamond property*.



Definition 11.3.36 (Euclidean)**.** A binary relation R on $W \times W$ is *Euclidean* iff $\forall x \forall y \forall z((xRy \wedge xRz) \rightarrow yRz)$.

Definition 11.3.37 (Directed)**.** A binary relation R on $W \times W$ is *directed* iff $\forall x \forall y \forall z((xRy \wedge xRz \wedge y \neq z) \rightarrow \exists x'(yRx' \wedge zRx'))$.

We now explore how some of these constraints on R correspond to certain modal formulas. The sense of correspondence that we are interested in is this: For various constraints on R, we can prove that some particular formula is valid on a class of frames if and only if that class of frames satisfies the constraint.[5]

Lemma 11.3.38. *The wff $\Box p \rightarrow p$ is valid on a frame $\mathfrak{F} = \langle W, R \rangle$ iff R is reflexive.*

Proof. Because the claim is an if-and-only-if claim, we must prove both that if the left-hand claim is true, the right-hand is true, and vice versa. We first prove the right-to-left claim and then the left-to-right claim.

$\Leftarrow$ We prove that if $\mathfrak{F} = \langle W, R \rangle$ is such that R is reflexive, then $\Box p \rightarrow p$ is valid on $\mathfrak{F}$. Assume that R is reflexive, but that $\Box p \rightarrow p$ is not valid in $\mathfrak{F}$. This means there is some model $\mathfrak{M}$ on $\mathfrak{F}$ and $w \in W$ such that $w \nvDash \Box p \rightarrow p$; that is, (1) $w \vDash \Box p$ and $w \nvDash p$, that is, (2) $w \vDash \neg p$. But, by (1), p is true at every world w sees, and wRw. So, (3) $w \vDash p$; and that is a contradiction with (2).

$\Rightarrow$ Proved contrapositively. Suppose $\mathfrak{F} = \langle W, R \rangle$ is such that R is not reflexive. Then, there is some $w \in W$ such that $\neg(wRw)$. Let $\mathfrak{M}$ be a model on $\mathfrak{F}$ where $V(p, w') = T$ for every world w' in W except for w. Then, (1) $w \nvDash p$. w either seems some world or it sees no worlds. If it sees no worlds, then (2) $w \vDash \Box p$, and hence $\Box p \rightarrow p$ is not valid; or (3) it does see at least one world, but by the way we constructed the particular model on this frame, every world it sees is a world where p is true, and thus $w \vDash \Box p$, and again, the wff $\Box p \rightarrow p$ is not valid.

□

Corollary 11.3.39. *The wff $p \rightarrow \Diamond p$ is valid on a frame $\mathfrak{F} = \langle W, R \rangle$ iff R is reflexive.*

Lemma 11.3.40. *The wff $\Box p \rightarrow \Diamond p$ is valid on a frame $\mathfrak{F} = \langle W, R \rangle$ iff R is serial.*

Proof. Again we must prove both the left-to-right and right-to-left claims.

$\Leftarrow$ Let $\mathfrak{F} = \langle W, R \rangle$ be such that R is serial, but suppose that $\Box p \rightarrow \Diamond p$ is not valid on $\mathfrak{F}$. Then there is a model $\mathfrak{M}$ on $\mathfrak{F}$ and some $w \in W$ such that $w \nvDash \Box p \rightarrow \Diamond p$; that is, (1) $w \vDash \Box p$ and $w \nvDash \Diamond p$, that is, (2) $w \vDash \neg \Diamond p$. If $w \vDash \neg \Diamond p$, then (3) $w \vDash \Box \neg p$. Since R

[5] Note that not every constraint on R corresponds to a modal formula: For example, there is no modal formula which is valid on all and only irreflexive frames [Hughes and Cresswell, 1996, p. 176].

is serial, there is a w' such that wRw'. By (1), $w' \vDash p$; by (3), $w' \vDash \neg p$. But this is a contradiction.

$\Rightarrow$ Proved contrapositively. Suppose that $\mathfrak{F} = \langle W, R\rangle$ is such that R is not serial. Then, there is some $w \in W$ that is a dead end. Let $\mathfrak{M}$ be any model on $\mathfrak{F}$, with the valuation arbitrary. Either (1) $w \vDash p$ or (2) $w \vDash \neg p$. Either way, (3) $w \vDash \Box p$ and (4) $w \vDash \Box\neg p$. By (4), (5) $w \vDash \neg \Diamond p$. But then by (3) and (5), (6) $w \nvDash \Box p \rightarrow \Diamond p$, showing that $\Box p \rightarrow \Diamond p$ is not valid on this frame.

□

Corollary 11.3.41. *The wff $\Diamond(p \vee \neg p)$ is valid on a frame $\mathfrak{F} = \langle W, R\rangle$ iff R has no last point.*

Proof. $\Diamond(p \vee \neg p)$ is equivalent to $\Box p \rightarrow \Diamond p$, and the frame conditions are equivalent as well. □

Lemma 11.3.42. *The wff $\Box p \rightarrow \Box\Box p$ is valid on a frame $\mathfrak{F} = \langle W, R\rangle$ iff R is transitive.*

Lemma 11.3.43. *The wff $p \rightarrow \Box\Diamond p$ is valid on a frame $\mathfrak{F} = \langle W, R\rangle$ iff R is symmetric.*

Proof. Proofs of these two lemmas are analogous to the proofs of Lemmas 11.3.38 and 11.3.40, and left as exercises to the reader. □

Lemma 11.3.44. *The wff* $\mathsf{Geach} := \Diamond\Box p \rightarrow \Box\Diamond p$ *is valid on a frame $\mathfrak{F} = \langle W, R\rangle$ iff R is convergent.*

Proof. Again we must prove both the left-to-right and right-to-left claims.

$\Leftarrow$ Let $\mathfrak{F} = \langle W, R\rangle$ be such that R is convergent, but suppose that $\Diamond\Box p \rightarrow \Box\Diamond p$ is not valid on $\mathfrak{F}$. Then there is a model $\mathfrak{M}$ on $\mathfrak{F}$ and some $w \in W$ such that $w \nvDash \Diamond\Box p \rightarrow \Box\Diamond p$; that is, (1) $w \vDash \Diamond\Box p$ and $w \nvDash \Box\Diamond p$, that is, (2) $w \vDash \neg\Box\Diamond p$. By (1), there is a world w' such that wRw' and $w' \vDash \Box p$. By (2), $w \vDash \Diamond\Box\neg p$, so there is another world w'' such that wRw'' and $w'' \vDash \Box\neg p$. Now, because R is convergent we know that there is a world v such that (3) $w'Rv$ and (4) $w''Rv$. From (3), we know that $v \vDash p$; but from (4) we get that $v \vDash \neg p$, which is a contradiction.

$\Rightarrow$ Proved contrapositively. Suppose that $\mathfrak{F} = \langle W, R\rangle$ is such that R is not convergent. Then, there is some $w, w', w'' \in W$ where wRw', wRw'', and there is no world that both w' and w'' both see. We define a model $\mathfrak{M}$ on $\mathfrak{F}$ with the following valuation: For every v such that $w'Rv$, $V(p, v) = T$. For every u such that $w''Ru$, $V(p, u) = F$.

Because there is no x that is seen by both w' and w'', this is a consistent valuation. It also ensures that $V(\Box p, w') = T$ and $V(\Box\neg p, w'') = T$. Since wRw', $w \vDash \Diamond\Box p$; because wRw'', $w \vDash \Diamond\Box\neg p$; thus, $w \nvDash \Diamond\Box p \to \Box\Diamond p$.

□

Lemma 11.3.45. *The wff* $\mathsf{E} := \Diamond p \to \Box\Diamond p$ *is valid on a frame* $\mathfrak{F} = \langle W, R\rangle$ *iff* R *is Euclidean.*

Proof. Again we must prove both the left-to-right and right-to-left claims.

⇐ Let $\mathfrak{F} = \langle W, R\rangle$ be such that R is Euclidean, but suppose that $\Diamond p \to \Box\Diamond p$ is not valid on $\mathfrak{F}$. Then there is a model $\mathfrak{M}$ on $\mathfrak{F}$ and some $w \in W$ such that $w \nvDash \Diamond p \to \Box\Diamond p$; that is, (1) $w \vDash \Diamond p$ and $w \nvDash \Box\Diamond p$, that is, (2) $w \vDash \neg\Box\Diamond p$. By (1), there is a world w' such that wRw' and $w' \vDash p$. By (2), $w \vDash \Diamond\Box\neg p$, so there is another world w'' such that wRw'' and $w'' \vDash \Box\neg p$. Now, because R is Euclidean, we know that (3) $w''Rw'$. From (3), it follows that $w' \vDash \neg p$, which contradicts (1).

⇒ Proved contrapositively. Suppose that $\mathfrak{F} = \langle W, R\rangle$ is such that R is not Euclidean. Then, there is some $w, w', w'' \in W$ where wRw' and wRw'', neither $w'Rw''$ or $w''Rw'$. We define a model $\mathfrak{M}$ on $\mathfrak{F}$ with the following valuation: Let $V(p, w') = T$; then $w \vDash \Diamond p$. Where $x \neq w'$, let $V(p, x) = F$. It then follows that every world that w'' can see makes $\neg p$ true, since the only world where p is true is w', which w'' cannot see by hypothesis. Therefore, $w'' \vDash \Box\neg p$, and hence $w \vDash \Diamond\Box\neg p$, which suffices to show that $\Diamond p \to \Box\Diamond p$ is false at w, and hence invalid on the frame.

□

To be added: a discussion of Goldblatt-Thomasson results

11.3.2 Neighborhood semantics

All logics interpreted on Kripke frames are *normal* modal logics, which is to say they satisfy the rule of necessitation and K. If we wish to study logics that do not validate K or the rule of necessitation, then we must use a different type of semantics. In this section, we introduce neighborhood semantics for non-normal modal logics (as well as normal modal logics, since it can handle both.)

11.3.3 The standard translation, part 2

Theorem 11.3.46.

$$\mathfrak{M} \vDash \varphi \quad \textit{iff} \quad \langle W, P_i, R\,;\, i \in \mathbb{N}\rangle \vDash \forall x\, \mathrm{ST}_x(\varphi)$$

Proof. To be written. □

To be completed

11.4 Proof theory

In this section, we outline a general method for reasoning about modality in a formal system of logic. The formal system is based on the language and the construction rules for wffs given in Definitions 11.2.1 and 11.2.2. To this we will add a selection of wffs as *axioms* (which wffs we pick will depend on the concept of modality that we are trying to explicate), and a set of *transformation rules* that allow us to change wffs into other wffs. Not every set of transformation rules will be a reasonable choice for study. In particular, we will only be interested in rules that are *sound*, that is, validity-preserving. We will discuss this further below. First we define a few basic notions and outline some desiderata.

Definition 11.4.1 (Axiomatic base)**.** An *axiomatic base* is a set of wffs Γ.

Axiomatic bases can contain both modal and non-modal wff (that is, ones that contain modal operators and ones that do not). We will sometimes talk about the 'modal base' of a logical system; these are simply the members of the axiomatic base that contain modal operators.

What we need to make an axiomatic base a formal and mechanistic system of proof is the following:

1. A mechanism to tell of a string whether it is a wff.
2. A mechanism to tell of a wff whether is an axiom.
3. A mechanism to determine if an application of a rule is a genuine application.

Definition 11.2.2 gives us a mechanism for determining (1). If an axiomatic base Γ is finite, then we have an easy mechanism for (2): For any given wff φ, simply check to see if it is in Γ. If it is, then it is an axiom. If it isn't, then it is not an axiom. In this chapter, we will only be concerned with logical systems that have finite modal bases; the non-modal part of the base may be infinite, but we will provide a finite mechanism for determining of

any non-modal wff whether it is an axiom or not. Meeting requirement (3) will be done by specifying our rules in such a way that there is no ambiguity in how they can be applied.

Given an axiomatic basis Γ and a set of transformation rules, we can define a notion of *proof*.

Definition 11.4.2 (Proof)**.** A *proof* is a numbered series of wffs, every wff of which is either an axiom or derived from one or more wffs earlier in the series by applications of one of the transformation rules.

Definition 11.4.3 (Theorem)**.** If a wff φ appears in a proof from an axiomatic base Γ, we write $\vdash_\Gamma \varphi$, and we say that φ is a *theorem*.

Where no ambiguity results, we will omit the subscript Γ.

Definition 11.4.4 (Provable equivalence)**.** If $\vdash \varphi \leftrightarrow \psi$, we say that φ and ψ are *provably equivalent*.

Definition 11.4.5 (System)**.** A *logical system* S is the set of theorems generated from an axiomatic base.

We will sometimes write $\vdash_{\mathrm{S}} \varphi$ for theoremhood in system S, rather than subscripting the axiomatic base.

Definition 11.4.6. Two systems S and S$'$ with different axiomatic bases are called *deductively equivalent* if they contain the same theorems.

Definition 11.4.7. If every theorem of S is a theorem of S$'$, then S *is contained in* S$'$ and S$'$ *contains* S and S$'$ *extends* S.

Definition 11.4.8. If S$'$ contains S and S and S$'$ are not deductively equivalent, then S$'$ *properly extends* S. S$'$ is the stronger system, and S is the weaker system.

Recall the definition of consistency given in Definition 8.5.3:

Definition 11.4.9 (Consistency)**.** Given a set S of axioms (logical and non-logical), a set of formulas Λ is S-*inconsistent* iff there is a finite $\Gamma = \{\varphi_0, \ldots, \varphi_n\} \subseteq \Lambda$ such that $\vdash_{\mathrm{S}} \neg(\varphi_0 \wedge \cdots \wedge \varphi_n)$, that is, in the system S is it possible to prove the negation of the conjunction of (a subset of) Λ. If this is not possible, then Λ is S-*consistent*.

We note some alternative definitions of consistency:

Definition 11.4.10. Other types of consistency:

- A system S is *absolutely consistent* if there is some wff φ such that $\nvdash_S \varphi$.
- A system S is *atomically consistent* if there is some atom p such that $\nvdash_S p$.
- A system S is *negation-consistent* if it is not the case that both $\vdash_S \varphi$ and $\vdash_S \neg\varphi$.

In classical contexts, where both the Law of Excluded Middle and the Principle of Non-Contradiction are present, when the underlying proof-system is complete with respect to the propositional connectives, some of these definitions are equivalent:

Theorem 11.4.11. *Let* S *be a logical system. The following are equivalent:*

1. S *is* S*-consistent.*
2. S *is absolutely consistent.*
3. S *is negation-consistent.*

Proof.

- (1) $\Rightarrow$ (3): If S is S-consistent, then there is no finite $\Gamma = \{\varphi_0, \ldots, \varphi_n\} \subseteq S$ such that $\vdash_S \neg(\varphi_0 \wedge \cdots \wedge \varphi_n)$. Suppose that S is negation-*in*consistent. Then there is a formula φ such that $\vdash_S \varphi$ and $\vdash_S \neg\varphi$. It follows that $\vdash_S \varphi \wedge \neg\varphi$. However, since S proves every propositional tautology, $\vdash_S \neg(\varphi \wedge \neg\varphi)$, so S is in fact S-inconsistent.
- (2) $\Rightarrow$ (1): If S is absolutely consistent, then there is some wff ψ such that $\nvdash_S \psi$. Suppose that S were S-*in*consistent. Then there is a finite $\Gamma = \{\varphi_0, \ldots, \varphi_n\} \subseteq S$ such that $\vdash_S \neg(\varphi_0 \wedge \cdots \wedge \varphi_n)$. However, since each $\varphi_n \in S$, $\vdash_S \varphi_n$, that is, every φ_n is an S-theorem. It therefore follows that $\vdash_S \varphi_0 \wedge \cdots \wedge \varphi_n$. Since we are assuming the PNC holds and the system contains all propositional tautologies, $\vdash_S ((\varphi_0 \wedge \cdots \wedge \varphi_n) \wedge \neg(\varphi_0 \wedge \cdots \wedge \varphi_n)) \rightarrow \psi$, and it follows that $\vdash_S \psi$, which contradicts the assumption of absolute consistency.
- (3) $\Rightarrow$ (2): If S is negation-consistent, then for every formula φ, either $\nvdash_S \varphi$ or $\nvdash_S \neg\varphi$, and hence S is absolutely consistent.

□

But for our purposes, the definition of consistency that we will make use of the most is one in Definition 8.5.3.

11.4.1 The system K

In this section we develop a proof system for the weakest normal[6] modal logic, the system K (named for Kripke). This system will form the foundation for all the other systems we look at in this chapter; every other system will be an extension of K. As noted at the beginning of this section, we need to identify the *axiomatic base* (or '*axioms*') of K and the acceptable transformation rules.

Definition 11.4.12 (Axioms of K)**.** The axioms of K are:

PL If φ is a valid wff of propositional logic, then φ is an axiom of K.

K $\Box(p \to q) \to (\Box p \to \Box q)$.

Note that PL is an infinite set, as there are infinitely many tautologies in propositional logic. However, because each propositional formula is finite in length, they will all have finite truth tables, and thus checking whether any given formula meets Definition 7.4.8 is a finite process. Therefore, this axiomatic basis meets the desideratum that we be able to tell, in a mechanistic way, whether something is an axiom of K or not; first, check if it is a propositional tautology. If it is, it is an axiom. If it is not, then check to see whether it is identical with K. If it isn't, then it is not an axiom. A list of propositional tautologies commonly used in modal proofs are listed in Table 7.10.

K is the only axiom of the system to contain a modal operator; because of this, we call it the "characteristic axiom" of K.

Every axiom of K is also a theorem of K, as any axiom can be written down in a proof at any time. We generate non-axiom theorems via the transformation rules (rules of inference).

The system K has three primitive transformation rules:

Rule 11.4.13 (Uniform Substitution (US))**.** The result of uniformly replacing atoms $p_1, \ldots, p_n$ by wffs $\beta_1, \ldots, \beta_n$ in a theorem is also a theorem.

We will often write the result of replacing atoms $p_1, \ldots, p_n$ uniformly with wffs $\beta_1, \ldots, \beta_n$ in formula φ by either $\varphi[\beta_1, \ldots, \beta_n/p_1, \ldots, p_n]$ or $\varphi[\beta_1/p_1, \ldots, \beta_n/p_n]$. We call both $\varphi[\beta_1, \ldots, \beta_n/p_1, \ldots, p_n]$ and $\varphi[\beta_1/p_1, \ldots, \beta_n/p_n]$ a *substitution instance* of φ.

Example 11.4.14. $\Box(r \to s) \to (\Box r \to \Box s)$ is a theorem, as it can be obtained from K by US, substituting r for p and s for q.

Example 11.4.15. $\Box((p \wedge q) \to (s \to r)) \to (\Box(p \wedge q) \to \Box(s \to r))$ is a theorem, as it can be obtained from K by US, substituting $p \wedge q$ for p and $s \to r$ for q.

[6] Cf. Definition 11.4.18.

Rule 11.4.16 (Modus ponens (MP))**.** If φ is a theorem and $\varphi \to \psi$ is a theorem, then so is ψ.

Neither US nor MP is a specifically modal rule; both of these hold for pretty much any logical system whatsoever.[7] Our third transformation rule is a specifically modal rule, and the only specifically modal rule:

Rule 11.4.17 (Rule of Necessitation (N))**.** If φ is a theorem, then $\Box\varphi$ is a theorem.

Definition 11.4.18. A *normal* modal logic or modal system is any system which is an extension of K, that is, any modal system whose axiomatic basis contains all the axioms of K and which contains at least US, MP, and N.

Non-normal modal logics drop either the K-axiom (such logics are called 'sub-K' logics) or one or more propositional tautologies (for example, intuitionistic modal logic does not include $p \vee \neg p$ as an axiom, see Chapter 19), or do not validate one of the three rules of inference (usually it is the specifically modal rule N that is dropped; however, some examples of logics that place restriction on US can also be found in the literature. Closure under MP is, however, almost universally required).

Before we start applying these rules to our axioms to derive new theorems, it's worth pausing to evaluate the quality of the axioms and rules that we've identified—there is no point in proving theorems from a set of axioms via a set of rules if we don't have some sort of guarantee of the quality of the theorems that we'll be deriving.

We already know that any non-modal axiom is going to be valid on the class of all frames, because the only non-modal axioms are the valid propositional wffs. We proved in Lemma 11.3.10 that K is valid in the class of all Kripke frames. All that is left to do is to show that the rules we've chosen preserve validity, that is, if the rules are applied to valid theorems, then the resulting theorem is also a valid on the class of all Kripke frames.

Lemma 11.4.19 (Soundness of MP)**.** *MP is sound with respect to the class of all Kripke frames.*

Proof. Rule MP states that if $\vdash \varphi$ and $\vdash \varphi \to \psi$, then $\vdash \psi$. To prove that this rule is sound, we must prove that validity is preserved, that is, if $\vDash \varphi$ and $\vDash \varphi \to \psi$ then $\vDash \psi$. We prove

[7] In fact, some logicians have even gone so far as to say that one can *define* a logical system to be a set of formulas that is closed under MP and US (that is, if φ and $\varphi \to \psi$ are in the set, then so is ψ, and that if φ is in the set, then every substitution instance of it is also in the set). While the vast majority of logical systems, both classical and non-classical, satisfy this definition, there are some important systems which do *not* validate US. A brief discussion of these systems, and an argument against including closure under US as a defining characteristic of a logic, can be found in [Uckelman et al., 2014].

this by assuming the antecedent of what we want to show along with the denial of the consequent, and show that this leads to a contradiction.

Suppose that (1) $\vDash \varphi$ and (2) $\vDash \varphi \to \psi$, but (3) $\nvDash \psi$. By (3), there is a frame $\mathfrak{F}$ and a model $\mathfrak{M}$ on $\mathfrak{F}$ and a world w in $\mathfrak{M}$ such that (4) $\mathfrak{M}, w \nvDash \psi$. However, (1) and (2) imply that (5) $\mathfrak{M}, w \vDash \varphi$ and (6) $\mathfrak{M}, w \vDash \varphi \to \psi$, and (4), (5), and (6) are jointly inconsistent. □

Lemma 11.4.20 (Soundness of US)**.** *US is sound with respect to the class of all Kripke frames.*

Proof. Rule US states that any substitution instance $\varphi[\beta_1, \ldots, \beta_n/p_1, \ldots, p_n]$ of a theorem φ is also a theorem. To prove that this rule is sound, we must prove that validity is preserved, that is, if $\vDash \varphi$ then $\vDash \varphi[\beta_1, \ldots, \beta_n/p_1, \ldots, p_n]$. We prove this contrapositively. Suppose that (1) $\vDash \varphi$ but (2) $\nvDash \varphi[\beta_1, \ldots, \beta_n/p_1, \ldots, p_n]$. By (2) there is a frame $\mathfrak{F}$ and a model $\mathfrak{M}$ on $\mathfrak{F}$ and a world w in $\mathfrak{M}$ such that (3) $\mathfrak{M}, w \nvDash \varphi[\beta_1, \ldots, \beta_n/p_1, \ldots, p_n]$. Let $\mathfrak{M}'$ be a model exactly like $\mathfrak{M}$ except that instead of V we use V', defined as follows:

$$\text{For } p_i,\ 1 \leq i \leq n, \begin{cases} V'(p_i, w) = T \text{ if } \mathfrak{M}, w \vDash \beta_i \\ V'(p_i, w) = F \text{ if } \mathfrak{M}, w \nvDash \beta_i \end{cases}$$

and for every other atom q, $V(q, w)$ is arbitrary. (In all other cases, $V = V'$.)

But then (4) $\mathfrak{M}', w \nvDash \varphi$, which contradicts (1). □

Example 11.4.21. Consider the formula $\varphi := r \to s$ and the substitution instance $\varphi[(p \wedge q)/r]$. Let $\mathfrak{M} = \langle W, R, V \rangle$ be defined as follows:

- $W = \{w\}$
- $R = \varnothing$
- $V(p, w) = V(q, w) = T$, $V(r, w) = V(s, w) = F$, and everything else arbitrary.

That is, $\mathfrak{M}$ consists of a single irreflexive world w where:

$$\begin{array}{lcl} \mathfrak{M}, w & \vDash & p \wedge q \\ \mathfrak{M}, w & \vDash & r \to s \\ \mathfrak{M}, w & \nvDash & (p \wedge q) \to r \end{array}$$

$$\begin{array}{l} w \ \bullet \\ p \wedge q \ \mathrm{T} \\ r \to s \ \mathrm{T} \\ (p \wedge q) \to r \ \mathrm{F} \end{array}$$

φ is valid on this model, but the substitution instance $\varphi[(p \wedge q)/r]$ is false at w. We use $\mathfrak{M}$ to construct a new model, $\mathfrak{M}'$ to show that φ, while valid on $\mathfrak{M}$, is *not* valid, tout court. Let $\mathfrak{M}' = \langle W', R', V' \rangle$ be constructed as follows:

- $W' = W$
- $R' = R$
- For r,

$$\begin{cases} V'(r, w) = T \text{ if } \mathfrak{M}, w \vDash p \wedge q \\ V'(r, w) = F \text{ if } \mathfrak{M}, w \nvDash p \wedge q \end{cases}$$

  and for every other atom p', $V'(p', w) = V(p', w)$, and for every other world, $V' = V$.

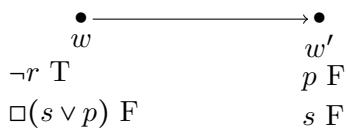
It is straightforward to check the $\mathfrak{M}', w \nvDash \varphi$, so φ is not valid.

Example 11.4.22. Consider the formula $\varphi := p \to q$ and the substitution instance $\varphi[\neg r/p, \Box(s \vee p)/q]$. We will give a model that makes $\varphi[\neg r/p, \Box(s \vee p)/q]$ false and show how to construct a model that make φ false.

Let $\mathfrak{M} = \langle W, R, V \rangle$ be defined as follows:

- $W = \{w, w'\}$
- $R = \{\langle w, w' \rangle\}$
- $V(r, w) = V(s, w') = V(p, w') = F$, and everything else arbitrary.

That is, $\mathfrak{M}$ consists of two irreflexive worlds w, w' where:



It is clear that $\neg r \to \Box(p \vee s)$, that is, $\varphi[\neg r/p, \Box(s \vee p)/q]$, is false at w in this model, and hence that $\nvDash \varphi[\neg r/p, \Box(s \vee p)/q]$. We use $\mathfrak{M}$ to construct a new model, $\mathfrak{M}'$ to show that φ is not valid. Let $\mathfrak{M}' = \langle W', R', V' \rangle$ be constructed as follows: $W' = W$ and $R' = R$ (that is, the frame stays the same), and then we define V' as follows:

$$\begin{cases} V'(p, w) = T \text{ if } \mathfrak{M}, w \vDash \neg r \\ V'(p, w) = F \text{ if } \mathfrak{M}, w \nvDash \neg r \\ V'(q, w) = T \text{ if } \mathfrak{M}, w \vDash \Box(s \vee p) \\ V'(q, w) = F \text{ if } \mathfrak{M}, w \nvDash \Box(s \vee p) \end{cases}$$

and for every other atom p' and every world w'', $V'(p', w'') = V(p', w'')$.

It is straightforward to check the $\mathfrak{M}', w \not\models \varphi$, so φ is not valid.

Lemma 11.4.23 (Soundness of N)**.** *N is sound with respect to the class of all Kripke frames.*

Proof. Rule N states that if $\vdash \varphi$ then $\vdash \Box\varphi$. To prove that this rule is sound, we must prove that validity is preserved, that is, if $\models \varphi$ then $\models \Box\varphi$. We prove this by assuming the antecedent of what we want to show along with the denial of the consequent, and show that this leads to a contradiction.

Suppose that (1) $\models \varphi$, but (2) $\not\models \Box\varphi$. By (2), there is a frame $\mathfrak{F}$ and a model $\mathfrak{M}$ on $\mathfrak{F}$ and a world w in $\mathfrak{M}$ such that (3) $\mathfrak{M}, w \not\models \Box\varphi$. By the truth conditions for $\Box$, (3) implies that there is a world w' such that wRw' and (4) $\mathfrak{M}, w' \not\models \varphi$. But (4) contradicts the assumption that $\models \varphi$. □

Theorems of K

We now prove a number of theorems of K.

Theorem 11.4.24.

$$\vdash_{\mathrm{K}} \Box(p \wedge q) \to (\Box p \wedge \Box q) \tag{K1}$$

Proof.

1	$(p \wedge q) \to p$	PL
2	$\Box((p \wedge q) \to p)$	N, 1
3	$\Box(p \to q) \to (\Box p \to \Box q)$	K
4	$\Box((p \wedge q) \to p) \to (\Box(p \wedge q) \to \Box p)$	US, $3[p \wedge q/p, p/q]$
5	$\Box(p \wedge q) \to \Box p$	MP, 2, 4
6	$(p \wedge q) \to q$	PL
7	$\Box((p \wedge q) \to q)$	N, 6
8	$\Box((p \wedge q) \to q) \to (\Box(p \wedge q) \to \Box q)$	US, $3[p \wedge q/p]$
9	$\Box(p \wedge q) \to \Box q$	MP, 7, 8
10	$(p \to q) \to ((p \to r) \to (p \to (q \wedge r)))$	PL
11	$(\Box(p \wedge q) \to \Box p) \to ((\Box(p \wedge q) \to \Box q) \to (\Box(p \wedge q) \to (\Box p \wedge \Box q)))$	US, $10[\Box(p \wedge q)/p, \Box p/q, \Box q/r]$
12	$(\Box(p \wedge q) \to \Box q) \to (\Box(p \wedge q) \to (\Box p \wedge \Box q))$	MP, 5, 11
13	$\Box(p \wedge q) \to (\Box p \wedge \Box q)$	MP, 9, 12

□

Theorem 11.4.25.

$$\vdash_{\mathrm{K}} (\Box p \wedge \Box q) \to \Box(p \wedge q) \tag{K2}$$

Proof.

1	$p \to (q \to (p \wedge q))$	PL
2	$\Box(p \to (q \to (p \wedge q)))$	N, 1
3	$\Box(p \to q) \to (\Box p \to \Box q)$	K
4	$\Box(p \to (q \to (p \wedge q))) \to (\Box p \to \Box(q \to (p \wedge q)))$	US, 3$[q \to (p \wedge q)/q]$
5	$\Box p \to \Box(q \to (p \wedge q))$	MP, 2, 4
6	$\Box(q \to (p \wedge q)) \to (\Box q \to \Box(p \wedge q))$	US, 3$[q/p, p \wedge q/q]$
7	$(p \to q) \to ((q \to (r \to s)) \to ((p \wedge r) \to s))$	PL
8	$(\Box p \to \Box(q \to (p \wedge q))) \to$ $((\Box(q \to (p \wedge q)) \to (\Box q \to \Box(p \wedge q))) \to ((\Box p \wedge \Box q) \to \Box(p \wedge q)))$	US, 7$[\Box p/p, \Box(q \to (p \wedge q))/q,$ $\Box q/r, \Box(p \wedge q)/s]$
9	$(\Box(q \to (p \wedge q)) \to (\Box q \to \Box(p \wedge q))) \to ((\Box p \wedge \Box q) \to \Box(p \wedge q))$	MP, 5, 8
10	$(\Box p \wedge \Box q) \to \Box(p \wedge q)$	MP, 6, 9

□

K1 and K2 are converses of each other. Since we can prove each of them, we can also prove the equivalence

$$\vdash_{\mathrm{K}} (\Box(p \wedge q) \to (\Box p \wedge \Box q)) \leftrightarrow ((\Box p \wedge \Box q) \to \Box(p \wedge q)) \tag{11.16}$$

which is in turn equivalent to:

$$\vdash_{\mathrm{K}} \Box(p \wedge q) \leftrightarrow (\Box p \wedge \Box q) \tag{K3}$$

K3 is sometimes called the "Law of $\Box$-Distribution". This theorem can be proven by concatenating the proofs of K1 and K2 (renumbering the second proof accordingly), and then appealing to the following propositional tautology:

$$(p \to q) \to ((q \to p) \to (p \leftrightarrow q))$$

with $\Box(p \wedge q)$ substituted in for p and $\Box p \wedge \Box q$ substituted in for q, and then applying MP twice. This results in a rather annoyingly long proof, so we will adopt the following convention that allows us to simplify proofs going forward:

Convention 11.4.26. Any theorem that has already been proved can be used in any future proof, justified by referring to the number of the theorem (e.g., K2).

This convention allows us to prove each theorem only once. With this convention, we can prove K3 in a much simpler fashion:

Proof of K3.

1	$\Box(p \wedge q) \rightarrow (\Box p \wedge \Box q)$	K1
2	$(\Box p \wedge \Box q) \rightarrow \Box(p \wedge q)$	K2
3	$(p \rightarrow q) \rightarrow ((q \rightarrow p) \rightarrow (p \leftrightarrow q))$	PL
4	$(\Box(p \wedge q) \rightarrow (\Box p \wedge \Box q)) \rightarrow$ $(((\Box p \wedge \Box q) \rightarrow \Box(p \wedge q)) \rightarrow (\Box(p \wedge q) \leftrightarrow (\Box p \wedge \Box q)))$	US, $3[\Box(p \wedge q)/p, \Box p \wedge \Box q/q]$
5	$((\Box p \wedge \Box q) \rightarrow \Box(p \wedge q)) \rightarrow (\Box(p \wedge q) \leftrightarrow (\Box p \wedge \Box q))$	MP, $1, 4$
6	$\Box(p \wedge q) \leftrightarrow (\Box p \wedge \Box q)$	MP, $2, 5$

□

Conventions in proofs

We will adopt some other conventions to simplify proofs. It is quite common to start a proof by asserting some propositional tautology in the first step and then applying rule N in the next step. Going forward, instead of writing these as separate steps, we will omit the explicit statement of the propositional axiom and simply write down the necessitated form of it. The justification for such a move will be PL+N. Similarly, it is quite common in a proof to write down a propositional tautology and then substitute into it some modal formula. Going forward, we will also omit the explicit statement of the propositional or modal axiom and write down the substitution instance of it. The justification for such a move will be PL+US or US+K.

In a similar fashion, we can often compress purely propositional reasoning. For instance, suppose we have a proof containing some wff, or perhaps more than one, that implies another wff on purely propositional grounds (i.e., no modal reasoning involved). In order to move from the initial wff(s) to the wff is implied, one would have write down the relevant propositional tautology or a substitution instance of it, and then apply MP. For instance, using purely propositional reasoning, we have that we can prove the formula on line m if we have already proven the wffs on lines i and j:

i	$\varphi \rightarrow \psi$	
j	$\gamma \rightarrow \psi$	
k	$(\varphi \rightarrow \psi) \rightarrow ((\gamma \rightarrow \psi) \rightarrow ((\varphi \vee \gamma) \rightarrow \psi))$	PL
l	$(\gamma \rightarrow \psi) \rightarrow ((\varphi \vee \gamma) \rightarrow \psi)$	MP, i, k
m	$(\varphi \vee \gamma) \rightarrow \psi$	MP, j, l

(The formula on line k is simply the tautology that corresponds to the $\vee$E rule of propositional proof theory.) Instead of requiring the spelling out of all these propositional steps, however, we will allow ourselves to skip straight from lines i and j to m, and justify it with PL. (In fact, we will see an instance of this in the proof of Theorem K4 below.)

Another way that we can simplify proofs is by introducing new rules. These rules may be either purely propositional, or involve modal reasoning. We first consider the following two purely propositional rules, which are analogues of Rules 7.5.21 and 7.5.22. The proofs of these rules are straightforward and left as exercises for the readers.

Rule 11.4.27 (Provable equivalence)**.** If $\vdash \varphi \to \psi$ and $\vdash \psi \to \varphi$, then $\vdash \varphi \leftrightarrow \psi$.

Rule 11.4.28 (Transitivity of $\to$)**.** If $\vdash \varphi \to \psi$ and $\vdash \psi \to \chi$, then $\vdash \varphi \to \chi$.

Thus, from K1 and K2 we can conclude K3 directly by appeal to Rule 11.4.27.

It is the specifically modal rules that are of greater interest to us. The next rules we introduce are specifically modal rules, and will be given their own reference numbers, e.g., DR1, DR2, etc. The 'DR' here stands for "derived rules"; see p. 288 for the definition of "derived rule" and how derived rules are distinguished from admissable rules.

Rule 11.4.29 (DR1)**.** If $\vdash \varphi \to \psi$ then $\vdash \Box\varphi \to \Box\psi$.

Proof.

1	$\varphi \to \psi$	Given
2	$\Box(\varphi \to \psi)$	N, 1
3	$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	US, K$[\varphi/p, \psi/q]$
4	$\Box\varphi \to \Box\psi$	MP, 2, 3

□

Here we use the convention at line 3, by simply writing down a substitution instance of K without first writing down K, but *not* to elide the step from 1 to 2, because 1 is not a tautology.

Rule 11.4.30 (DR2)**.** If $\vdash \varphi \leftrightarrow \psi$ then $\vdash \Box\varphi \leftrightarrow \Box\psi$.

Proof.

1	$\varphi \leftrightarrow \psi$	Given
2	$\psi \to \varphi$	PL, 1
3	$\Box\psi \to \Box\varphi$	DR1, 2
4	$\varphi \to \psi$	PL, 1
5	$\Box\varphi \to \Box\psi$	DR1, 4
6	$\Box\varphi \leftrightarrow \Box\psi$	Rule 11.4.27, 3, 5

□

The next rule that we prove is one that essentially allow us to omit series of steps in proofs that are basically steps of substitution and *modus ponens*. Recall from above that US requires us to substitute formulas in for atoms *uniformly*, that is, the same way in all occurrences. Thus,

$$\Box((p \wedge r) \to q) \to (\Box(p \wedge r) \to \Box q) \tag{11.17}$$

is a substitution instance of K, but

$$\Box(p \to q) \to (\Box(p \wedge r) \to \Box q) \tag{11.18}$$

is not, because we have replaced only one occurrence of p by $p \wedge r$. The next derived rule that we introduce allows us to essentially perform *nonuniform* substitution, but only in a certain context: The formula that we substitute in must be provably equivalent (cf. Definition 11.4.4) to the formula being substituted out. (The reason why we don't want to allow cases such as (11.18) is that p and $p \wedge r$ are not provably equivalent.)

Rule 11.4.31 (Substitution of equivalences (SubsEquiv))**.** If $\vdash \varphi$ and $\vdash \chi \leftrightarrow \gamma$, and ψ differs from φ only in having γ in some places where φ has χ, then $\vdash \psi$. That is, if χ and γ are provably equivalent, then we can replace any occurrence of χ in a theorem φ with γ and the result, ψ, will still be a theorem.

There are two ways in which SubsEquiv differs from US:

- The formula that you are substituting for need not be an atom (as it must in US);
- The substitution need not be performed uniformly (as it must be in US).

Example 11.4.32. Given that $\vdash \neg\neg\varphi \leftrightarrow \varphi$, if $\vdash \varphi \to (\varphi \vee \psi)$, SubsEquiv allows us to derive both

$$\varphi \to (\neg\neg\varphi \vee \psi) \tag{11.19}$$

and

$$\neg\neg\varphi \to (\varphi \vee \psi) \tag{11.20}$$

as well as

$$\neg\neg\varphi \to (\neg\neg\varphi \vee \psi) \tag{11.21}$$

A more complicated example:

Example 11.4.33. Given that $\vdash \neg\neg\varphi \leftrightarrow \varphi$ and $\vdash \neg(\varphi \wedge \psi) \leftrightarrow (\neg\varphi \vee \neg\psi)$, we can derive via SubsEquiv

$$\neg(\neg\varphi \wedge \neg\psi) \leftrightarrow (\varphi \vee \psi) \tag{11.22}$$

Before we prove the soundness of this rule, we prove an intermediate result:

Lemma 11.4.34. *Every modal formula is equivalent to one containing only $\neg$, $\vee$, and $\Box$.*

Proof. Per Corollary 7.4.38, every propositional formula is equivalent to one containing just $\neg$ and $\vee$. Given the equivalence of $\Diamond$ and $\neg\Box\neg$, it follows that every modal formula is equivalent to one containing just $\neg$, $\vee$, and $\Box$. □

Lemma 11.4.35 (Soundness of SubsEquiv)**.** *SubsEquiv is sound with respect to the class of all Kripke frames.*

Proof. To prove this, we must show that if $\vDash \varphi$ and $\vDash \chi \leftrightarrow \gamma$, and ψ differs from φ only in having γ where φ has χ, then $\vDash \psi$.

The soundness of this rule is based on the following propositional equivalences and implications:

$$\begin{array}{lcl} (p \leftrightarrow q) & \leftrightarrow & (\neg p \leftrightarrow \neg q) \\ (p \leftrightarrow q) & \rightarrow & ((p \vee r) \leftrightarrow (q \vee r)) \\ (p \leftrightarrow q) & \rightarrow & ((r \vee p) \leftrightarrow (r \vee q)) \end{array}$$

Because these are all propositional tautologies, they are all axioms of K.

Suppose now that $\vdash_{\mathrm{K}} \chi \leftrightarrow \gamma$. By US, MP, and DR2, and the three propositional axioms listed above, it follows that:

$$\begin{array}{l} \vdash_{\mathrm{K}} \neg\chi \leftrightarrow \neg\gamma \\ \vdash_{\mathrm{K}} \chi \vee \beta \leftrightarrow \gamma \vee \beta \\ \vdash_{\mathrm{K}} \beta \vee \chi \leftrightarrow \beta \vee \gamma \\ \vdash_{\mathrm{K}} \Box\chi \leftrightarrow \Box\gamma \end{array}$$

It follows from previous soundness results that all of these are valid as well. By Lemma 11.4.34 just proven, every modal formula is equivalent to a formula built up from atoms with $\neg$, $\vee$, and $\Box$ alone.

Let α be a formula that is constructed from χ by $\neg$, $\vee$, and $\Box$ alone, and β be a formula that is constructed from γ by $\neg$, $\vee$, and $\Box$ alone in exactly the same way. It follows that if $\vDash \chi \leftrightarrow \gamma$, then $\vDash \alpha \leftrightarrow \beta$, by propositional reasoning, and from this that $\vdash_{\mathrm{K}} \alpha$ if and only if $\vdash_{\mathrm{K}} \beta$. Thus, by substituting provably equivalent formulas, validity is preserved. □

Once we have SubsEquiv at our disposal, we add another simplifying convention. SubsEquiv allows, among other substitutions, us to replace $\neg\neg\varphi$ with φ and vice versa (that is,

to apply double negation to any subformula at any time). This can be useful if we wish to apply DeMorgan's laws to formulas that don't quite meet the syntactic requirements—for instance,

$$\neg(\varphi \wedge \psi)$$

is equivalent to

$$\neg\varphi \vee \neg\psi$$

but strictly speaking the one cannot be derived from the other using the equivalences given in Table 7.10 without first converting

$$\neg(\varphi \wedge \psi)$$

to

$$\neg(\neg\neg\varphi \wedge \neg\neg\psi)$$

and then, after DeMorgan's has been applied, remove the doubled negations again. This seems unnecessarily fiddly, and therefore in the presence of SubsEquiv we will often omit applications of double negation.

Because such terms as "DeMorgan's" and "Double Negation" are already familiar and commonly used, we will sometimes simply call an application of SubsEquiv by the name of the equivalency being substituted.

With these conventions and new derived rules in hand, we will state a few more theorems, leaving proofs of some of them for the reader:

Theorem 11.4.36.

$$\vdash_{\mathrm{K}} (\Box p \vee \Box q) \to \Box(p \vee q) \tag{K4}$$

Proof.

1	$p \to (p \vee q)$	PL
2	$q \to (p \vee q)$	PL
3	$\Box p \to \Box(p \vee q)$	DR1, 1
4	$\Box q \to \Box(p \vee q)$	DR1, 2
5	$(\Box p \vee \Box q) \to \Box(p \vee q)$	PL, 3, 4

□

Note that from lines 3 and 4 to 5, we are silently suppressing the statement of a substitution instance of a propositional tautology along with two applications of MP.

Note 11.4.37. The converse of K4 is not a theorem, as demonstrated in Figure 11.6:

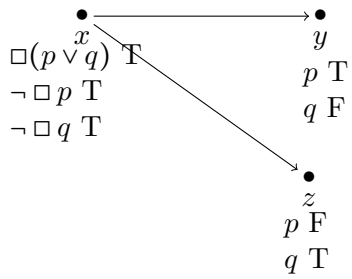


Figure 11.6: $x \not\models \Box(p \vee q) \to (\Box p \vee \Box q)$

Theorem 11.4.38.

$$\vdash_{\mathrm{K}} \Box p \leftrightarrow \neg \Diamond \neg p \tag{K5}$$

Proof.

1	$p \leftrightarrow \neg\neg p$	PL
2	$\Box p \leftrightarrow \neg\neg \Box p$	US, $1[\Box p/p]$
3	$\Box p \leftrightarrow \neg\neg \Box \neg\neg p$	SubsEquiv, $1, 2$
4	$\Box p \leftrightarrow \neg \Diamond \neg p$	def. of $\Diamond, 3$

□

Corollary 11.4.39. $\vdash_{\mathrm{K}} \Diamond p \leftrightarrow \neg \Box \neg p$.

Proof. Immediately from the def. of $\Diamond$ and SubsEquiv. □

Theorem K5 and Corollary 11.4.39 together we will refer to as principles of "modal interchange" (MI), which we will make great use of in conjunction with SubsEquiv. In particular, we will use it to prove another derived rule:

Rule 11.4.40 (DR3)**.** If $\vdash \varphi \to \psi$ then $\vdash \Diamond\varphi \to \Diamond\psi$.

Proof.

1	$\varphi \to \psi$	Given
2	$\neg\psi \to \neg\varphi$	Contraposition, 1
3	$\Box\neg\psi \to \Box\neg\varphi$	DR1, 2
4	$\neg \Box \neg\varphi \to \neg \Box \neg\psi$	Contraposition, 3
5	$\Diamond\varphi \to \Diamond\psi$	SubsEquiv, MI, 4

□

DR3 is the $\Diamond$ analogue of DR1 (Rule 11.4.29). One can also prove a $\Diamond$ analogue of DR2 (Rule 11.4.30); this is left as an exercise to the reader.

Theorem 11.4.41.

$$\vdash_{\mathrm{K}} \Diamond(p \vee q) \leftrightarrow (\Diamond p \vee \Diamond q) \tag{K6}$$

Proof.

1	$\Box(\neg p \wedge \neg q) \leftrightarrow (\Box\neg p \wedge \Box\neg q)$	US, K3$[\neg p/p, \neg q/q]$
2	$\neg\Box(\neg p \wedge \neg q) \leftrightarrow \neg(\Box\neg p \wedge \Box\neg q)$	PL, 1
3	$\neg\Box\neg(p \vee q) \leftrightarrow \neg(\Box\neg p \wedge \Box\neg q)$	DeMorgan's, 2
4	$\neg\Box\neg(p \vee q) \leftrightarrow \neg(\neg\Diamond p \wedge \neg\Diamond q)$	MI, 3
5	$\neg\Box\neg(p \vee q) \leftrightarrow \neg\neg(\Diamond p \vee \Diamond q)$	DeMorgan's, 4
6	$\neg\Box\neg(p \vee q) \leftrightarrow (\Diamond p \vee \Diamond q)$	DN, 5
7	$\Diamond(p \vee q) \leftrightarrow (\Diamond p \vee \Diamond q)$	SubsEquiv, 6

□

Theorem 11.4.42.

$$\vdash_{\mathrm{K}} \Diamond(p \to q) \leftrightarrow (\Box p \to \Diamond q) \tag{K7}$$

Proof.

1	$\Diamond(\neg p \vee q) \leftrightarrow (\Diamond\neg p \vee \Diamond q)$	US, K6$[\neg p/p]$
2	$\Diamond(\neg p \vee q) \leftrightarrow (\neg\Box p \vee \Diamond q)$	MI, 1
3	$\Diamond(p \to q) \leftrightarrow (\Box p \to \Diamond q)$	SubsEquiv, 3

□

K7 is our first theorem that mixes operators, involving both $\Box$ and $\Diamond$.

Theorem 11.4.43.

$$\vdash_{\mathrm{K}} \Diamond(p \wedge q) \to (\Diamond p \wedge \Diamond q) \tag{K8}$$

Proof.

1	$(\Box\neg p \vee \Box\neg q) \to \Box(\neg p \vee \neg q)$	US, K4$[\neg p/p, \neg q/q]$
2	$\neg\Box(\neg p \vee \neg q) \to \neg(\Box\neg p \vee \Box\neg q)$	Contraposition, 1
3	$\Diamond\neg(\neg p \vee \neg q) \to \neg(\Box\neg p \vee \Box\neg q)$	MI, 2
4	$\Diamond(p \wedge q) \to \neg(\Box\neg p \vee \Box\neg q)$	DeMorgan's, 3
5	$\Diamond(p \wedge q) \to (\neg\Box\neg p \wedge \neg\Box\neg q)$	DeMorgan's, 4
6	$\Diamond(p \wedge q) \to (\Diamond p \wedge \Diamond q)$	MI, 5

□

Theorem 11.4.44.

$$\vdash_{\mathrm{K}} \Box(p \vee q) \to (\Box p \vee \Diamond q) \tag{K9}$$

Proof. Left as an exercise for the reader. The proof follows from K using US, Contraposition, and DeMorgan's. □

The attentive reader will note that many (all) of the proofs above crucially involve US. However, it is not always immediately apparent which substitution instances are going to be relevant when in any given proof. We give a method which, while not foolproof, is successful in many instances. First, we define some terminology.

Definition 11.4.45 (PL-transform)**.** The PL-*transform* $\tau(\varphi)$ of a formula φ is the result of deleting all modal operators from φ.

Note that the PL-transform of a propositional wff φ will always be identical to φ. Some examples of formulas and their PL-transforms are given in Table 11.7

Formula	PL-**Transform**
$(p \wedge q) \to r$	$(p \wedge q) \to r$
$\Box(p \to q) \to (\Box p \to \Box q)$	$(p \to q) \to (p \to q)$
$(\Box(p \to q) \wedge \Diamond(p \wedge r)) \to \Diamond(q \wedge r)$	$((p \to q) \wedge (p \wedge r)) \to (q \wedge r)$
$(\Box(p \to q) \wedge \Diamond(p \wedge \Diamond r)) \to \Diamond(q \wedge \Box r)$	$((p \to q) \wedge (p \wedge r)) \to (q \wedge r)$
$\Diamond \Box \Box \Diamond p \to \Box \Diamond \Diamond \Box p$	$p \to p$

Table 11.7: Some formulas and their PL-transforms

When you are asked to prove that a certain wff φ is a theorem, one useful way to start is to identify whether the PL-transform of φ is a tautology. If it is, then assert it as the first line of your proof; then, identify theorems and rules already derived which contain similarly shaped formulas—that is, if you are asked to prove something with an antecedent that is a disjunction, identify which theorems and rules have disjunctive antecedents; if you are asked to prove something with both $\Box$ and $\Diamond$ in the consequent, look to see which theorems and rules have $\Box$ and $\Diamond$ in the consequent.

Alternative axiomatic bases

Taking K to be defined as the logic generated from K plus *modus ponens*, substitution, and N is standard. However, there are alternative axiomatic bases that can be given for K. We

give two alternative axiomatic bases here, and sketch how one would go about proving that these alternative systems are equivalent to K.

Theorem 11.4.46. *Let the axiomatic basis of* K^* *be the same as for* K *except that* N *is replaced by the axiom* $\Box\top : \Box(p \to p)$ *and the rule*

$$\mathsf{R}^* \quad \textit{If } \vdash \varphi \to \psi \textit{ then } \vdash \Box\varphi \to \Box\psi$$

Then K *and* K^* *have the same theorems.*

Theorem 11.4.47. *Let the axiomatic basis of* K^{**} *be the same as for* K *but with* N *and* K *replaced by* $\Box\top$, R^* *and*

$$\mathsf{K2}^* \quad (\Box p \wedge \Box q) \to \Box(p \wedge q)$$

then K *and* K^{**} *have the same theorems.*

Both of these results are proven in the same way. In order for two axiomatic bases to be equivalent, the list of theorems that they generate must be equivalent. We already know that K can already prove $\Box\top$ (since $p \to p$ is a tautology and K constains rule N, $\Box(p \to p)$ is a theorem of K) and that R^* is a derived rule of K (namely, it is $\mathsf{DR1}$ above). Thus, all that must be shown in the proof of Theorem 11.4.46 is that from *modus ponens*, substitution, $\Box\top$, and R^*, one can prove N. Once this is proven, then all that it remains to prove Theorem 11.4.47 is the following:

- That $\mathsf{K2}^*$ is a theorem of K.
- That K is a theorem of $\mathrm{K2}^*$.
- That N is a derived rule of $\mathrm{K2}^*$.

Admissable vs. derivable rules

We will prove (in §11.5.2) that the system K can prove every formula which is valid on the class of all Kripke frames, that is, is K-valid. One consequence of this fact is that K has no theorems of the form $\Diamond\varphi$, because every possibility statement can be falsified in a Kripke frame that has at least one dead-end world. A consequence of *this* fact is that the following rule is a sound rule of K:

Rule 11.4.48. If $\vdash_{\mathrm{K}} \Diamond\varphi$ then $\vdash_{\mathrm{K}} \varphi$.

This rule is sound because the antecedent will never be satisfied.

We introduce this rule here because it allows us to illustrate a distinction between two types of rules that a system can have, beyond the basic rules of US, MP, and N: *admissible* rules and *derivable* rules.

Definition 11.4.49 (Admissible rules)**.** Let S be a system and let $\varphi_1, \ldots, \varphi_n, \varphi$ be some formulas in the language of S. Denote an arbitrary substitution of any of these formulas by, e.g., $\varphi[s]$. A rule of the form:

$$\text{If} \quad \vdash \varphi_1 \wedge \cdots \wedge \varphi_n \quad \text{then} \quad \vdash \varphi$$

is *admissible in* S if for every substitution instance $\varphi_1[s], \ldots \varphi_n[s]$ of $\varphi_1, \ldots, \varphi_n$, $\varphi[s]$ is in S only if $\varphi_1[s], \ldots \varphi_n[s]$ is in S.

Definition 11.4.50 (Derivable rules)**.** Let S be a system and let $\varphi_1, \ldots, \varphi_n, \varphi$ be some wffs in the language of S. A rule of the form:

$$\text{If} \quad \vdash \varphi_1 \wedge \cdots \wedge \varphi_n \quad \text{then} \quad \vdash \varphi$$

is *derivable in* S if there is a proof of φ from the assumptions $\varphi_1 \wedge \cdots \wedge \varphi_n$ in S.

That is, if a rule can be defined without appealing to anything other than steps involving the basic rules defined above, then we say that the rule is a *derived rule*. Derived rules are called this because while they are not primitive transformation rules, their application can always be justified by appeal to a derivation. The addition of derived rules does not change what it is we can prove from an axiomatic basis, they merely serve to simplify our proofs: Any theorem that can be proven with the help of a derived rule can also be proven without appeal to that rule. Admissable rules, on the other hand, capture certain structural properties of logical systems that cannot be expressed via proofs. While the addition derived rules to a logic do not change which theorems can be proven, the addition of admissable rules that are *not* derivable can.

Corollary 11.4.51. *Every derivable rule is admissible.*

DR1, DR2, and DR3 above are all *derived rules*, as indicated in their name and by the fact that we can give a proof of each of the rules. Rule 11.4.48 is admissible, but *not* derivable.

Definition 11.4.52 (Structural completeness)**.** A system S is *structurally complete* if every admissible rule is also derivable.

Corollary 11.4.53. K *is not structurally complete.*

In the next section, we will see an extension of K where Rule 11.4.48 is *not* admissible (and hence is also not derivable). This shows that while every theorem of a logic is a theorem of every extension of the logic, not every rule of a logic will still be an acceptable rule of every extension of the logic; it is only the derivable rules of a system that are guaranteed to still hold in any extension of the logic. This is because when we add new axioms, we never *remove* theorems; we can still prove whatever we could prove before. However, since we can also possibly prove *more*, admissible rules may become inadmissible.

11.4.2 The systems T and D

In §11.3.1, we mentioned a number of modal principles that, depending on the notion of necessity involved, have some intuitive pull. One of these principles is the principle that necessity implies truth (or its contrapositive, that truth implies possibility):

$$\Box p \to p \tag{T}$$

T is sometimes called the "axiom of necessity". If the type of necessity that we are interested in is factive, i.e., something like logical necessity, or epistemic necessity, then this seems to be a plausible principle that, e.g., if something is logically necessary then it is true, or that if something is known then it is true. It *isn't* a plausible principle for all interpretations of modality; for example, belief is not generally factive, for just because something is believed doesn't mean it is true.

T is *not* a theorem of K, which we can easily demonstrate by showing that it is not K-valid. Let $\mathfrak{F}$ be a frame that has a world w which is a dead end (that is, nothing is accessible to it, not even itself). Let $\mathfrak{M}$ be a model that assigns $V(p, w) = F$; then $\Box p$ is true at w but p is not, and hence T is not valid in that model. More precisely, recall the result proven in Lemma 11.3.38 that shows that T is valid on all and only *reflexive* frames. Because K is sound, we know that $\nvdash_{\mathrm{K}} \mathsf{T}$.

This means that if we wish to have a system which includes this principle, we must extend K by adding T as an axiom. We call the resulting system T:

Definition 11.4.54 (System T)**.** System T is system K + T.

T is the characteristic axiom of T. T is an extension of K because everything that was a theorem of K is still a theorem of T. With the addition of the new axiom, however, we can prove new theorems that are not theorems of K (in addition to being able to prove T itself), showing that T is a proper extension of K per Definition 11.4.8.

Some theorems of T that are not theorems of K are the following:

Theorem 11.4.55 (T1)**.** $\vdash_{\mathrm{T}} p \to \Diamond p$.

Proof.

1	$\Box\neg p \to \neg p$	US, T$[\neg p/p]$
2	$p \to \neg\,\Box\,\neg p$	Contraposition, 1
3	$p \to \Diamond p$	MI, 2

Note that in line (2) we have silently surpressed an application of DN. □

Theorem 11.4.56 (T2)**.** $\vdash_{\mathrm{T}} \Diamond(p \to \Box p)$.

Proof.

1	$\Box p \to \Diamond\,\Box\,p$	US, T1$[\Box p/p]$
2	$\Diamond(p \to \Box p) \leftrightarrow (\Box p \to \Diamond\,\Box\,p)$	US, K7$[\Box p/q]$
3	$\Diamond(p \to \Box p)$	MP, 1, 2

□

Lemma 11.3.38 shows that the logic T is *sound* with respect to the class of all reflexive frames. In §11.5.2, we will also show that it is *complete* with respect to this class.

Because T2 is a theorem of T, we can show that Rule 11.4.48 is not admissible in T. Suppose that it were. Then:

1	$\Diamond(p \to \Box p)$	T2
2	$p \to \Box p$	Rule 11.4.48
3	$\Box p \to p$	T
4	$\Box p \leftrightarrow p$	PL, 2, 3

But this principle, $\Box p \leftrightarrow p$, is not provable in T, because we can define a reflexive model in which it is false (see Figure 11.8).

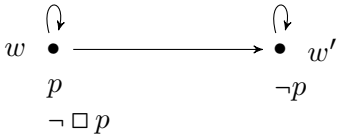


Figure 11.8: $\Box p \leftrightarrow p$ is false at w

Given that $\vdash_{\mathrm{T}} \Box p \to p$, we also have, by US, such theorems as $\vdash_{\mathrm{T}} \Box\,\Diamond\,p \to \Diamond p$ and $\vdash_{\mathrm{T}} \Box\,\Box\,\Diamond\,\Diamond\,\Box p \to \Box\,\Diamond\,\Diamond\,\Box\,p$. What about their converses? Are they theorems of T? If

$$\mathrm{K} \subset \mathrm{D} \subset \mathrm{T}$$

Figure 11.9: Inclusion relations of systems so far

they are not, are they plausible modal principles? As it turns out, the answer to the former question is 'no' and the answer to the latter is 'it depends'; we discuss this further in §11.4.3.

The principle that necessity implies truth is quite a strong principle, appropriate for only certain types of necessity—logical necessity, for example, but also epistemic necessity (since knowledge is typically taken to imply truth). It would not, however, be appropriate for doxastic necessity, as belief does not always imply truth, nor would it be appropriate for deontic necessity: For while "ought implies can", it is not the case that "ought implies is". The latter slogan would be expressed by the T axiom, which indicates that in order to define a logic appropriate for reasoning about deontic necessity, we need a weaker principle. The principle that "ought implies can" can be put in somewhat more modal terms: If it is necessary that one do p, then it is at least possible that one do p, that is:

$$\Box p \to \Diamond p \tag{D}$$

Approaching deontic operators in a modal fashion in this way was first explored in modern times in [von Wright, 1951].

If we add the wff $\Box\varphi \to \Diamond\varphi$ to K, we obtain the system D:

Definition 11.4.57 (System D)**.** System D is $\mathrm{K} + \mathsf{D}$.

Because validity of axiom D corresponds to the frame property seriality (cf. Lemma 11.3.40, D is characterized by the class of serial frames.

D is, like T, a proper extension of K. This is because D can be falsified on any frame that is not serial, and hence given that K is sound, D is not a theorem of K. Because reflexivity implies seriality, but not the other way around, D is *not* an extension of T, but rather the other way around: T is an extension of D because every theorem of D is already a theorem of T. The inclusion relations for the systems we've investigated so far are given in Figure 11.9.

Instead of taking D as our characteristic axiom, we could have taken another wff,

$$\Diamond(p \to p) \tag{D1}$$

This is because from D1 we can prove D:

1	$\Diamond(p \to p)$	D1
2	$\Diamond(p \to p) \leftrightarrow (\Box p \to \Diamond p)$	US, K7$[p/q]$
3	$\Box p \to \Diamond p$	MP, 1, 2

We can also prove D1 from D:

Theorem 11.4.58 (D1)**.** $\vdash_{\mathrm{D}} \Diamond(p \to p)$.

Proof.

1	$p \to p$	PL
2	$\Box(p \to p)$	N, 1
3	$\Box(p \to p) \to \Diamond(p \to p)$	US, D$[p \to p/p]$
4	$\Diamond(p \to p)$	MP, 2, 3

□

Thus, D and D1 are provably equivalent, and the systems which have these axioms as their bases are deductively equivalent. We can also show, from D1, that any system which has as a theorem a wff of the form $\Diamond\varphi$ *must* be either D or an extension of D:

Lemma 11.4.59. *If* $\vdash_{\mathrm{S}} \Diamond\varphi$ *for some system* S *and wff* φ*, then* $\mathrm{D} \subseteq \mathrm{S}$.

Proof.

1	$\Diamond\varphi$	Given
2	$q \to (p \to p)$	PL
3	$\varphi \to (p \to p)$	US$[\varphi/q]$, 2
4	$\Diamond\varphi \to \Diamond(p \to p)$	DR3, 3
5	$\Diamond(p \to p)$	MP, 1, 4

□

The following result, concerning validity in the class $\mathfrak{C}^{\mathrm{D}}$ of serial frames, also involves possibility statements:

Lemma 11.4.60. *If* $\mathfrak{C}^{\mathrm{D}} \vDash \Diamond\varphi$ *then* $\mathfrak{C}^{\mathrm{D}} \vDash \varphi$.

Proof. Suppose that $\mathfrak{C}^{\mathrm{D}} \vDash \Diamond\varphi$ but $\mathfrak{C}^{\mathrm{D}} \nvDash \varphi$. Then there is a serial model $\mathfrak{M} = \langle W, R, V\rangle$ with $w \in W$ such that $\mathfrak{M}, w \nvDash \varphi$. Construct a new model $\mathfrak{M}' = \langle W', R', V'\rangle$ such that $W' = W \cup \{x\}$, $R' = R \cup \langle x, w\rangle$, and $V' = V$. (That is, take the original model, and add a new world that only sees the world where φ is false, and nothing else, and where nothing is true at that world.) Note that the new model is also serial; we've added one world, and

it can see at least one other world. Since xRw and nothing else, $\mathfrak{M}, x \vDash \Box\neg\varphi$. But then $\mathfrak{M}, x \vDash \neg \Diamond \varphi$, and this contradicts the assumption that $\Diamond\varphi$ is D-valid. $\square$

Finally, we prove a result concerning D and hence any system extending it, that will be useful later on, concerning a relationship between valid PL-transforms (cf. Definition 11.4.45) of modal wffs and provability in D. First, recall Definition 11.2.3, of a *modal constant wff* (that is, a wff which is built up from $\top$, $\bot$, and the logical operators, including the modal operators, alone).

Lemma 11.4.61. *Let φ be a modal constant wff. Then:*

- *If $\vDash \tau(\varphi)$, then $\vdash_{\mathrm{D}} \varphi$;*
- *If $\nvDash \tau(\varphi)$, then $\vdash_{\mathrm{D}} \neg\varphi$.*

That is, if the PL-transform of a modal constant wff is valid, then it will be provable in D*, and any system extending* D*, and if it is invalid, then its negation will be provable in* D*, and any system extending* D*.*

Proof. The proof is by induction on the complexity of φ. First, note that every constant wff is equivalent to one built up from $\bot$ by $\neg$, $\vee$, and $\Box$. Thus it is sufficient to prove the Lemma to prove the following:

1. it holds for $\bot$;
2. if it holds for φ then it holds for $\neg\varphi$;
3. if it holds for φ and ψ then it holds for $\varphi \vee \psi$;
4. if it holds for φ then it holds for $\Box\varphi$.

We prove the atomic case and the modal case in turn, and leave the remaining propositional cases as exercises for the reader.

(1) $\tau(\bot) = \bot$, and since $\neg \bot = \top$, it follows that $\nvDash \bot$. But since $\top$ is a tautology, it is an axiom of D, so $\vdash_{\mathrm{D}} \top$, that is, $\vdash_{\mathrm{D}} \neg \bot$.

(4) Assume (1) the result holds for φ, and assume further (2) that $\vDash \tau(\Box\varphi)$. By definition, $\tau(\Box\varphi) = \tau(\varphi)$, hence $\vDash \tau(\varphi)$. By assumption (1), $\vdash_{\mathrm{D}} \varphi$. By N, $\vdash_{\mathrm{D}} \Box\varphi$, which is what we needed to show.

Next, assume (3) that $\not\vdash \tau(\Box\varphi)$. By definition, $\tau(\Box\varphi) = \tau(\varphi)$, hence $\not\vdash \tau(\varphi)$. By assumption (1), $\vdash_{\mathrm{D}} \neg\varphi$. We can then derive the following:

1	$\neg\varphi$	Assumption
2	$\Box\neg\varphi$	N, 1
3	$\Box\neg\varphi \rightarrow \Diamond\neg\varphi$	US, D$[\neg\varphi/p]$
4	$\Diamond\neg\varphi$	MP, 2, 3
5	$\neg\,\Box\,\varphi$	MI, 4

Which is what we required.

(2), (3) Left as exercises for the reader.

□

Note the crucial application of D in step (4); without this axiom, this result would not be provable.

11.4.3 The systems S4, S5, and B

In this section, we introduce three related systems which all developed out of similar motivations having to do with how we interpret embedded and iterated modalities. Let's begin with some definitions and preliminary results:

Definition 11.4.62 (Modality)**.** A *modality* is any unbroken sequence of zero or more monadic operators (e.g., $\neg$, $\Diamond$, and $\Box$).

We indicate the zero case by $-$.

Definition 11.4.63 (Iterated modality)**.** An *iterated modality* is any modality that has at least two modal operators (e.g., $\Box$ and $\Diamond$).

Definition 11.4.64 (Standard form)**.** A modality with at most one negation is said to be in *standard form.*

Lemma 11.4.65. *In any system which validates MI, every modality is equivalent to a modality that contains at most one negation.*

Proof. Any modality not in standard form can be reduced to one in standard form through repeated applications of MI and Double Negation. □

Example 11.4.66.

$$\Box\neg\Diamond\Box\neg\Box\neg\Diamond\Diamond\neg p$$

is equivalent to

$$\Box\neg\Diamond\Box\Diamond\Diamond\Diamond\neg p$$

is equivalent to

$$\Box\Box\Diamond\Box\Box\Box\neg\neg p$$

is equivalent to

$$\Box\Box\Diamond\Box\Box\Box p$$

Definition 11.4.67. A modality is *affirmative* or *positive* if its standard form contains no negation. A modality is *negative* if its standard form contains one negation.

Example 11.4.68. The modality in Example 11.4.66 is positive.

Some simple examples of formulas involving iterated modalities are given in Figure 11.10, along with their natural readings in English. But iterated modalities can often be quite a bit more complex than the ones given in Figure 11.10, and they can be hard for us to interpret in natural language, much less have any intuitions about their truth or falsity under a given interpretation of the modal operators.

$\Box\varphi\rightarrow\Box\Box\varphi$	"If φ is necessary, then it is necessarily necessary".
$\Box\Box\varphi\rightarrow\Box\varphi$	"If it is necessary that φ is necessary, then φ is necessary".
$\Diamond\varphi\rightarrow\Box\Diamond\varphi$	"If φ is possible, then it's necessarily possible".
$\Diamond\Diamond\varphi\rightarrow\Diamond\varphi$	"If φ is possibly possible, then φ is possible".
$\Diamond\Box\varphi\rightarrow\Box\varphi$	"If φ is possibly necessary, then φ is necessary".

Figure 11.10: Some examples of wffs involving iterated modalities

One way to deal with this is to investigate classes of frames which validate so-called *reduction principles*, that is, principles that allow us to convert complex iterated modalities into simpler, yet still equivalent, ones.

Definition 11.4.69. Two modalities A and B are *equivalent* in a system S iff $\vdash_{\mathrm{S}} Ap\leftrightarrow Bp$.

Definition 11.4.70. If A and B are equivalent modalities and A has fewer modal operators and negations than B, then B is *reducible* to A.

If $\vdash_{\mathrm{S}} Ap \leftrightarrow Bp$ and B is reducible to A, then we say that $Ap \leftrightarrow Bp$ is a *modal reduction principle* in system S.

The logic T, with T as its characteristic axiom, does not validate any reduction principles other than the modal interchange principles (e.g., $\Box p \leftrightarrow \neg\Diamond\neg p$ and the like) [Hughes and Cresswell, 1996, pp. 52, 70]. It does, however, validate the following implications:

$$\begin{array}{ll} (1) & \Box \Diamond p \rightarrow \Diamond p \\ (2) & \Box p \rightarrow \Diamond \Box p \\ (3) & \Diamond p \rightarrow \Diamond \Diamond p \\ (4) & \Box \Box p \rightarrow \Box p \end{array}$$

(1) and (4) are obtained from T via US and (2) and (3) are obtained from $\mathsf{T1}$ by US. *If* the converses of any of these were also valid, we would have a modal reduction principle. However, none of the converses of (1)–(4) are valid in T, which is easily shown by the production of a reflexive countermodel where they are falsified.

Thus, we will want to look at stronger logics which do include reduction principles. These stronger logics can be obtained by adding the converse of one or more of (1)–(4) above (we'll indicate these converses by (1′)–(4′)):

$$\begin{array}{ll} (1') & \Diamond p \rightarrow \Box \Diamond p \\ (2') & \Diamond \Box p \rightarrow \Box p \\ (3') & \Diamond \Diamond p \rightarrow \Diamond p \\ (4') & \Box p \rightarrow \Box \Box p \end{array}$$

First, note that in K (and hence in T), (1′) and (2′) are provably equivalent; they can each be obtained from the other by contraposition and the substitution of $\neg p$ for p. Therefore, if we add one of (1′) or (2′) to T, we will immediately have the other one.

Second, note also that in T, (3′) and (4′) are provably equivalent, in a similar fashion. So similarly, if we add one of (3′) or (4′) to the other, then we will immediately have the other one.

Finally, we will show below (Theorem 11.4.81) that in T, 1′ implies 4′, but not vice versa.

Thus, there are two distinct ways that we could add reduction principles to T: By adding (1′) $\Diamond p \rightarrow \Box \Diamond p$ or by adding (4′) $\Box p \rightarrow \Box \Box p$. (If we added (2′), the result would be deductively equivalent to adding (1′); and similarly for (3′) and (4′)). If we add (1′) to T, the resulting system is known as S5; if we add (4′) to T, the resulting system is known as S4.[8] That is:

[8]The names S4 and S5 were given by Lewis and Langford because they were the fourth and fifth systems of modal logic to be considered in [Lewis and Langford, 1932a, p. 501].

Definition 11.4.71. The logic S4 is the logic T + the axiom $\mathsf{4} := \Box p \to \Box\Box p$.

It follows from Lemma 11.3.42 that S4 is the set of all formulas valid on every frame which is both reflexive and transitive.

Definition 11.4.72. The logic S5 is the logic T + the axiom $\mathsf{E} := \Diamond p \to \Box \Diamond p$.

It follows from Lemma 11.3.45 that S5 is the set of all formulas valid on every frame which is both reflexive and Euclidean. Below, we will see an alternative characterisation of S5. First, we look at the reduction laws that these two logics can prove.

Iterated modalities in S4

S4 validates four reduction principles, which are given in Figure 11.11. We prove two of them; the other two are straightforward corollaries using substitution and contraposition.

$$\begin{array}{llll} (1) & \Box p \leftrightarrow \Box\Box p & (3) & \Box\Diamond p \leftrightarrow \Box\Diamond\Box\Diamond p \\ (2) & \Diamond p \leftrightarrow \Diamond\Diamond p & (4) & \Diamond\Box p \leftrightarrow \Diamond\Box\Diamond\Box p \end{array}$$

Figure 11.11: Reduction laws in S4

Theorem 11.4.73 ($\mathsf{S4}(1)$)**.** $\vdash_{\mathrm{S4}} \Box p \leftrightarrow \Box\Box p$.

Proof.

1	$\Box\Box p \to \Box p$	$\mathsf{US}, \mathsf{T}[\Box p/p]$
2	$\Box p \to \Box\Box p$	$\mathsf{4}$
3	$\Box p \leftrightarrow \Box\Box p$	$\mathrm{PL}, 1, 2$

□

Corollary 11.4.74 ($\mathsf{S4}(2)$)**.** $\vdash_{\mathrm{S4}} \Diamond p \leftrightarrow \Diamond\Diamond p$.

Proof. Immediate from the proof of $\mathsf{S4}(1)$ by substitution and contraposition. □

Theorem 11.4.75 ($\mathsf{S4}(3)$)**.** $\vdash_{\mathrm{S4}} \Box\Diamond p \leftrightarrow \Box\Diamond\Box\Diamond p$.

Proof.

1	$\Box \Diamond p \rightarrow \Diamond \Box \Diamond p$	$\mathsf{US}, \mathsf{T1}[\Box \Diamond p/p]$
2	$\Box \Box \Diamond p \rightarrow \Box \Diamond \Box \Diamond p$	$\mathsf{DR1}, 1$
3	$\Box \Diamond p \rightarrow \Box \Diamond \Box \Diamond p$	$\mathsf{SubsEquiv}, \mathsf{S4(1)}, 2$
4	$\Box \Diamond p \rightarrow \Diamond p$	$\mathsf{US}, \mathsf{T}[\Diamond p/p]$
5	$\Diamond \Box \Diamond p \rightarrow \Diamond \Diamond p$	$\mathsf{DR3}, 4$
6	$\Diamond \Box \Diamond p \rightarrow \Diamond p$	$\mathsf{S4(2)}, 5$
7	$\Box \Diamond \Box \Diamond p \rightarrow \Box \Diamond p$	$\mathsf{DR1}, 6$
8	$\Box \Diamond p \leftrightarrow \Box \Diamond \Box \Diamond p$	$\mathrm{PL}, 3, 7$

□

Note that in the above proof, we've slightly suppressed the substitution of S4(1) at line 3.

Corollary 11.4.76 (S4(4))**.** $\vdash_{\mathrm{S4}} \Diamond \Box p \leftrightarrow \Diamond \Box \Diamond \Box p$.

Proof. Immediate from the proof of S4(3) by substitution and contraposition. □

These reduction principles allow us to reduce any two identical modalities to just one copy; and also to reduce any double repetition of alternating modalities with just a single instance of the pair of alternating modalities. Thus, even though the reduction principles only involve modalities of length 4 or shorter, we can take modalities of arbitrary lengths and reduce them in S4.

Example 11.4.77. In S4, the following modality:

$$\Box \Diamond \Box \Box \Diamond \Box \Diamond \Box \Diamond \Diamond$$

is equivalent to the following:

$\Box \Diamond \Box \Box \Diamond \Box \Diamond \Box \Diamond$	S4(2)
$\Box \Diamond \Box \Diamond \Box \Diamond \Box \Diamond$	S4(1)
$\Box \Diamond \Box \Diamond \Box \Diamond$	S4(3)
$\Box \Diamond \Box \Diamond$	S4(3)
$\Box \Diamond$	S4(3)

The final entry in the list above, $\Box\Diamond$, cannot be reduced any further.

This raises the following question: How many distinct (that is, non-equivalent) modalities are there in S4? In T, there were infinitely many—any string of modalities is non-equivalent to any other distinct string in T—because T has no reduction laws. Given the modal reduction laws in Figure 11.11, the situation is very different in S4. The following lemma makes precise the situation in S4.

Lemma 11.4.78. *There are 14 distinct modalities in* S4 *that are not equivalent to each other, the following seven and their negations:*

$$(i)-,\ (ii)\Box,\ (iii)\Diamond,\ (iv)\,\Box\,\Diamond,\ (v)\,\Diamond\,\Box,\ (vi)\ \Box\,\Diamond\Box,\ (vii)\,\Diamond\,\Box\Diamond$$

Proof. It is easy to demonstrate that these are all distinct (i.e., that none can be reduced to another): This can be done by giving a reflexive and transitive model that makes one true and another false for every pair of modalities in the list. (Because this would be $2^{14} = 16,384$ cases to consider, we leave this as an exercise for the reader.)

We also need to show that these are the *only* distinct modalities, i.e., that any other modality is either identical to or reducible to one of these fourteen. We consider just the seven positive modalities, as the case of the negative modalities is parallel: Reduce each of the negative cases to their standard form, and then substitute in $\neg p$ for p and perform double negation.

Consider the zero case, (i). If we prefix $\Box$ to (i), then the result is (ii); if we prefix $\Diamond$ to (i), then the result is (iii).

Consider the single-operator modalities, (ii) and (iii). If we prefix (ii) with $\Box$, the result is requivalent to (ii) by $\mathsf{S4(1)}$. If we prefix (ii) with $\Diamond$, the result is (v). If we prefix (iii) with $\Box$, the result is (iv). If we prefix (iii) with $\Diamond$, it is equivalent to (iii) by $\mathsf{S4(2)}$.

Consider the two operator modalities, (iv) and (v). If we prefix (iv) with $\Diamond$, the result is (vii). If we prefix (iv) with $\Box$, the result is equivalent to (iv) by $\mathsf{S4(1)}$. If we prefix (v) with $\Diamond$, the result is equivalent to (v) by $\mathsf{S4(2)}$. If we prefix (v) with $\Box$, the result is (vi).

Consider the three operator modalities, (vi) and (vii). If we prefix (vi) with $\Diamond$, the result is equivalent to (iv) by $\mathsf{S4(3)}$. If we prefix (vi) with $\Box$, the result is equivalent to (vi) by $\mathsf{S4(1)}$. If we prefix (vii) with $\Diamond$, the result is equivalent to (vii) by $\mathsf{S4(2)}$. If we prefix (vii) with $\Box$, the result is equivalent to (v) by $\mathsf{S4(4)}$; as a result, there are no irreducible four-modality affirmative modalities.

Any modality of length 5 or greater will contain within it some 2- or 4-length modality to which the reduction laws can be applied to reduce it (perhaps successively) to a 4-length modality or shorter. □

The implication relations between the seven positive modalities is given in Figure 11.12 (cf. [Prior, 1957, p. 124]). The case for the negative modalities is proven similarly, and a similar diagram can be given for the implication relations between them, except that φ is replaced with $\neg\varphi$ and the direction of the arrows is changed.

There is more that can be said here about S4, specifically concerning its applications in philosophy, mathematics, and computer science. I will write more when I have time.

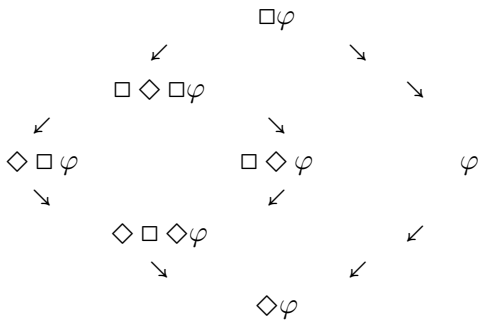


Figure 11.12: Implication relations of positive modalities in S4.

Iterated modalities in S5

In S4, we took the weaker of the two reduction principles not provable in T and added it to T; in S5, we take the stronger of the two, the axiom $\mathsf{E} \coloneqq \Diamond p \to \Box \Diamond p$.

In S5, we can prove four reduction laws (see Figure 11.13). Two of them were theorems of S4. The other two, involving mixed modalities, we prove, and then we show that the reduction laws from S4 are provable in S5 as well.

Theorem 11.4.79 (S5(1))**.** $\vdash_{\mathrm{S5}} \Diamond p \leftrightarrow \Box \Diamond p$.

Proof.

1	$\Box \Diamond p \to \Diamond p$	$\mathsf{US}, \mathsf{T}[\Diamond p/p]$
2	$\Diamond p \to \Box \Diamond p$	E
3	$\Box \Diamond p \leftrightarrow \Diamond p$	$\mathrm{PL}, 1, 2$

□

Corollary 11.4.80 (S5(2))**.** $\vdash_{\mathrm{S5}} \Box p \leftrightarrow \Diamond \Box p$.

Proof. Immediate from the previous result by substitution and contraposition. □

Note that the method by which this and earlier corollaries can be proven is a general method based on purely proposition methods and MI: For any theorem whose main connective is a conditional, one can reverse the direction of the arrow while exchanging □s for ◇s and vice versa. When the theorem is a biconditional, one must reverse both arrows, which cancels out; thus, one can simply exchange the □s for ◇s, leaving the rest of the formula unchanged.

Though as presented, we've built S5 directly onto T, bypassing S4, S5 is actually an extension of S4, which we show by showing that in S5 the axiom $\mathsf{4} := \Box p \rightarrow \Box\Box p$ can be proven:

Theorem 11.4.81. $\vdash_{\mathrm{S5}} \Box p \rightarrow \Box\Box p$.

Proof.

1	$\Box p \rightarrow \Diamond \Box p$	$\mathsf{US}, \mathsf{T1}[\Box p/p]$
2	$\Diamond \Box p \leftrightarrow \Box \Diamond \Box p$	US, Theorem 11.4.79$[\Box p/p]$
3	$\Diamond \Box p \rightarrow \Box \Diamond \Box p$	PL, 2
4	$\Diamond \Box p \rightarrow \Box\Box p$	$\mathsf{SubsEquiv}, \mathsf{S5(2)}, 3$
5	$\Box p \rightarrow \Box\Box p$	$\mathsf{Trans}, 1, 4$

□

From this it follows that the proofs of the reduction laws in S4 also hold in S5, giving us four reduction laws that allow us to reduce two modalities to one.

(1)	$\Box p \leftrightarrow \Box\Box p$	(3)	$\Diamond p \leftrightarrow \Box \Diamond p$
(2)	$\Diamond p \leftrightarrow \Diamond \Diamond p$	(4)	$\Box p \leftrightarrow \Diamond \Box p$

Figure 11.13: Reduction laws in S5

(1) and (2) are already present in S4; (3) and (4) are the new additions specific to S5. Specifically, the laws listed in Figure 11.13 state that in any pair of adjacent modal operators, you can always delete the left-most one. This gives us the following result concerning the number of distinct modalities in S5:

Lemma 11.4.82. *There are six non-equivalent modalities in* S5*:*

$$(i)-,\ (ii)\Box,\ (iii)\Diamond$$

and their negations,

$$(iv)\neg,\ (v)\neg\Box,\ (vi)\neg\Diamond$$

Proof. Left as an exercise for the reader; it is similar to the proof of Lemma 11.4.78. □

The ability to reduce complex iterated modalities to a single modal operator (with or without a negation) is one of the attractive features of S5, and it is one of the reasons why S5 is taken to be the logic which characterizes our notion of logical necessity. With these reduction rules, we can prove a number of useful theorems in S5, including the following distribution principles:

Theorem 11.4.83 (S5(3))**.** $\vdash_{\mathrm{S5}} \Box(p \vee \Box q) \leftrightarrow (\Box p \vee \Box q)$.

Proof.

1	$\Box(p \vee \Box q) \rightarrow (\Box p \vee \Diamond \Box q)$	US, K9$[\Box q/q]$
2	$\Box(p \vee \Box q) \rightarrow (\Box p \vee \Box q)$	Reduction Law $4, 1$
3	$(\Box p \vee \Box \Box q) \rightarrow \Box(p \vee \Box q)$	US, K4$[\Box q/q]$
4	$(\Box p \vee \Box q) \rightarrow \Box(p \vee \Box q)$	Reduction Law $1, 3$
5	$(\Box p \vee \Box q) \leftrightarrow \Box(p \vee \Box q)$	PL$2, 4$

□

Theorem 11.4.84 (S5(4))**.** $\vdash_{\mathrm{S5}} \Box(p \vee \Diamond q) \leftrightarrow (\Box p \vee \Diamond q)$.

Proof. Proven from S5(3) via US. □

Theorem 11.4.85 (S5(5))**.** $\vdash_{\mathrm{S5}} \Diamond(p \wedge \Diamond q) \leftrightarrow (\Diamond p \wedge \Diamond q)$.

Proof. Proven from S5(3) via contraposition, MI, and US. □

Theorem 11.4.86 (S5(6))**.** $\vdash_{\mathrm{S5}} \Diamond(p \wedge \Box q) \leftrightarrow (\Diamond p \wedge \Box q)$.

Proof. Proven from S5(5) via US. □

We now give an alternative axiomatic basis of S5, before discussing which class of frames S5 corresponds to. Let $\mathsf{B} \coloneqq p \rightarrow \Box \Diamond p$. (Axiom B takes its name from the philosopher and mathematician L.E.J. Brouwer, and sometimes called the 'Brouwer axiom', or the 'Brouwerian axiom' or the 'Brouwersche axiom'.[9])

Theorem 11.4.87. *The logic* S5 *is also axiomatised by* $\mathsf{K} + \mathsf{T} + 4 + \mathsf{B}$.

Proof. To show this, we must show that from $\mathsf{K}+\mathsf{T}+4+\mathsf{B}$ we can prove E, and from $\mathsf{K}+\mathsf{T}+\mathsf{E}$ we can prove B (we have already shown that we can prove 4). □

Per Lemma 11.3.43, B is valid on a class of frames iff the frame is symmetric. Putting this condition together with the frame conditions induced by the other two axioms, we have the following corollary:

Corollary 11.4.88. *The logic* S5 *is characterised by the class of frames which is reflexive, transitive, and symmetric.*

[9] Brouwer was Dutch.

Recall from Definition 11.3.24 that relations which are reflexive, transitive, and symmetric are called *equivalence relations*. Thus we can also say that S5 is characterised by the class of frames where R is an equivalence relation. Note that it does not follow from this that S5 is characterised by the class of frames where R is a *universal* relation (cf. Definition 11.3.25), for the frame in Figure 11.14 is a frame for S5 but does not have a universal relation, since the xs are not related to the ws.

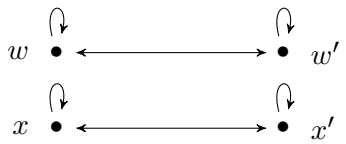


Figure 11.14: An S5 frame where R is not universal

S5 is one of the most fruitful and productive modal logics with a wide range of applications in philosophy, mathematics, and computer science. I will say more about these when I have time.

System B

Definition 11.4.89. The logic B is axiomatised by $\mathsf{K} + \mathsf{B}$.

By Lemma 11.3.43, B is characterised by the class of frames where R is symmetric.

Useful theorems of B include the following (all of which we present without proof but encourage the reader to prove for themself):

Lemma 11.4.90 (B1)**.** $\vdash_{\mathrm{B}} \Diamond \Box p \to p$.

Lemma 11.4.91 (B2)**.** $\vdash_{\mathrm{B}} (\Diamond \Box p \wedge \Diamond \Box q) \to \Box \Diamond (p \wedge q)$.

Lemma 11.4.92 (B3)**.** $\vdash_{\mathrm{B}} \Diamond \Box p \to \Box \Diamond p$.

Any system containing B has the following as a derived rule:

Rule 11.4.93 (DR4)**.** If $\vdash \Diamond\varphi \to \psi$ then $\vdash \varphi \to \Box\psi$.

Proof.

1	$\Diamond\varphi \to \psi$	Given
2	$\Box \Diamond \varphi \to \Box\psi$	DR1
3	$\varphi \to \Box \Diamond \varphi$	B$[\varphi/p]$
4	$\varphi \to \Box\psi$	Trans, 2, 3

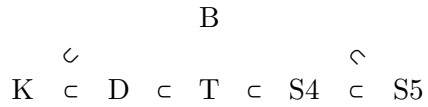


Figure 11.15: Inclusion relations of systems so far

□

Since S5 contains B, DR4 is a derived rule in S5 as well. We can also use DR4 to generate S5 from S4:

Lemma 11.4.94. S4 + DR4 = S5.

Proof.

1	$\Diamond\Diamond p \rightarrow \Diamond p$	S4(2)
2	$\Diamond p \rightarrow \Box\Diamond p$	DR4, 1

□

Other modal systems

It is not difficult to verify that each of T, 4, and B are independent of each other (for there are reflexive frames which are neither transitive nor symmetric; transitive frames which are neither reflexive nor symmetric; and symmetric frames which are neither reflexive nor transitive). Thus, we can also add either axiom 4 or axiom B to K, and obtain distinct logics, K4 and KB, just as we added T to K to obtain T (we can also add multiple axioms to get, e.g., KTB, etc.).

In Figure 11.15 we give the inclusion relation for the primary logics introduced in this section.

11.4.4 The systems Triv and Ver; Spinozist metaphysics and modal logic

In the previous section, we identified a number of reduction laws that could be obtained by adding to T the converses of certain theorems of T. What we did not consider was the case of adding the converse of T itself, namely:

$$p \rightarrow \Box p \qquad \text{(Converse-T)}$$

In this section, we explore what happens if we add wffs like this one as an axiom to one of our already defined systems. The result will be two extremely strong modal systems, Triv

(for 'trivial') and Ver (for '*verum*'.) The first system trivialises modality; the second one verifies all modality. While at first it might appear that neither system has interest beyond the formal/technical, at the end of this section we will briefly motivate a context in which Triv might actually be the desired modal logic.

To return to Converse-T, there is certainly good intuitive reason for not adding this as an axiom, to T or any other system: For it says that the truth implies necessity, which seems to get the order of things all wrong. But this is a semantic concern; nothing prevents us, on the axiomatic side of things, from adding Converse-T to any axiomatic base. Furthermore, we know that the result of adding Converse-T to any of the modal systems we've studied so far will result in a new system. This is because Converse-T isn't even a theorem of S5. We can show that it is not a theorem of S5 by giving an S5-countermodel (in Figure 11.16); if it is not a theorem of S5, it is not a theorem of any logic contained in S5, and so far, we have only considered logics that are contained in S5.

Figure 11.16: An S5 model falsifying $p \to \Box p$

In this section, we will look at a variety of systems that can be obtained from adding super-strong modal axioms such as Converse-T to different axiomatic bases. First we consider what happens if we add Converse-T to some system that is at least as strong as D. If we do, then we can prove the following:

$$p \leftrightarrow \Box p \qquad \text{(Triv)}$$

Proof.

1	$p \to \Box p$	Converse-T
2	$\Box p \to \Diamond p$	D
3	$p \to \Diamond p$	$1, 2,$ Trans
4	$\neg p \to \Diamond \neg p$	US$[\neg p/p], 3$
5	$\neg p \to \neg \Box p$	MI$, 4$
6	$\Box p \to p$	Contraposition$, 5$
7	$\Box p \leftrightarrow p$	PL$, 1, 6$

□

This means that adding the converse of T to any system at least as strong as D will get you the theorem Triv. We call this wff Triv because it *trivialises* the concept of modality,

making it entirely reducible to truth. There is no longer any distinction between what is necessary and what is (merely) true; one can always shift from one to the other.

But though Triv trivialises the concept of modality, it doesn't make it inconsistent. That is to say, we can add Triv to a system—even to S5—and the resulting system will still be consistent (in the sense of 'consistent' defined in Definition 11.4.9). This is because we can give an S5 model (in Figure 11.17) where Triv is true at w. So even though $p \to \Box p$ can be falsified on an S5 frame, adding it to S5 (or to any of the weaker systems contained in S5) would not result in an inconsistent system.

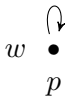


Figure 11.17: An S5 model that makes Triv true at w

The first of the two systems that we discuss in this section is the system Triv containing Triv. This system has two natural axiomatic bases:

Definition 11.4.95 (Triv)**.** The logic Triv is the logic K + Triv := $p \leftrightarrow \Box p$ or D+Converse-T := $p \to \Box p$.

It is straightforward to see that these two axiomatic bases are equivalent. First, we showed above how adding Converse-T to any system containing D allows us to prove Triv. In order to show the other direction of the equivalence, we must show that from K + Triv, we can prove both D and Converse-T:

Lemma 11.4.96. *(1)* K + Triv $\vdash \Box p \to \Diamond p$ *and (2)* K + Triv $\vdash p \to \Box p$.

Proof.

1.

1	$p \leftrightarrow \Box p$	Triv
2	$\neg p \to \neg \Box p$	PL, 1
3	$\neg p \to \Diamond \neg p$	MI, 2
4	$p \to \Diamond p$	US$[\neg p/p]$, DN, 3
5	$\Box p \to p$	PL, 1
6	$\Box p \to \Diamond p$	Trans, 4, 5

2.

1	$p \leftrightarrow \Box p$	Triv
2	$p \to \Box p$	PL, 1

□

As we pointed out above, while Triv may be a trivial modal logic, it is not an inconsistent one, since $p \leftrightarrow \Box p$ is satisfiable. What is more, there is a class of models that characterises Triv:

Lemma 11.4.97. *System* Triv *corresponds to the class* $\mathfrak{C}$ *of frames where every world can see itself, and nothing more. That is,* $\mathfrak{F} \in \mathfrak{C}$ *iff* $\mathfrak{F} \vDash p \leftrightarrow \Box p$.

Proof.

($\Rightarrow$) Suppose $\mathfrak{F} \in \mathfrak{C}$, but that $\mathfrak{F} \nvDash p \leftrightarrow \Box p$. Then there is a model $\mathfrak{M}$ on $\mathfrak{F}$ and a world w where $\mathfrak{M}, w \nvDash p \leftrightarrow \Box p$. That is, either (1) $\mathfrak{M}, w \vDash p \wedge \neg \Box p$ or (2) $\mathfrak{M}, w \vDash \neg p \wedge \Box p$. Consider (1); it follows that $\mathfrak{M}, w \vDash \Diamond \neg p$, that is, there is a w' s.t. wRw' and $\mathfrak{M}, w' \vDash \neg p$. But the only world that w sees is w, and $\mathfrak{M}, w \vDash p$, which is a contradiction. Case (2) is argued analogously.

($\Leftarrow$) There are two cases. First, suppose that $\mathfrak{F}$ is such that there is a world w that can see at least one world w' where $w \neq w'$. Define a model $\mathfrak{M}$ on $\mathfrak{F}$ by letting $V(p, w) = T$ and $V(p, v) = F$ for all other worlds v (including w'). Then $\mathfrak{M}, w \vDash p$ but $\mathfrak{M}, w \nvDash \Box p$, and hence $\mathfrak{M}, w \nvDash p \leftrightarrow \Box p$, as required. Next, suppose that $\mathfrak{F}$ is such that there is a world w that can see no other world. Define a model $\mathfrak{M}$ on $\mathfrak{F}$ by letting $V(p, w) = F$ and $V(p, v) = T$ for all other worlds v. Since w is a deadend, $\mathfrak{M}, w \vDash \Box \varphi$ for all formulas $\Box \varphi$, including $\Box p$. This makes the antecedent of one of the conditionals true and its consequent false, so $\mathfrak{M}, w \nvDash p \leftrightarrow \Box p$, as required. □

Corollary 11.4.98. T $\subset$ Triv*, because every* Triv *frame is reflexive.*

In the final result we will show concerning Triv, we will use the notion of a PL-transform defined in Definition 11.4.45. First, we define the concept of "collapsing" into propositional logic:

Definition 11.4.99. Any system where every wff φ is equivalent to a wff φ' containing no modal operators is said to *collapse into* PL.

If any arbitrary modal formula is equivalent to a non-modal one, then we have essentially erased modality from our system and we are back in a purely propositional setting. A system which collapses into PL effectively removes the expressive power of the modal operators. As should not be surprising, Triv is such a system:

Lemma 11.4.100. Triv *collapses into* PL.

Proof. We show that Triv collapses into PL by showing that in Triv, every formula φ is equivalent to its PL-transform. We prove this by induction on the complexity of φ. The basis case is when φ is an atom. Then $\tau(p) = p$, and $\vdash p \leftrightarrow p$, so $\vdash p \leftrightarrow \tau(p)$

Now assume that $\vdash \gamma \leftrightarrow \tau(\gamma)$ and $\vdash \psi \leftrightarrow \tau(\psi)$. The case where $\varphi := \gamma \wedge \psi$ or $\varphi := \neg\psi$ are left as exercises for the reader, as they are straightforward, and only the modal case really interests us. Let $\varphi := \Box\psi$. We need to show that $\vdash \varphi \leftrightarrow \tau(\varphi)$, which is to say, $\vdash \Box\psi \leftrightarrow \tau(\Box\psi)$. First, note that $\tau(\Box\psi) = \psi$, so we can replace the righthand side of the equivalence with ψ: $\vdash \Box\psi \leftrightarrow \psi$. But this just as a substitution instance of Triv, and hence we have the desired equivalence. □

Triv is quite a strong system (we will discuss just exactly how strong it is below), but it is not the only option out there, if we want to see just how far we can push the limits of modal logic and still be able to say something sensible about the concept of modality. Recall that we gave two ways of defining Triv: We can obtain it either by adding Triv to K, or we could add Converse-T to any system containing D. This raises a natural question: What happens if we add Converse-T directly to K? Exploring this question is left as an exercise to the reader: What can you prove, with this new axiom? Can you prove any correspondence results between that axiom and some well-defined frame condition?

Instead of answering that question, we will look at a slightly different question, namely: What happens if we add not $p \rightarrow \Box p$ to K but rather $\Box p$?

The resulting system is one that validates (or verifies) every necessity claim, which is why the system is called Ver, and the wff $\Box p$ is sometimes called axiom Ver. This system is perhaps even more difficult than Triv to see how it could be a philosophically useful or interesting system; but logicians need not always care about applicability when developing new logics. It is sufficient to note that adding Ver to K is consistent: Take any frame with single dead-end world, then $\Box p$ is valid on that frame; but p is not, and hence the addition of $\Box p$ to K is not inconsistent.

Ver also collapses into propositional logic, but in a slightly different way than Triv did. Instead of using the PL-transform, we introduce another transform, σ:

Definition 11.4.101. The transform $\sigma(\varphi)$ is defined recursively:

- $\sigma(p) = p$
- $\sigma(\neg\varphi) = \neg\sigma(\varphi)$
- $\sigma(\varphi \wedge \psi) = \sigma(\varphi) \wedge \sigma(\psi)$
- $\sigma(\Box\varphi) = \top$

- $\sigma(\Diamond\varphi) = \bot$

That is, in this transform, we replace every formula of the form $\Box\varphi$ with an arbitrary tautology ($\top$), and we replace every formula of the form $\Diamond\varphi$ with an abritrary contradiction ($\bot$), because—recall from Corollary 11.3.9—everything is necessary at a dead end and nothing is possible. We can then prove the following:

Lemma 11.4.102. *For every wff φ,*

$$\vdash_{\text{Ver}} \varphi \leftrightarrow \sigma(\varphi)$$

Proof. Exercise for the reader. □

While both Triv and Ver collapse into propositional logic, they do it in quite different ways, and for quite different reasons. While many of the other systems that we've looked at have built upon each other and are included in each other, this is not the case with Ver and Triv: Neither system is a subsystem of the other. This is because any system containing both Ver and Triv is inconsistent: For from Ver we can prove $\Box\varphi$ for any wff φ, and from Triv we can prove $\Box\varphi \leftrightarrow \varphi$, from which we can derive, by MP, φ for every wff φ. Thus, Ver and Triv are incompatible. More than this, they are both *maximal*:

Definition 11.4.103. A system S is *maximal* if $\mathrm{S} + \varphi$ is inconsistent when $\nvdash_{\mathrm{S}} \varphi$.

That is, one cannot add any non-theorem to the system without making it inconsistent. Both Triv and Ver are maximal systems:

Lemma 11.4.104. Triv *is maximal.*

Lemma 11.4.105. Ver *is maximal.*

The proof of the maximality of these two systems depends on one further, final result about both Triv and Ver; we have seen that they are inconsistent with each other, and that they are maximal, what we will show now is that they are the *only* maximal normal modal systems; every other normal modal logic is contained either in Triv or in Ver:

Theorem 11.4.106. *Every normal (cf. Definition 11.4.18) modal system* S *is contained in either* Triv *or* Ver*.*

As [Hughes and Cresswell, 1996, p. 67] note, some systems are contained in *both*; for instance, K is contained in both Triv and Ver, but so also is $\mathrm{K} + \mathsf{4}$ and B.

We'll first prove Theorem 11.4.106, and then the two maximality results. In order to prove Theorem 11.4.106, we'll first prove two intermediary results:

Lemma 11.4.107. *Every consistent system which contains* D *is contained in* Triv.

Proof. Suppose that D ⊆ S, but that S ⊈ Triv. Then there is some wff φ such that $\vdash_{\mathrm{S}} \varphi$ but $\nvdash_{\mathrm{Triv}} \varphi$. We now show that S is inconsistent, contrary to assumption. Since $\nvdash_{\mathrm{Triv}} \varphi$, its PL-transform $\tau(\varphi)$ is not valid (suppose it were; then since it is a propositional tautology, and would be a theorem of Triv; however, by Lemma 11.4.100, every wff is equivalent in Triv to its PL-transform, which would mean that $\vdash_{\mathrm{Triv}} \varphi$, making Triv inconsistent). Take a valuation V that makes $\tau(\varphi)$ false, and generate a substitution instance of φ where for any atom p, $\top$ is substituted in for p if $V(p) = T$, and for any atom p, $\bot$ is substituted in for p if $V(p) = F$. (1) The result, call it φ', is a theorem of S, because it is derived from a theorem by US. (2) It is also a modal constant formula which has the same truth value as $\tau(\varphi)$, namely, false, so it too is invalid. By Lemma 11.4.61, since $\nvDash \varphi'$, then $\vdash_{\mathrm{D}} \neg\varphi'$. But since D ⊆ S, it follows that $\vdash_{\mathrm{S}} \neg\varphi'$, which is a contradiction. □

Lemma 11.4.108. *Every consistent extension* S *of* K *which is not contained in* Ver *contains* D.

We postpone the proof of this lemma until the next section, as it uses a construction which will be familiar after we've presented the method in §11.5.1. However, once we have both of these results, we are in a position to prove Theorem 11.4.106:

Proof of Theorem 11.4.106. Immediately from the previous two results. Let S be a consistent system extending K. If S ⊆ Ver, then we are done. If S ⊈ Ver, then by Lemma 11.4.108, D ⊆ S. If D ⊆ S, then by Lemma 11.4.107, S ⊆ Triv, and we are done. □

And we are now in a position to prove Lemmas 11.4.104 and 11.4.105. We do just one case; the other is proven entirely analogously.

Proof of Lemma 11.4.104. Suppose that Triv is not maximal. Then there is some wff φ where $\nvdash_{\mathrm{Triv}} \varphi$. Since φ is not a theorem, it follows that Triv + φ is not contained in Triv. By Theorem 11.4.106, if Triv + φ is consistent, then Triv + φ ⊆ Ver. But then Ver is inconsistent, because we would be able to derive p in it, as we showed above. So Triv + φ must already be inconsistent, in which case it is trivially maximal. □

The inclusion diagram for all systems considered so far is given in Figure 11.18.

What we have see in this section is two very strong systems, that trivialise the concept in modality in diferent ways, both resulting in a collapse into propositional logic. It might be argued that such systems are irrelevant, and that there is no purpose in studying them.

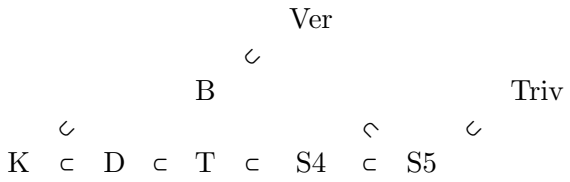


Figure 11.18: Inclusion relations of systems so far

However, there are philosophical motivations for considering systems where truth and necessity are equivalent. For example, Triv, which collapses necessity and truth, is relevant to the study of Spinozist metaphysics, as discussed [Jarrett, 1978]. If there is only one possible world, which is the case with Spinoza's monad, then the logic of a class of frames where each frame only has one world is exactly the right logic to capture the notion of necessity in that metaphysical framework.

This is a topic I hope to expand upon.

11.5 Completeness

Completeness is the agreement between a set of theorems and derivation rules and a set of formulas valid on a class of models. If a syntactic, derivation system S is complete with respect to a class of models, then every formula valid in the class of models can be derived in the corresponding derivation system. (The converse, that of derivability implying validity in a class of models, is soundness; it is much easier to prove and we will not consider it here, since this is covered in basic modal logic courses.)

In the case of modal logic, we are generally interested in the class of models which are models based on frames satisfying a certain frame condition. Thus, we will consider classes of frames rather than classes of models when discussing completeness. Recall that for a class of frames $\mathfrak{C}$, we say that a formula φ is valid in $\mathfrak{C}$ iff φ is valid on every frame $\mathfrak{F} \in \mathfrak{C}$, that is, for any model based on that frame, $\mathfrak{M}, w \vDash \varphi$. A system S is *complete* with respect to a class of frames $\mathfrak{C}$ iff every $\mathfrak{C}$-valid formula φ is a theorem of S.

11.5.1 An effective completeness proof for S5

In this section, we give a proof that S5 is complete with respect to the class $\mathfrak{C}$ of frames that are reflexive, transitive, and symmetric. That is, our main result in this section is the following:

Theorem 11.5.1. *Let $\mathfrak{C}$ be the class of reflexive, transitive, and symmetric frames. If $\mathfrak{C} \vDash \varphi$, then $\vdash_{\mathrm{S5}} \varphi$.*

The method by which we prove this theorem is *effective* in the sense that for every wff φ valid in $\mathfrak{C}$, we will give a method for construction a proof of φ in S5. The presentation of the method in this section is greatly indebted to [Hughes and Cresswell, 1996, ch. 5].

We begin with a number of definitions. First we define a normal form for modal formulas which is analogous to the Conjunctive and Disjunctive Normal Forms defined for propositional formulas (cf. Definitions 7.4.22 and 7.4.23): Modal Conjunctive Normal Form.

Definition 11.5.2 (Modal Conjunctive Normal Form)**.** A wff φ is said to be in *modal conjunctive normal form* if it is a conjunction (possibly degenerate) in which each conjunct is a disjunction (possibly degenerate) of (a) propositional formulas, (b) $\Box\alpha$, or (c) $\Diamond\alpha$, where α is a non-modal (i.e., propositional) formula.

(Recall that a degenerate conjunction or disjunction is one that has only one conjunct or disjunct. Thus, the disjunction $p \vee q$ can be thought of as a degenerate conjunction, as it is equivalent to $(p \vee q) \wedge (p \vee q)$; and the conjunction $p \wedge q$ can be thought of as a degenerate disjunction, as it is equivalent to $(p \wedge q) \vee (p \wedge q)$.)

We give a mechanism for converting an arbitrary wff into a wff in MCNF in Lemma 11.5.8; for the time being, we simply list some examples of formulas that are in MCNF, and some that are not:

Example 11.5.3. The following wffs are all in Modal CNF:

$$p$$
$$p \vee \Diamond r$$
$$(p \vee \Box s) \wedge \Diamond r$$
$$(\Box p \vee \Diamond \neg (p \vee r) \vee \neg (p \to q)) \wedge (\Diamond p \vee r)$$

Example 11.5.4. The following wffs are not in Modal CNF:

$$(\Diamond (p \vee q) \wedge r) \vee \Box s$$
$$\Box \Diamond p$$
$$\Box (\Diamond \neg r \vee q \vee s) \wedge (\Box q \vee \Diamond p)$$

We introduce now a refinement on MCNF, which orders the disjuncts of each conjunct in a particular order:

Definition 11.5.5 (Ordered MCNF)**.** Take a wff φ that is in modal CNF.

(1) Using the rules of commutivity, arrange the disjuncts of each conjunct so that all of the unmodalized disjuncts are listed first, then all disjuncts prefixed with $\Box$, and then all the disjuncts prefixed with $\Diamond$. The resulting wff will have the following structure:

$$\beta \vee \Box\gamma_1 \vee \cdots \vee \Box\gamma_n \vee \Diamond\delta_1 \vee \cdots \vee \Diamond\delta_m \qquad (\varphi')$$

where $n \geq 0, m \geq 0$, and β and all the γ_i and δ_i are purely propositional.

(2) Using the law of $\Diamond$-over-$\vee$ distribution, replace

$$\Diamond\delta_1 \vee \cdots \vee \Diamond\delta_m$$

with

$$\Diamond(\delta_1 \vee \cdots \vee \delta_m)$$

which we will now simply call $\Diamond\delta$. φ' is now in the following form:

$$\beta \vee \Box\gamma_1 \vee \cdots \vee \Box\gamma_n \vee \Diamond\delta \qquad (\varphi'')$$

where each of $\beta, \gamma_1, \ldots, \gamma_n, \delta$ are all purely propositional.

The resulting formula φ'' is said to be in *ordered modal conjunctive normal form* (ordered-MCNF).

In propositional logic, every wff is equivalent to a wff in CNF. A similar result for modal logic, i.e., that every wff is equivalent to a wff in Modal CNF, is not generally provable. In fact, in most systems, it is not the case that every theorem can be reduced to an equivalent wff in Modal CNF; for example, in S4, the following is not equivalent to any Modal CNF wff:

$$\Diamond(p \wedge \Diamond\neg p)$$

A crucial result for the construction of the method in this chapter is the following:

Theorem 11.5.6. *Every theorem of* S5 *is equivalent to a wff in (Ordered) Modal Conjunctive Normal Form.*

We refrain from proving this until later below. The proof goes via the notion of the 'modal degree' of a wff.

Definition 11.5.7 (Modal degree)**.** The *modal degree* of a formula φ is defined inductively:

1. The modal degree of an atomic proposition is 0.

2. If φ is of degree n, then $\neg\varphi$ is of degree n.

3. If φ is of degree n and ψ is of degree m, then $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \rightarrow \psi$ are all of degree $\mathsf{max}\{n, m\}$.

4. If φ is of degree n, then $\Box\varphi$ and $\Diamond\varphi$ are of degree $n + 1$.

That is, the modal degree of a formula can be computed by looking at the number of modalities that are within the scope of another modality. A wff of first-degree has one modal operator that is not within the scope of any other modality; a wff of second-degree has at least one modal operator that has within its scope a wff of first-degree.

Note that the modal degree of a formula should not be confused with the number of modalities in an iterated modality (though there is a connection between the two in that a wff with a standard-form iterated modality of length n will have modal degree at least n). This is because a wff can have modal degree n without having any iterated modalities, as the following formula of modal degree 3 illustrates:

$$\Diamond(p \wedge \Box(\Diamond\neg r \vee s)) \tag{11.23}$$

Here, there are no iterated modalities but there are what we can call *bracketed* modalities, that is, modal operators that are within the scope of other modal operators via parentheses.

There is a simple visual method for calculating the modal degree of a formula. First, label every atom with 0:

$$\begin{array}{cccccccccc} \Diamond & (p & \wedge & \Box & (\Diamond & \neg & r & \vee & s)) \\ & 0 & & & & & 0 & & 0)) \end{array}$$

Once we have the atoms labeled, we can calculate the degrees of each more complex formula according to Definition 11.5.7. If r is of degree 0, then so too is $\neg r$:

$$\begin{array}{cccccccccc} \Diamond & (p & \wedge & \Box & (\Diamond & \neg & r & \vee & s)) \\ & 0 & & & & 0 & 0 & & 0)) \end{array}$$

However, when we add $\Diamond$ to the subformula $\neg r$, that increases in the modal degree by 1:

$$\begin{array}{cccccccccc} \Diamond & (p & \wedge & \Box & (\Diamond & \neg & r & \vee & s)) \\ & 0 & & & 1 & 0 & 0 & & 0)) \end{array}$$

Next, we look at the disjunction $\Diamond\neg r \vee s$, and take the maximum of the modal degrees of the two disjuncts, that is $\mathsf{max}\{1,0\}$:

$$\begin{array}{cccccccccc} \Diamond & (p & \wedge & \Box & (\Diamond & \neg & r & \vee & s)) \\ & 0 & & & 1 & 0 & 0 & 1 & 0)) \end{array}$$

This disjunction is necessitated by the addition of $\Box$, which means we increase the modal degree by 1:

$$\begin{array}{cccccccccc} \Diamond & (p & \wedge & \Box & (\Diamond & \neg & r & \vee & s)) \\ & 0 & & 2 & 1 & 0 & 0 & 1 & 0)) \end{array}$$

Next, we have again a propositional combination, the conjunction, so we take the maximum of the modal degrees of the two conjuncts, that is $\mathsf{max}\{2,0\}$:

$$\begin{array}{cccccccccc} \Diamond & (p & \wedge & \Box & (\Diamond & \neg & r & \vee & s)) \\ & 0 & 2 & 2 & 1 & 0 & 0 & 1 & 0)) \end{array}$$

Finally, that entire formula is within the scope of a further modal operator, $\Diamond$, so we increase the modal degree again by 1:

$$\begin{array}{cccccccccc} \Diamond & (p & \wedge & \Box & (\Diamond & \neg & r & \vee & s)) \\ 3 & 0 & 2 & 2 & 1 & 0 & 0 & 1 & 0)) \end{array}$$

The number underneath the main connective of a formula will be the modal degree of the entire formula.

The reduction laws in S5 prove that any wff of modal degree n where the degree is induced by an iterated modality can be reduced, in S5, to a first-degree formula. However, we can in fact prove a further, stronger result in S5:

Lemma 11.5.8. *Every wff of higher than first-degree is reducible in* S5 *to a first-degree formula.*

Proof. We give an effective procedure for converting any second-degree wff φ to a wff φ' of first-degree; repeated applications of this procedure will be sufficient to demonstrate the desired result. Our procedure will make use of the already-proven modal equivalences collected in Figure 11.19 and various propositional equivalences.

1. Convert all occurences of $\varphi \to \psi$ to $\neg\varphi \vee \psi$.

2. Apply DeMorgan's laws, the law of double negation, and MI to eliminate any negations before $\Diamond$, $\Box$, or (.

3. Reduce all iterated modalities according to the S5-reduction laws (Figure 11.13).

4. If, after applying steps 1–3, we do not yet have a first-degree wff, then it can only be because there is a modality in the scope of another modality that is *not* an iterated modality. That is, it is of the form $\Box\varphi$ or $\Diamond\varphi$ where φ contains a modality and is either

1	MI	$\neg \Box \neg p \leftrightarrow \Diamond p$	$\neg \Diamond \neg p \leftrightarrow \Box p$
2	$\Box$-over-$\wedge$	$\Box(p \wedge q) \leftrightarrow (\Box p \wedge \Box q)$	
3	$\Diamond$-over-$\vee$	$\Diamond(p \vee q) \leftrightarrow (\Diamond p \vee \Diamond q)$	
4	$\Box$-over-$\vee$	$\Box(p \vee \Box q) \leftrightarrow (\Box p \vee \Box q)$	$\Box(p \vee \Diamond q) \leftrightarrow (\Box p \vee \Diamond q)$
5	$\Diamond$-over-$\wedge$	$\Diamond(p \wedge \Diamond q) \leftrightarrow (\Diamond p \wedge \Diamond q)$	$\Diamond(p \wedge \Box q) \leftrightarrow (\Diamond p \wedge \Box q)$

Figure 11.19: Relevant S5 reduction rules

a conjunction or a disjunction. Consider the case of $\Box\varphi$ first. If φ is a conjunction, then $\Box$ distributes over it via (2) in Figure 11.19. If φ is a disjunction, then if it has a modalised disjunct, $\Box$ distributes over it via the rules in (4). If no disjunction is modalised, but at least one disjunction contains a modality, then use the distrubtion law $p \vee (q \wedge r) \leftrightarrow (p \vee q) \wedge (p \vee r)$ to convert the disjunction into a conjunction and then apply (2). The case for $\Diamond\varphi$ proceeds similarly, but uses the distribution law $p \wedge (q \vee r) \leftrightarrow (p \wedge q) \vee (p \wedge r)$.

5. If the result is still not a first-degree formula, apply the steps again. Because each pass through the algorithm removes modal operators and each wff has a finite number of modal operators, this process will eventually terminate, and the resulting wff will be in MCNF.

□

Example 11.5.9. $\Box((\Diamond p \vee q) \rightarrow \neg(\Box p \vee \Diamond q))$ is a second-degree wff and hence is reducible in S5 to an equivalent first-degree wff.

1. Eliminate implications:
$$\Box(\neg(\Diamond p \vee q) \vee \neg(\Box p \vee \Diamond q))$$

2. Apply DeMorgan's, double negation, and MI:
$$\Box((\Box\neg p \wedge \neg q) \vee (\Diamond\neg p \wedge \Box\neg q))$$

3. Distribute $\vee$ over $\wedge$:
$$\Box(((\Box\neg p \wedge \neg q) \vee \Diamond\neg p) \wedge ((\Box\neg p \wedge \neg q) \vee \Box\neg q))$$

4. Distribute $\Box$ over $\wedge$:
$$\Box((\Box\neg p \wedge \neg q) \vee \Diamond\neg p) \wedge \Box((\Box\neg p \wedge \neg q) \vee \Box\neg q)$$

5. Distribute $\Box$ over $\vee$:

$$(\Box(\Box\neg p \wedge \neg q) \vee \Diamond\neg p) \wedge (\Box(\Box\neg p \wedge \neg q) \vee \Box\neg q)$$

6. Distribute $\Box$ over $\wedge$:

$$((\Box\,\Box\,\neg p \wedge \Box\neg q) \vee \Diamond\neg p) \wedge ((\Box\,\Box\,\neg p \wedge \Box\neg q) \vee \Box\neg q)$$

7. Reduce iterated modalities.

$$((\Box\neg p \wedge \Box\neg q) \vee \Diamond\neg p) \wedge ((\Box\neg p \wedge \Box\neg q) \vee \Box\neg q)$$

This wff is now a first-degree wff.

We are now in a position to prove Theorem 11.5.6:

Proof of Theorem 11.5.6. First, given Lemma 11.5.8, all we need to show is that any wff φ of degree 1 or 0 can be reduced in S5 to (O)MCNF. We thus restrict our attention to those cases, and give an effective procedure for reducing them to (O)MCNF:

1. If φ is a propositional wff, then it is of modal degree 0 and it is in (O)MCNF by definition.

2. If φ is a first-degree wff, then it is a truth-function of wffs each of which is either a 0-degree wff or is of the form $\Box\psi$ or $\Diamond\psi$, where ψ is a 0-degree wff. By treating each of these wffs as atoms, we can then convert φ to CNF (this is always possible by Theorem 7.4.27). We then use MI to convert $\neg\Box$ and $\neg\Diamond$ to $\Diamond\neg$ and $\Box\neg$, respectively. The result, φ', is in MCNF.

3. If φ' is not already in OMCNF, it can be converted to OMCNF at this point, if desired, using the procedure given in Definition 11.5.5.

□

We now give an example of how Theorem 11.5.6 works:

Example 11.5.10.

$$(\Box\neg p \wedge \Box\neg q) \vee (\Diamond\neg p \wedge \Box\neg q)$$

is a first-degree wff and hence is reducible in S5 to OMCNF.

We treat each of $\Box\neg p$, $\Box\neg q$, $\Diamond\neg p$, and $\Box\neg q$ (again) as atoms. First we distribute $\Diamond\neg p\wedge\Box\neg q$ over $\Box\neg p\wedge\Box\neg q$:

$$(\Box\neg p\vee(\Diamond\neg p\wedge\Box\neg q))\wedge(\Box\neg q\vee(\Diamond\neg p\wedge\Box\neg q))$$

We now distribute the left conjunct:

$$((\Box\neg p\vee\Diamond\neg p)\wedge(\Box\neg p\vee\Box\neg q))\wedge(\Box\neg q\vee(\Diamond\neg p\wedge\Box\neg q))$$

Now the right conjunct:

$$((\Box\neg p\vee\Diamond\neg p)\wedge(\Box\neg p\vee\Box\neg q))\wedge((\Box\neg q\vee\Diamond\neg p)\wedge(\Box\neg q\vee\Box\neg q))$$

By associativity, this is equivalent to:

$$(\Box\neg p\vee\Diamond\neg p)\wedge(\Box\neg p\vee\Box\neg q)\wedge(\Box\neg q\vee\Diamond\neg p)\wedge(\Box\neg q\vee\Box\neg q)$$

This is now in MCNF, and is also already in OMCNF because each in each conjunct we have grouped all the disjuncts with $\Box$ before the disjuncts with $\Diamond$.

The next example puts together the two results—we take an arbitrary high-degree formula and reduce it to a first-degree formula, and then convert it to OMCNF.

Example 11.5.11. $\Box(\Diamond\Diamond p\to p)\to\Box(p\to\Box p)$ is a third-degree wff and hence is reducible in S5 to a first-degree wff and thence to OMCNF.

Reduction to first degree:

1. Eliminate implications:

$$\neg\,\Box\,(\neg\,\Diamond\,\Diamond p\vee p)\vee\Box(\neg p\vee\Box p)$$

2. Apply DeMorgan's, double negation, and MI.

$$\Diamond(\Diamond\,\Diamond\,p\wedge\neg p)\vee\Box(\neg p\vee\Box p)$$

3. Distribute modalities.

$$(\Diamond\,\Diamond\,p\wedge\Diamond\neg p)\vee(\Box\neg p\vee\Box p)$$

4. Reduce iterated modalities.

$$(\Diamond p\wedge\Diamond\neg p)\vee(\Box\neg p\vee\Box p)$$

Conversion to OMCNF:

1. Distribute $\vee$ over $\wedge$:

$$(\Diamond p \vee (\Box \neg p \vee \Box p)) \wedge (\Diamond \neg p \vee (\Box \neg p \vee \Box p))$$

2. Remove superfluous parentheses:

$$(\Diamond p \vee \Box \neg p \vee \Box p) \wedge (\Diamond \neg p \vee \Box \neg p \vee \Box p)$$

3. Re-order disjuncts:

$$(\Box \neg p \vee \Box p \vee \Diamond p) \wedge (\Box \neg p \vee \Box p \vee \Diamond \neg p)$$

This is now in OMCNF.

The purpose of all of this converting of wffs into equivalent, but simpler, wffs with a rigid structure is to provide us a means of identifying properties of wffs which will allow us to construct S5-proofs of them. To that end, we define a test that we will subject wffs in OMCNF to such that ones that pass the test will be provable in S5 and ones that don't will not be. Furthermore, for the ones that fail the test, we will give a method for constructing a reflexive, symmetric, and transitive model in which they are false.

Definition 11.5.12. Let φ be a wff in OMCNF. First, note that every wff in OMCNF is of the form $C_1 \wedge \cdots \wedge C_k$ where each C_i $(1 \leq i \leq k)$ is of the form:

$$\beta \vee \Box \gamma_1 \vee \cdots \vee \Box \gamma_n \vee \Diamond \delta$$

For each C_i in φ we will form $\leq n+1$ disjunctions (n if C_i contains no β, $n+1$ if C_i contains β) by taking δ and disjoining it to each of the β and γ_i, e.g.:

$$\begin{array}{l} \beta \vee \delta \\ \gamma_1 \vee \delta \\ \vdots \\ \gamma_n \vee \delta \end{array}$$

We say that C_i *passes the test* if at least one of the disjunctions so constructed is a tautology. We say that a wff φ *passes the test* if every conjunct C_i of φ passes the test.

Example 11.5.13. $(\Box \neg p \vee \Box p \vee \Diamond p) \wedge (\Box \neg p \vee \Box p \vee \Diamond \neg p)$ passes the test. The wff has two conjuncts:

- $C_1 := \Box \neg p \vee \Box p \vee \Diamond p$

- $C_2 := \Box\neg p \vee \Box p \vee \Diamond\neg p$.

Take C_1. We form two disjunctions, (1) $\neg p \vee p$ and (2) $p \vee p$. Because (1) is a tautology, C_1 passes the test.

Take C_2. We form two disjunctions, (1) $\neg p \vee \neg p$ and (2) $p \vee \neg p$. Because (2) is a tautology, C_2 passes the test.

Because both conjuncts pass the test, $(\Box\neg p \vee \Box p \vee \Diamond p) \wedge (\Box\neg p \vee \Box p \vee \Diamond\neg p)$ passes the test.

Example 11.5.14. $(\Box\neg p \vee \Diamond\neg p) \wedge (\Box\neg p \vee \Box\neg q) \wedge (\Box\neg q \vee \Diamond\neg p) \wedge (\Box\neg q \vee \Box\neg q)$ does not pass the test. The wff has four conjuncts:

- $C_1 := \Box\neg p \vee \Diamond\neg p$.
- $C_2 := \Box\neg p \vee \Box\neg q$.
- $C_3 := \Box\neg q \vee \Diamond\neg p$.
- $C_4 := \Box\neg q \vee \Box\neg q$.

Take C_1. We form one disjunction, (1) $\neg p \vee \neg p$. We do not need to go any further, for (1) is not a tautology, so C_1 doesn't pass the test, so the entire wff does not pass the test.

We are now in a position to prove Theorem 11.5.1. We do so via the following two lemmas:

Lemma 11.5.15. *Let $\mathfrak{C}$ be the class of reflexive, transitive, and symmetric frames. If $\mathfrak{C} \vDash \varphi$, then φ in OMCNF passes the test.*

Lemma 11.5.16. *If φ in OMCNF passes the test, then $\vdash_{\mathrm{S5}} \varphi$.*

Proof of Lemma 11.5.15. We prove this lemma contrapositively. That is, suppose φ does *not* pass the test. Then we construct a reflexive, transitive, and symmetric model with a world where φ is false.

If φ does not pass the test, then there is some conjunct C_i of φ that does not pass the test (if there is more than one, this does not matter, we only need one). C_i is of the form

$$\beta \vee \Box\gamma_1 \vee \cdots \vee \Box\gamma_n \vee \Diamond\delta$$

and none of $\beta \vee \delta$, $\gamma_1 \vee \delta$, $\ldots \gamma_n \vee \delta$ is a tautology, and each is a purely propositional formula.

Let $\mathfrak{F}$ be a reflexive, transitive, and symmetric frame with at least n+1 worlds, $w_0, \ldots, w_n$. We define a model $\mathfrak{M}$ on $\mathfrak{F}$ that is such that $V(\beta, w_0) = F$ and $V(\delta, w_0) = F$. (Since $\beta \vee \delta$ is

not a tautology, we know that there is at least one valuation that makes it false, and we simply take that valuation to apply to w_0.) Similarly, we let $V(\gamma_1, w_1) = F$ and $V(\delta, w_1) = F$, and so on for each of the γ_n. For any other world, let the valuation be identical to the valuation at w_0.

Now, $V(\beta, w_0) = F$, by definition. Because w_0 sees all the other w_i, including at least one world which falsifies each γ_i, it follows that $V(\Box\gamma_i, w_0) = F$ for every γ_i. Further, because no world makes δ true, it follows that $V(\Diamond\delta, w_0) = F$. Hence, $V(\beta \vee \Box\gamma_1 \vee \cdots \vee \Box\gamma_n \vee \Diamond\delta, w_0) = F$, showing that φ is not S5-valid. □

Before we prove Lemma 11.5.16, we give an example of how this method works.

Example 11.5.17. We showed above that $(\Box\neg p \vee \Diamond\neg p) \wedge (\Box\neg p \vee \Box\neg q) \wedge (\Box\neg q \vee \Diamond\neg p) \wedge (\Box\neg q \vee \Box\neg q)$ does not pass the test, and in particular that C_1 fails the test. We have one disjunction, $\neg p \vee \neg p$. Let $\mathfrak{M}$ be a model that has exactly one world, w_1, which can see itself. This model (depicted in Figure 11.20) is reflexive and degenerately symmetric and transitive. By setting $V(p, w_1) = T$, we obtain $V(\neg p, w_1) = F$, and hence that $V(\neg p \vee \neg p, w_1) = F$. Further, $V(\Box\neg p, w_1) = F$ and $V(\Diamond\neg p, w_1) = F$, so C_1 is false at w_1, and hence the entire formula is false at w_1.

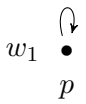


Figure 11.20: A countermodel for $(\Box\neg p \vee \Diamond\neg p) \wedge (\Box\neg p \vee \Box\neg q) \wedge (\Box\neg q \vee \Diamond\neg p) \wedge (\Box\neg q \vee \Box\neg q)$

Proof of Lemma 11.5.16. To prove this, we give a procedure that converts conjuncts that pass the test into S5-theorems. Let φ be in OMCNF and let C_i be a conjunct of φ that passes the test. That means, one of $\beta \vee \delta$, $\gamma_1 \vee \delta$,…,$\gamma_n \vee \delta$ is valid.

Suppose that it is $\beta \vee \delta$ that is valid. We then prove C_i as follows:

1	$\beta \vee \delta$	PL
2	$\delta \rightarrow \Diamond\delta$	T1
3	$\beta \vee \Diamond\delta$	1, 2, PL
4	$\beta \vee \Box\gamma_1 \vee \cdots \vee \Box\gamma_n \vee \Diamond\delta$	3, PL

Suppose that it is one of the $\gamma_i \vee \delta$, that is valid (suppose it is the one with γ_j as a disjunct). We then prove C_i as follows:

1	$\gamma_j \vee \delta$	PL
2	$\Box(\gamma_j \vee \delta)$	N
3	$\Box\gamma_j \vee \Diamond\delta$	K9
4	$\beta \vee \Box\gamma_1 \vee \cdots \vee \Box\gamma_n \vee \Diamond\delta$	3, PL

□

Proof of Theorem 11.5.1. It follows immediately from Lemmas 11.5.15 and 11.5.16 that if φ is valid on the class of reflexive, transitive, and symmetric frames, then there is a proof of φ in S5. □

In order to get to the point where we could prove this theorem, we had to take quite a meandering path through a number of definitions and lemmas, so before we conclude this section we'll put all the pieces back together with an example.

Example 11.5.18. Pick an arbitrary wff φ, say, the following:

$$\Box \Diamond (\Diamond p \vee (q \to \neg p)) \to \Diamond(\Diamond q \to \neg \Diamond p) \tag{11.24}$$

We want to determine whether this formula is a theorem of S5.

It is a formula of modal degree 3, so by Lemma 11.5.8 it can be reduced to an equivalent 1st-degree wff.

Reduction to first degree:

1. Eliminate implications:

$$\neg \Box \Diamond(\Diamond p \vee (\neg q \vee \neg p)) \vee \Diamond(\neg \Diamond q \vee \neg \Diamond p)$$

2. Apply MI.

$$\Diamond \Box \neg(\Diamond p \vee (\neg q \vee \neg p)) \vee \Diamond(\Box\neg q \vee \Box\neg p)$$

3. Apply DeMorgan's and double negation.

$$\Diamond \Box (\neg \Diamond p \wedge \neg(\neg q \vee \neg p)) \vee \Diamond(\Box\neg q \vee \Box\neg p)$$

4. Apply DeMorgan's, double negation, and MI again.

$$\Diamond \Box (\Box\neg p \wedge (q \wedge p)) \vee \Diamond(\Box\neg q \vee \Box\neg p)$$

5. Distribute modalities ($\Box$ over $\wedge$).

$$\Diamond(\Box\,\Box\,\neg p \wedge \Box(q \wedge p)) \vee \Diamond(\Box\neg q \vee \Box\neg p)$$

6. Distribute modalities again ($\Diamond$ over $\wedge$-with-modalized conjunct).

$$(\Diamond\,\Box\,\Box\neg p \wedge \Box(q \wedge p)) \vee \Diamond(\Box\neg q \vee \Box\neg p)$$

7. Distribute modalities again ($\Box$ over $\wedge$).

$$(\Diamond\,\Box\,\Box\neg p \wedge (\Box q \wedge \Box p)) \vee \Diamond(\Box\neg q \vee \Box\neg p)$$

8. Distribute modalities again ($\Diamond$ over $\vee$).

$$(\Diamond\,\Box\,\Box\neg p \wedge (\Box q \wedge \Box p)) \vee (\Diamond\,\Box\,\neg q \vee \Diamond\,\Box\,\neg p)$$

9. Reduce iterated modalities.

$$(\Box\neg p \wedge (\Box q \wedge \Box p)) \vee (\Box\neg q \vee \Box\neg p)$$

This is a first-degree wff. According to Theorem 11.5.6, this is equivalent to a wff in OMCNF. We convert it now.

Conversion to OMCNF:

1. Distribute $\vee$ over $\wedge$:

$$(\Box\neg p \vee (\Box\neg q \vee \Box\neg p)) \wedge (\Box q \vee (\Box\neg q \vee \Box\neg p)) \wedge (\Box p \vee (\Box\neg q \vee \Box\neg p))$$

2. Remove superfluous parentheses:

$$(\Box\neg p \vee \Box\neg q \vee \Box\neg p) \wedge (\Box q \vee \Box\neg q \vee \Box\neg p) \wedge (\Box p \vee \Box\neg q \vee \Box\neg p)$$

This is in MNCF, and is trivially in OMCNF, since every disjunct in each conjunction is under the scope of $\Box$. We now perform the test on each conjunct.

Performing The Test:

- $C_1 := \Box\neg p \vee \Box\neg q \vee \Box\neg p$. Because there is no $\Diamond$-disjunct, we form three degenerate disjunctions, one for each γ_i:

$$\begin{array}{ll} \gamma_1 & \neg p \\ \gamma_2 & \neg q \\ \gamma_3 & \neg p \end{array}$$

  None of these are tautologies, so the first conjunct fails the test.

- C_2 and C_3: Because we have already found one conjunct that fails the test, we don't need to test either of these.

Since the first conjunct fails the test, by Lemma 11.5.15, we can construct a reflexive, transitive, and symmetric countermodel that falsifies φ, by falsifying each of the γ_i in the first conjunct. We therefore need two worlds, w_1 and w_2, one where γ_1 is false and one where γ_2 is false (because $\gamma_1 = \gamma_3$, we can simplify our model by identifying the world where γ_1 is false with the world where γ_3 is false). Such a model is given in Figure 11.21.

Figure 11.21: A countermodel showing that $\Box \Diamond (\Diamond p \vee (q \to \neg p)) \to \Diamond(\Diamond q \to \neg \Diamond p)$ is not valid in S5.

Because $w_1 \vDash p$, $w_1 \vDash \neg \Box \neg p$; and since $w_2 \vDash q$, $w_1 \vDash \neg \Box \neg q$. Thus, $w_1 \vDash \neg \Box \neg p \wedge \neg \Box \neg q$, and therefore C_1 is false at w_1. As a result, the entire conjunction is false and, converting that conjunction back into the original φ, φ is false.

Finally, we give an example that does work.

Example 11.5.19. Consider the following formula:

$$\Diamond(p \wedge q) \to \Box(\Box(\Box p \to \Box q) \to \Diamond q) \tag{11.25}$$

We want to determine whether this formula is a theorem of S5. It is a formula of modal degree 3, so by Lemma 11.5.8 we can reduce it to an equivalent 1st-degree wff.

Reduction to first degree:

1. Eliminate implications:

$$\neg \Diamond (p \wedge q) \vee \Box(\neg \Box (\neg \Box p \vee \Box q) \vee \Diamond q)$$

2. Apply MI:

$$\Box\neg(p \wedge q) \vee \Box(\Diamond\neg(\Diamond\neg p \vee \Box q) \vee \Diamond q)$$

3. Apply DeMorgan's and MI again:

$$\Box(\neg p \vee \neg q) \vee \Box(\Diamond(\Box p \wedge \Diamond\neg q) \vee \Diamond q)$$

4. Distribute the modality in red:

$$\Box(\neg p \vee \neg q) \vee \Box((\Diamond \Box p \wedge \Diamond \neg q) \vee \Diamond q)$$

5. Distribute the modality in blue:

$$\Box(\neg p \vee \neg q) \vee (\Box(\Diamond \Box p \wedge \Diamond \neg q) \vee \Diamond q)$$

6. Distribute it again:

$$\Box(\neg p \vee \neg q) \vee ((\Box \Diamond \Box p \wedge \Box \Diamond \neg q) \vee \Diamond q)$$

7. Reduce iterated modalities:

$$\Box(\neg p \vee \neg q) \vee ((\Box p \wedge \Diamond \neg q) \vee \Diamond q)$$

This formula is now a first-degree wff. According to Theorem 11.5.6, this is equivalent to a wff in OMCNF. We conver it now:

Conversion to OMCNF:

1. Distribute $\wedge$ over $\vee$:

$$\Box(\neg p \vee \neg q) \vee ((\Box p \vee \Diamond q) \wedge (\Diamond \neg q \vee \Diamond q))$$

2. Distribute $\vee$ over $\wedge$:

$$((\Box p \vee \Diamond q) \vee \Box(\neg p \vee \neg q)) \wedge ((\Diamond \neg q \vee \Diamond q) \vee \Box(\neg p \vee \neg q))$$

3. Associativity and communtativity:

$$(\Box p \vee \Diamond q \vee \Box(\neg p \vee \neg q)) \wedge (\Diamond \neg q \vee \Diamond q \vee \Box(\neg p \vee \neg q))$$

4. Order disjuncts:

$$(\Box p \vee \Box(\neg p \vee \neg q) \vee \Diamond q) \wedge (\Box(\neg p \vee \neg q) \vee \Diamond \neg q \vee \Diamond q)$$

5. Combine $\Diamond$ wffs into one:

$$(\Box p \vee \Box(\neg p \vee \neg q) \vee \Diamond q) \wedge (\Box(\neg p \vee \neg q) \vee \Diamond(\neg q \vee q))$$

This is in OMCNF, and there are two conjuncts. We can now perform the test.

Performing the test:

- $C_1 := \Box p \vee \Box(\neg p \vee \neg q) \vee \Diamond q$. Because there is no purely-propositional disjunct and two $\Box$ disjuncts, we form two disjunctions:

$$\begin{array}{l|l} \gamma_1 \vee \delta & p \vee q \\ \gamma_2 \vee \delta & \neg p \vee \neg q \vee q \end{array}$$

  Because γ_2 is a tautology, this conjunct passes the test.

- $C_2 := \Box(\neg p \vee \neg q) \vee \Diamond(\neg q \vee q)$. Because there is no purely-propositional disjunct and one $\Box$ disjunct, we form one disjunction:

$$\gamma_1 \vee \delta \mid \neg p \vee \neg q \vee \neg q \vee q$$

  Because this is a tautology, the second conjunct also passes the test.

Because both conjuncts pass the test, we can prove each conjunct in S5. The first conjunct:

1	$\neg p \vee \neg q \vee q$	PL
2	$\Box((\neg p \vee \neg q) \vee q)$	$\mathsf{N}, 1$
3	$\Box((\neg p \vee \neg q) \vee \Diamond q$	$\mathsf{K9}, 2$
4	$\Box p \vee \Box(\neg p \vee \neg q) \vee \Diamond q$	PL, 3

The second conjunct:

5	$\neg p \vee \neg q \vee \neg q \vee q$	PL
6	$\Box(\neg p \vee \neg q \vee \neg q \vee q)$	$\mathsf{N}, 5$
7	$\Box(\neg p \vee \neg q) \vee \Diamond(\neg q \vee q)$	$\mathsf{K9}, 6$

The rest of the proof can be completed by propositional reasoning and by reversing all of the equivalences used to generate the OMCNF form.

Earlier, we noted that we were going to delay the proof of Lemma 11.4.108 until after we'd introduced the procedure just given. We prove it now.

Proof of Lemma 11.4.108. Suppose that S is consistent extension of K and that there is some φ such that $\vdash_{\mathrm{S}} \varphi$ but $\nvdash_{\mathrm{Ver}} \varphi$. We will show that S must contain D (and hence, by the previous result, is contained in Triv).

Recall (Lemma 11.4.59) that any system S which has a theorem of the form $\Diamond\varphi$ will contain D; so what we need to show is that $\vdash_{\mathrm{S}} \Diamond\varphi$ for some φ. Note that every propositional modal wff is built up out of wffs of three kinds: (a) ones with no modal operators, (b) ones of

the form $\Box\psi$, and (c) ones of the kind $\Diamond\psi$ (where, unlike above, we allow ψ to be of any modal degree). As a result, any modal wff φ can be reduced to a wff in (propositional) Conjunctive Normal Form, where every disjunct in each conjunct is either (a) purely propositional, (b) of the form $\Box\psi$, or (c) of the negation of a wff of form $\Diamond\psi$. Further, using modal interchange, commutativity, and K8, we can convert this CNF wff into a wff φ' which has only one subformula beginning with $\Diamond$ in any individual conjunct, and $\vdash_{\mathrm{K}} \varphi \leftrightarrow \varphi'$.

Now, φ' is a conjunction of the form

$$C_1 \wedge \cdots \wedge C_n$$

where each C_i is one of the following:

1. A purely propostional wff,

2. A disjunction with at least on disjunct of the form $\Box\psi$,

3. A wff of the form $\Diamond\psi$,

4. A wff of the form $\gamma \vee \Diamond\psi$ where γ is purely propositional.

Since $\vdash_{\mathrm{K}} \varphi \leftrightarrow \varphi'$, and $\vdash_{\mathrm{S}} \varphi$, it follows that $\vdash_{\mathrm{S}} \varphi'$, and—more importantly—it proves each individual conjunct of φ'. Since $\nvdash_{\mathrm{Ver}} \varphi$, this means that at least one of the conjuncts, which is a theorem of S is not a theorem of Ver. The question is: Which one?

It cannot be a conjunct of the form (1): The only purely propositional wffs that are theorems of S are ones that are tautologies; but these are all theorems of Ver as well.

It cannot be a conjunct of the form (2), since every wff of the form $\Box\psi$ is a theorem of Ver, and hence any disjunction containing $\Box\psi$ will also be a theorem of Ver.

If it is of the form (3), then we are done: We have shown that S has a theorem of the form $\Diamond\psi$, and hence contains D.

If it is of the form (4), by similar reasoning to the first case we know that γ cannot be valid, or the conjunct would be a theorem of Ver. Therefore, it *must* be $\Diamond\psi$ that is a theorem of S, and again, we are done. □

11.5.2 Canonical models

The completeness method given in the previous section is not generalisable, as it crucially depends upon the reduction laws in S5. In this system we introduce a general method of

proving completeness of normal modal logics, and use it to prove the completeness of K, D, T, S4, B, S5 (again), and other logics.[10] This method is the method of *canonical models*.

Recall that when we defined a notion of a *system* above (cf. Definition 11.4.5), we defined it as a set of theorems arising from some axiomatic basis. However, we've *also* seen examples where two different axiomatic bases gave rise to the same sets of theorems, making the two systems deductively equivalent. This means that, for all intents and purposes, we can simply consider systems which are deductively equivalent as being *the same system*, even though they might have arisen from different axiomatic bases. In particular, since every axiom of a system is also a theorem of the system, we can think of every system as a set of theorems that can be divided into two parts: $\mathrm{K} + \Lambda$, where Λ is the set of formulas in S that are not theorems of K. (Where S is K itself, Λ will be the empty set.) In what follows, we will use 'system' in this fashion, to refer to the entire set of theorems, without caring about the specific axiomatic basis that generated them.

A canonical model $\mathfrak{M}^{\mathrm{S}}$ for a modal logic S is a special type of model where every theorem of S is valid in $\mathfrak{M}^{\mathrm{S}}$ (that is, for $\lambda \in \mathrm{S}$ and $w \in W$, $\mathfrak{M}^{\mathrm{S}}, w \vDash \lambda$) and every non-theorem is falsified in the model (i.e., for every $\gamma \notin \mathrm{S}$, there is a $w \in W$ such that $\mathfrak{M}^{\mathrm{S}}, w \vDash \neg\gamma$). If the frame of the model is a frame for the logic, then the logic is called *canonical*. If the frame of the canonical model is a frame for the logic, then the frame is in some class of frames $\mathfrak{C}$. Since every formula valid in $\mathfrak{C}$ is valid on the canonical model $\mathfrak{M}^{\mathrm{S}}$, and the only formulas which are valid on the canonical model are those which are theorems of S. Thus, we will have that φ is valid in the canonical model of S iff $\vdash_{\mathrm{S}} \varphi$.

We now outline how to construct such a model for a system S and prove that S is canonical. There are three parts of a model, the set of possible worlds W, the accessibility relation R between worlds, and the valuation V, and we will treat each part in turn.

Worlds

When we first introduced the notion of possible worlds in the beginning of this chapter, we explicitly stated that we were not going to put any constraints on what possible worlds could be. When we construct a canonical model, we *do* impose constraints. In a canonical model, we will take as worlds sets of wff—but not just any old set of wff will count as a world. Instead, we are interested in a specific type of set of wff, where the wffs in the set mirror the wffs that are true at that world. Thus, we will be developing W and V in tandem to reflect this goal.

[10] Though note that this method is *also* not universal; there are complete normal modal logics which are *not* canonical and whose completeness cannot be proven via this method. We discuss such systems in §11.5.3.

Recall (cf. Definition 11.4.9) that for a given system S, a set of formulas Λ is S-*inconsistent* iff there is a finite $\Gamma = \{\varphi_0, \ldots, \varphi_n\} \subseteq \Lambda$ such that $\vdash_S \neg(\varphi_0 \wedge \cdots \wedge \varphi_n)$.

Definition 11.5.20 (Maximality)**.** A set of formulas Λ is *maximal* iff for every φ, either $\varphi \in \Lambda$ or $\neg\varphi \in \Lambda$.

Definition 11.5.21 (Maximal consistent sets)**.** A set of formulas Λ is a *maximal* S-*consistent set* iff it is both S-consistent and maximal.

That is, in a maximal consistent set Λ, for every formula φ, either it or its negation is included in Λ, and the addition of any further formula to Λ would allow one to prove the negation of the conjunction of a subset of Λ.

Lemma 11.5.22. *Let Λ be a maximal* S*-consistent set of formulas. Then*

1. *For all formulas φ, either $\varphi \in \Lambda$ or $\neg\varphi \in \Lambda$ and not both.*
2. *$\varphi \vee \psi \in \Lambda$ iff either $\varphi \in \Lambda$ or $\psi \in \Lambda$.*
3. *$\varphi \wedge \psi \in \Lambda$ iff both $\varphi \in \Lambda$ and $\psi \in \Lambda$.*
4. *If $\varphi \in \Lambda$ and $\varphi \to \psi \in \Lambda$ then $\psi \in \Lambda$.*

Proof.

1. Because Λ is maximal, we know that at least one of φ and $\neg\varphi$ is in Λ. If both were in Λ, then $\{\varphi, \neg\varphi\}$ would be a finite subset of Λ. But $\vdash_S \neg(\varphi \wedge \neg\varphi)$ for any φ, and Λ would be S-inconsistent, contrary to hypothesis.

2. Suppose that $\varphi \vee \psi \in \Lambda$ but that $\varphi \notin \Lambda$ and $\psi \notin \Lambda$. By (1), this implies that both $\neg\varphi \in \Lambda$ and $\neg\psi \in \Lambda$. However, this means that $\{\varphi \vee \psi, \neg\varphi, \neg\psi\}$ is a finite subset of Λ, but this subset is also S-inconsistent. Now, suppose that either $\varphi \in \Lambda$ or $\psi \in \Lambda$, but that $\varphi \vee \psi \notin \Lambda$. If $\varphi \vee \psi \notin \Lambda$, then by (1), $\neg(\varphi \vee \psi) \in \Lambda$. Again, this means that either $\{\varphi, \neg(\varphi \vee \psi)\}$ or $\{\psi, \neg(\varphi \vee \psi)\}$ is a finite subset of Λ, but both of these subsets are S-inconsistent.

3. Proven in a fashion analogous to (2).

4. Suppose that $\varphi \in \Lambda$ and $\varphi \to \psi \in \Lambda$, but that $\psi \notin \Lambda$. From (1) it follows that $\neg\psi \in \Lambda$. But then $\{\varphi, \varphi \to \psi, \neg\psi\}$ is a finite subset of Λ, and it is S-inconsistent.

□

Lemma 11.5.23. *Let Λ be a maximal* S*-consistent set of formulas. Then*

1. *If $\vdash_{\mathrm{S}} \varphi$, then $\varphi \in \Lambda$.*
2. *If $\alpha \in \Lambda$ and $\vdash_{\mathrm{S}} \alpha \to \beta$, then $\beta \in \Lambda$.*

Proof.

1. Suppose $\vdash_{\mathrm{S}} \varphi$. Then $\{\neg\varphi\}$ is S-inconsistent, since $\vdash_{\mathrm{S}} \neg\neg\varphi$. Thus, $\neg\varphi \notin \Lambda$, since Λ is S-consistent. Since Λ is maximal, it follows that $\varphi \in \Lambda$.
2. Suppose that $\alpha \in \Lambda$ and $\vdash_{\mathrm{S}} \alpha \to \beta$. By (1), since $\vdash_{\mathrm{S}} \alpha \to \beta$, it follows that $\alpha \to \beta \in \Lambda$. But then by Lemma 11.5.22(4), $\beta \in \Lambda$, as required.

□

This gives us the first part of our desiderata for worlds in the canonical model: Every maximal S-consistent set of wffs contains every theorem of S. This means that if we take as our worlds these maximal consistent sets, then if our valuations mirror the formulas that make up each world, we are assured that the resulting model will validate every theorem of S.

However, that is only part of it: We also need that every non-theorem is falsified somewhere in the model. That means that for every non-theorem ψ, there is a world in the model where ψ is false. If we identify worlds with the sets of formulas true at the world, this means there must be a world Λ in the model where $\neg\psi \in \Lambda$ (since these worlds are maximal consistent sets, it follows that $\psi \notin \Lambda$).

Now, if ψ is not a theorem of S, this means that $\{\neg\psi\}$ is S-consistent. For if $\neg\psi$ were S-inconsistent, then $\vdash_{\mathrm{S}} \neg\neg\psi$, which implies that $\vdash_{\mathrm{S}} \psi$, contra assumption. What we need, then, to show that for every non-theorem ψ there is a world in which ψ does not occur, is the following result:

Theorem 11.5.24. *Any* S*-consistent set of formulas Γ can be extended to a maximal* S*-consistent set of formulas Λ. (That is, if Γ is* S*-consistent, there is a maximal* S*-consistent set Λ such that $\Gamma \subseteq \Lambda$.)*

Proof. Let Γ be an S-consistent set. We construct a series of sets of formulas starting with Γ and ending with a maximal consistent set Λ.

Because we start from a countably infinite set of atomic propositions (cf. Definition 11.2.1), and every formula is finite in length, there will be only countably many syntactically distinct modal wffs. Therefore we can take all the wffs of modal propositional logic and

arrange them in some determinate order, e.g., $\varphi_1, \varphi_2, \varphi_3, \ldots$. We then define a sequence of sets $\Lambda_0, \Lambda_1, \Lambda_2, \ldots$ as follows:

1. $\Lambda_0 = \Gamma$.

2. $\Lambda_{n+1} = \Lambda_n \cup \{\varphi_{n+1}\}$ if that is consistent; otherwise, $\Lambda_{n+1} = \Lambda_n \cup \{\neg\varphi_{n+1}\}$.

If Λ_n is consistent then so too will Λ_{n+1}. We prove this fact contrapositively: If Λ_{n+1} is not consistent, then the inconsistency already occurred in Λ_n.

If Λ_{n+1} is not consistent, then this means that both (1) $\Lambda_n \cup \{\varphi_{n+1}\}$ is not consistent and (2) $\Lambda_n \cup \{\neg\varphi_{n+1}\}$ is not consistent. If (1) is not consistent, then there are some wff $\delta_1, \ldots \delta_i \in \Lambda_n$ such that:

$$\vdash_{\mathrm{S}} \neg(\delta_1 \wedge \cdots \wedge \delta_i \wedge \varphi_{n+1}) \tag{11.26}$$

If (2) is not consistent, then there are some wff $\gamma_1, \ldots, \gamma_j \in \Lambda_n$ such that

$$\vdash_{\mathrm{S}} \neg(\gamma_1 \wedge \cdots \wedge \gamma_j \wedge \neg\varphi_{n+1}) \tag{11.27}$$

By propositional logic, it follows that:

$$\vdash_{\mathrm{S}} (\delta_1 \wedge \cdots \wedge \delta_i) \rightarrow \neg\varphi_{n+1} \tag{11.28}$$

and

$$\vdash_{\mathrm{S}} (\gamma_1 \wedge \cdots \wedge \gamma_j) \rightarrow \varphi_{n+1} \tag{11.29}$$

Combining (11.28) and (11.29), it follows that:

$$\vdash_{\mathrm{S}} \neg(\delta_1 \wedge \cdots \wedge \delta_i \wedge \gamma_1 \wedge \cdots \wedge \gamma_j) \tag{11.30}$$

But that is just the same as saying that Λ_n is inconsistent. Thus, constructing each Λ_i in this fashion maintains consistency.

We then define Λ to be the union of all the Λ_n.

(1) Λ is consistent: For if it weren't, then there would be a finite subset of Λ that was inconsistent. But every finite subset is already contained in one of the Λ_i, and thus that Λ_i would be inconsistent, contrary to hypothesis.

(2) Λ is maximal: For if it weren't, then there would be some φ_i such that neither φ_i nor $\neg\varphi_i$ were in Λ. But our method of construction assures us that either $\varphi_i \in \Lambda_i$ or $\neg\varphi_i \in \Lambda_i$, and, since $\Lambda_i \subseteq \Lambda$, in Λ as well.

□

Corollary 11.5.25. *If* $\nvdash_{\mathrm{S}} \varphi$*, then there is a maximal* S*-consistent set containing* $\neg\varphi$*.*

Proof. If $\nvdash_{\mathrm{S}} \varphi$, then $\{\neg\varphi\}$ is consistent. For if it weren't consistent, then $\vdash_{\mathrm{S}} \neg\neg\varphi$. But then $\vdash_{\mathrm{S}} \varphi$, which contradicts the assumption that $\nvdash_{\mathrm{S}} \varphi$. Since $\{\neg\varphi\}$ is S-consistent, by Theorem 11.5.24, there is a maximal S-consistent set containing $\neg\varphi$. □

We then define W^{S} in the canonical model $\mathfrak{M}^{\mathrm{S}}$ to be the set of all maximal S-consistent sets.

Accessibility relation

Next, we must define what it means for one world (i.e., maximal consistent set) to see another world in the canonical model.

In an ordinary model, which formulas are necessary at a world w falls out of which worlds are accessible to w—if φ is true at every world accessible to w, then $\Box\varphi$ will be true at w. That is, we define accessibility and then determine necessity. In a canonical model, we go the other way around: We look at the structure of the modal sentences in a given maximal consistent set Γ and use that to determine which other sets are accessible to Γ, by using the truth conditions for necessity to define R. Informally, our goal is to define R such that $\Gamma R\Delta$ iff for every wff φ, if $\Box\varphi \in \Gamma$ then $\varphi \in \Delta$. We now define this precisely:

Definition 11.5.26. Let Λ be a set of modal wffs. We then define $\Box^{-}(\Lambda)$ to be the set of wffs φ such that $\Box\varphi \in \Lambda$, that is:

$$\Box^{-}(\Lambda) = \{\varphi : \Box\varphi \in \Lambda\}$$

(That is, $\Box^{-}(\Lambda)$ is the set of formulas φ where $\Box\varphi \in \Lambda$.)

We then define R^{S} in $\mathfrak{M}^{\mathrm{S}}$ so that $\Gamma R\Delta$ iff $\Box^{-}(\Gamma) \subseteq \Delta$.

Before we construct the final part of our canonical model, the valuation, we prove a few results about R defined in this way.

A consequence of the standard definition of truth is that if φ is true at w' and wRw', then $\Diamond\varphi$ is true at w. Another way of stating this result is that if $\Diamond\varphi$ is true at w, there must be a w' such that wRw' and φ is true at w'. The following results show an analogue of this result, namely, that if $\varphi \in \Delta$ and $\Gamma R\Delta$, then $\Diamond\varphi \in \Gamma$, i.e., if $\Diamond\varphi \in \Gamma$ then there is a Δ such that $\Gamma R\Delta$ and $\varphi \in \Delta$.

Lemma 11.5.27. *Let Λ be an* S*-consistent set containing $\neg\Box\varphi$ for some φ. Then $\Box^{-}(\Lambda) \cup \{\neg\varphi\}$ is* S*-consistent.*

Proof. We prove this by contraposition, showing that if $\Box^-(\Lambda) \cup \{\neg\varphi\}$ is S-inconsistent then so to is Λ.

Suppose that $\Box^-(\Lambda) \cup \{\neg\varphi\}$ is S-inconsistent. This means that there is some finite subset $\{\gamma_1, \ldots, \gamma_n\}$ of $\Box^-(\Lambda)$ such that:

1	$\neg(\gamma_1 \wedge \cdots \wedge \gamma_n \wedge \neg\varphi)$	by assumption
2	$(\gamma_1 \wedge \cdots \wedge \gamma_n) \rightarrow \varphi$	PL
3	$\Box(\gamma_1 \wedge \cdots \wedge \gamma_n) \rightarrow \Box\varphi$	DR1
4	$(\Box\gamma_1 \wedge \cdots \wedge \Box\gamma_n) \rightarrow \Box\varphi$	K3, SubsEq
5	$\neg(\Box\gamma_1 \wedge \cdots \wedge \Box\gamma_n \wedge \neg\Box\varphi)$	PL

But $\{\Box\gamma_1, \ldots, \Box\gamma_n, \neg\Box\varphi)\} \subseteq \Lambda$, and hence Λ is inconsistent, contra assumption. □

Corollary 11.5.28. $\Box^-(\Lambda) \cup \{\neg\varphi\}$ *can be extended to a maximal* S*-consistent set, and hence is included in a world in* W^{S}*.*

This ensures that our definition of R respects the usual truth conditions for modal formulas.

Valuation

The final step is to define V^{S} in $\mathfrak{M}^{\mathrm{S}}$. When defining W^{S} and R^{S}, we did so in a way that respects the standard truth conditions. We can therefore generate the valuation from the atoms that are included in the maximal consistent sets. That is, we define $V^{\mathrm{S}}(p, \Gamma) = T$ iff $p \in \Gamma$.

Putting all of these components together, we can define a *canonical model*:

Definition 11.5.29 (Canonical model)**.** $\mathfrak{M}^{\mathrm{S}} = \langle W^{\mathrm{S}}, R^{\mathrm{S}}, V^{\mathrm{S}}\rangle$ is a *canonical model* for a system S if

- W^{S} is the set of all maximal S-consistent sets of wff.
- $R^{\mathrm{S}} \subseteq W^{\mathrm{S}} \times W^{\mathrm{S}}$ is such that $\Delta R^{\mathrm{S}} \Gamma$ iff $\Box^-(\Delta) \subseteq \Gamma$.
- V^{S} is such that $p \in \Gamma$ iff $V^{\mathrm{S}}(p, \Gamma) = T$.

It remains to show that this atomic valuation respects the more complex formulas in Γ. That is, we need to show the following fundamental result of canonical models which proves that truth at a world and membership in a world coincide in the canonical model:

Theorem 11.5.30 (Truth and Membership)**.** *Let* $\mathfrak{M}^S = \langle W^{\mathrm{S}}, R^{\mathrm{S}}, V^{\mathrm{S}}\rangle$ *be the canonical model for* S*. Then for any wff* φ *and* $\Gamma \in W^{\mathrm{S}}$*,* $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \varphi$ *iff* $\varphi \in \Gamma$*.*

Proof. By induction on the complexity of φ. For ease of reading, we omit the superscripts in this proof.

Basis case: $\mathfrak{M}^{\mathrm{S}} \vDash p$ iff $V^{\mathrm{S}}(p, \Gamma) = T$ iff $p \in \Gamma$, by definition of V^{S}.

Inductive cases: Suppose that the result holds for arbitrary φ and ψ. Then:

- Suppose that $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \neg\varphi$. $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \neg\varphi$ iff $\mathfrak{M}^{\mathrm{S}}, \Gamma \nvDash \varphi$, by the definition of truth. But since the result holds for φ, $\mathfrak{M}^{\mathrm{S}}, \Gamma \nvDash \varphi$ iff $\varphi \notin \Gamma$. Since Γ is maximal and consistent, $\varphi \notin \Gamma$ iff $\neg\varphi \in \Gamma$. So $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \neg\varphi$ iff $\neg\varphi \in \Gamma$.

- Suppose that $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \varphi \wedge \psi$. $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \varphi \wedge \psi$ iff $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \varphi$ and $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \psi$. By hypothesis, $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \varphi$ iff $\varphi \in \Gamma$ and $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \psi$ iff $\psi \in \Gamma$. By Lemma 11.5.22, $\varphi \in \Gamma$ and $\psi \in \Gamma$ iff $\varphi \wedge \psi \in \Gamma$, as required. The cases of $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \varphi \vee \psi$ and $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \varphi \rightarrow \psi$ are similar.

- Suppose that $\Box\varphi \in \Gamma$. By definition of R, we know that $\varphi \in \Delta$ for every Δ such that $\Gamma R \Delta$. By hypothesis, it thus follows that $\mathfrak{M}^{\mathrm{S}}, \Delta \vDash$ for every Δ such that $\Gamma R \Delta$. It therefore follows that $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \Box\varphi$, by the definition of truth.

  Now suppose that $\Box\varphi \notin \Gamma$. Because Γ is maximal, it follows that $\neg\Box\varphi \in \Gamma$. By Lemma 11.5.27, we know that $\Box^{-}(\Gamma) \cup \{\neg\varphi\}$ is consistent, and hence is contained in some maximal consistent set Λ. But then by definition of R, $\Gamma R \Lambda$. By hypothesis, since $\neg\varphi \in \Lambda$, $\mathfrak{M}^{\mathrm{S}}, \Lambda \nvDash \varphi$. It then follows that $\mathfrak{M}^{\mathrm{S}}, \Gamma \nvDash \Box\varphi$ as well.

  The case of $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \Diamond\varphi$ is similar.

□

Corollary 11.5.31. *$\mathfrak{M}^{\mathrm{S}} \vDash \varphi$ iff $\vdash_{\mathrm{S}} \varphi$.*

Proof. ($\Leftarrow$) By Lemma 11.5.23, if $\vdash_{\mathrm{S}} \varphi$, then every maximal S-consistent set Λ contains φ. By Theorem 11.5.30, if $\varphi \in \Lambda$, then $\mathfrak{M}^{\mathrm{S}}, \Lambda \vDash \varphi$; but this is true for every Λ, and hence $\mathfrak{M}^{\mathrm{S}} \vDash \varphi$.

($\Rightarrow$) If $\nvdash^{\mathrm{S}} \varphi$, then $\{\neg\varphi\}$ is S-consistent (see proof of Lemma 11.5.25). By Theorem 11.5.24, there is a maximal S-consistent set Λ containing $\neg\varphi$. By Theorem 11.5.30, $\Lambda \vDash \neg\varphi$ and hence $\Lambda \nvDash \varphi$, so it follows that $\mathfrak{M}^{\mathrm{S}} \nvDash \varphi$. □

Proving canonicity

Thus, we have now defined the canonical model $\mathfrak{M}^{\mathrm{S}}$ for a system S in such a way that (a) every theorem of S is valid in $\mathfrak{M}^{\mathrm{S}}$ and (b) no non-theorem of S is valid in $\mathfrak{M}^{\mathrm{S}}$. We now sketch how these canonical models can be used to prove completeness.

To show that a logic S is complete with respect to a class of frames $\mathfrak{C}$, we must show that if $\mathfrak{C} \vDash \varphi$, then $\vdash_{\mathrm{S}} \varphi$. Our method proceeds contrapositively: We show that if $\nvdash_{\mathrm{S}} \varphi$, then $\mathfrak{C} \nvDash \varphi$ — that is, there is a frame $\mathfrak{F} \in \mathfrak{C}$, and a model $\mathfrak{M}$ on $\mathfrak{F}$ and a world w in $\mathfrak{M}$ such that $\mathfrak{M}, w \nvDash \varphi$. The proofs of completeness that we discuss in this section rely crucially on the canonical models we constructed. These canonical models are special because of what we noted above — the only things valid on the model are things which are theorems of S — if something is not a theorem of S, then there is a maximal S-consistent set that contains its negation, and that set is one of the worlds in the model, and the truth-and-membership lemma allows us to get from membership in a set to truth at that world. So, from the fact that $\nvdash_{\mathrm{S}} \varphi$ we get that there is a world Γ of the canonical model such that $\mathfrak{M}^{\mathrm{S}}, \Gamma \vDash \neg\varphi$, and hence $\mathfrak{M}^{\mathrm{S}}, \Gamma \nvDash \varphi$.

But this just shows us that we have a *model* of a particular type. What we need to show completeness is that everything valid on a *frame* in the relevant class of frames is a theorem — or, to put it another way, every *non*-theorem is falsified in some world in some model on some frame in the class.

Now, we've got a model (the canonical model) where every non-theorem is falsified in some world. *If* we can show that the frame for $\mathfrak{M}^{\mathrm{S}}$ is a frame for S, then we have shown that no non-theorem φ of S is valid in the class of frames $\mathfrak{C}^{\mathrm{S}}$ for S, because we can always find a model based on a frame in $\mathfrak{C}^{\mathrm{S}}$ that has a world where φ is false (namely, the canonical model).

So, we need to show that the frame of the canonical model for S is a member of the class of frames characterising S; once we have shown that, then we have shown that every non-theorem of S can be falsified in a model on a frame for S, and hence cannot be valid in the class of frames for S.

Where S is K, the proof is easy:

Theorem 11.5.32. *The frame $\mathfrak{F}^{\mathrm{K}}$ of the canonical model $\mathfrak{M}^{\mathrm{K}}$ of the system* K *is member of the class of* K*-frames.*

Proof. The class of K-frames is the class of all frames; since every frame is a frame, the frame $\mathfrak{F}^{\mathrm{K}}$ is a frame for K. □

When a class of frames is characterised by a particular frame condition, we show that the frame of $\mathfrak{M}^{\mathrm{S}}$ has the relevent frame condition. We consider two cases: Classes of frames defined only by universal conditions and classes of frames defined by at least one existential condition.

Classes of frames with universal conditions (such as reflexivity, transitivity, symmetry, etc.) are straightforward, and all can be proven in a similar fashion.

Theorem 11.5.33. *The frame* $\mathfrak{F}^{\mathrm{T}}$ *of the canonical model* $\mathfrak{M}^{\mathrm{T}}$ *of the system* T *is a frame for* T*.*

Proof. T is characterised by the class of reflexive frames, i.e., where $\forall x(xRx)$. To show that $\mathfrak{F}^{\mathrm{T}}$ is reflexive, we must show that $\forall\Gamma(\Gamma R\Gamma)$. By the definition of R^{T}, we need to show that $\Box^{-}(\Gamma) \subseteq \Gamma$, that is, if $\Box\varphi \in \Gamma$ then $\varphi \in \Gamma$.

Assume that $\Box\varphi \in \Gamma$. By Lemma 11.5.23(1), $\Box\varphi \to \varphi \in \Gamma$; but then by Lemma 11.5.22(4), $\varphi \in \Gamma$, which is what we needed to show. □

Corollary 11.5.34. *The logic* T *is complete with respect to the class of reflexive frames.*

Theorem 11.5.35. *The frame* $\mathfrak{F}^{\mathrm{S4}}$ *of the canonical model* $\mathfrak{M}^{\mathrm{S4}}$ *of the system* S4 *is a frame for* S4*.*

Proof. S4 is characterised by the class of reflexive and transitive frames, i.e., where $\forall x(xRx)$ and $\forall x\forall y\forall z((xRy \wedge yRz) \to xRz)$. Since $\vdash_{\mathrm{S4}} \mathsf{T}$, the previous results apply to S4 as well, so we have already proven that $\mathfrak{F}^{\mathrm{S4}}$ is reflexive. To show that $\mathfrak{F}^{\mathrm{S4}}$ is transitive, we must show that $\forall\Gamma\Delta\Lambda((\Gamma R\Delta \wedge \Delta R\Lambda) \to \Gamma R\Lambda)$. By the definition of R^{S4}, we need to show that if $\Box^{-}(\Gamma) \subseteq \Delta$ and $\Box^{-}(\Delta) \subseteq \Lambda$, then $\Box^{-}(\Gamma) \subseteq \Lambda$). Assume (1) if $\Box\varphi \in \Gamma$ then $\varphi \in \Delta$ and (2) if $\Box\varphi \in \Delta$ then $\varphi \in \Lambda$. We now assume (3) that $\Box\varphi \in \Gamma$ and show that $\varphi \in \Lambda$.

By Lemma 11.5.23(1), $\Box\varphi \to \Box\Box\varphi \in \Gamma$, and by Lemma 11.5.22(4), $\Box\Box\varphi \in \Gamma$. By (1), it follows that $\Box\varphi \in \Delta$, and by (2) that $\varphi \in \Lambda$, which is what we needed to show. □

Corollary 11.5.36. *The logic* S4 *is complete with respect to the class of reflexive and transitive frames.*

Theorem 11.5.37. *The frame* $\mathfrak{F}^{\mathrm{B}}$ *of the canonical model* $\mathfrak{M}^{\mathrm{B}}$ *of the system* B *is a frame for* B*.*

Proof. B is characterised by the class of symmetric frames, i.e., where $\forall xy(xRy \to yRx)$. To show that $\mathfrak{F}^{\mathrm{B}}$ is symmetric, we must show that $\forall\Gamma\Delta(\Gamma R\Delta \to \Delta R\Gamma)$. By the definition of R^{B}, we need to show that if $\Box^{-}(\Gamma) \subseteq \Delta$ then $\Box^{-}(\Delta) \subseteq \Gamma$, that is, if (1) if $\Box\varphi \in \Gamma$ then $\varphi \in \Delta$, then (2) if $\Box\varphi \in \Delta$, then $\varphi \in \Gamma$. Assume (1) and (3) that $\Box\varphi \in \Delta$. Suppose finally (4) that $\varphi \notin \Gamma$. By maximality, $\neg\varphi \in \Gamma$. By Lemma 11.5.23(1), $\neg\varphi \to \Box\Diamond\neg\varphi \in \Gamma$, and so it follows that $\neg\varphi \to \Box\neg\Box\varphi \in \Gamma$. By Lemma 11.5.22(4), we have that $\Box\neg\Box\varphi \in \Gamma$. By (1), $\neg\Box\varphi \in \Delta$, and so it follows that $\Box\varphi \notin \Delta$; but this contradicts (3). This completes the proof. □

Corollary 11.5.38. *The logic* B *is complete with respect to the class of symmetric frames.*

Theorem 11.5.39. *The frame* $\mathfrak{F}^{\mathrm{S5}}$ *of the canonical model* $\mathfrak{M}^{\mathrm{S5}}$ *of the system* S5 *is a frame for* S5*.*

Proof. Immediate from the preceding results. □

Corollary 11.5.40. *The logic* S5 *is complete with respect to the class of reflexive, transitive, and symmetric frames.*

Classes of frames where the frame condition has an existential quantifier requires a bit more work, because they require us to show that a particular type of set exists (in order to show that the required possible world, related to the other worlds in the required way, exists). Sometimes the existence of the relevant possible world can be shown directly; in other cases, the world must be constructed by Theorem 11.5.24.

We give examples of both of these options.

Theorem 11.5.41. *The frame* $\mathfrak{F}^{\mathrm{D}}$ *of the canonical model* $\mathfrak{M}^{\mathrm{D}}$ *of the system* D *is a frame for* D*.*

Proof. D is characterised by the class of serial frames, i.e., where $\forall x \exists y(xRy)$. To show that $\mathfrak{F}^{\mathrm{D}}$ is serial, we must pick an arbitrary Γ and demonstrate the existence of a Δ which is such that $\Gamma R \Delta$.

First, note that $\vdash_{\mathrm{D}} \Diamond(p \vee \neg p)$, so by Lemma 11.5.23(1), $\Diamond(p \vee \neg p) \in \Gamma$. It follows from the Truth and Membership Theorem (11.5.30) that $\mathfrak{M}^{\mathrm{D}}, \Gamma \vDash \Diamond(p \vee \neg p)$, and hence, by the definition of truth, that there is a Δ such that $\Gamma R \Delta$ and $\mathfrak{M}^{\mathrm{D}}, \Delta \vDash p \vee \neg p$, i.e., $p \vee \neg p \in \Delta$. Thus, we have proven that $\mathfrak{F}^{\mathrm{D}}$ is serial. □

Corollary 11.5.42. *The logic* D *is complete with respect to the class of serial frames.*

Consider the class $\mathfrak{C}$ of convergent frames (cf. Definition 11.3.35). The property of convergence corresponds to the modal axiom $\mathsf{Geach} \coloneqq \Diamond \Box p \to \Box \Diamond p$ (cf. Lemma 11.3.44). We consider the canonical model $\mathfrak{M}^{\mathrm{C}}$ for the system C. We want to show that the frame of the canonical model is a frame for C, that is, it is convergent.

Theorem 11.5.43. *The frame* $\mathfrak{F}^{\mathrm{C}}$ *of the canonical model* $\mathfrak{M}^{\mathrm{C}}$ *of the system* C *is convergent.*

Proof. Assume that there are maximal C-consistent sets Δ, Γ, Γ' such that $\Delta R \Gamma$ and $\Delta R \Gamma'$. We need to prove the existence of a maximal C-consistent set Λ that is such that $\Gamma R \Lambda$ and $\Gamma' R \Lambda$. If there is such a set Λ, then by the definition of R^{C}, it will contain $\{\alpha \colon \Box\alpha \in \Gamma\} \cup \{\beta \colon \Box\beta \in \Gamma'\}$. If this set (call it Λ) is consistent, then it will be contained in a maximal C-consistent set, and we will be done. So we need to prove its consistency.

Assume that Λ is C-inconsistent. Then there are $\Box\alpha_1, \ldots, \Box\alpha_n \in \Gamma$ and $\Box\beta_1, \ldots, \Box\beta_m \in \Gamma'$ such that $\vdash_{\mathrm{C}} \neg(\alpha_1 \wedge \cdots \wedge \alpha_n \wedge \beta_1 \wedge \cdots \wedge \beta_m)$. Let $\alpha \coloneqq \alpha_1 \wedge \cdots \wedge \alpha_n$ and $\beta \coloneqq \beta_1 \wedge \cdots \wedge \beta_m$.

Then we can rewrite this as $\vdash_{\mathrm{C}} \neg(\alpha \wedge \beta)$. By propositional logic, this is the same as $\vdash_{\mathrm{C}} \alpha \rightarrow \neg\beta$. By basic modal logic, we can prove $\vdash_{\mathrm{C}} \Diamond\alpha \rightarrow \Diamond\neg\beta$, and again by basic modal logic, $\vdash_{\mathrm{C}} \Diamond\alpha \rightarrow \neg \Box \beta$. Now, note that if $\Delta R\Gamma$ and $\Box\alpha \in \Gamma$, then $\Diamond \Box \alpha \in \Delta$. By the characteristic axiom C, it follows that $\Box \Diamond \alpha \in \Delta$, and since $\Delta R\Gamma'$, $\Diamond\alpha \in \Gamma'$. But then $\neg \Box \beta \in \Gamma'$, which contradicts our earlier assumption about Γ'. Thus, Λ is in fact consistent, and by Lindenbaum's lemma it is contained in a maximal C-consistent set, Δ', which is the set we needed.

Therefore, the frame for the canonical model of C is a C-frame. □

Corollary 11.5.44. *The system* K+Geach *is complete with respect to the class of convergent frames.*

It should be noted that these proofs all follow roughly the same sort of structure. We outline the structure here, and recommend readers go back through each of the proofs to see how the structure is implemented, and to use this as scaffolding for their own proofs (cf. Exercises 124 and 125):

1. Rewrite the frame condition in terms of Γ, Δ, Λ, etc.
2. Use the definition of R^{S} to rewrite in terms of $\Box^{-}(\Gamma)$, etc.
3. Rewrite the $\Box^{-}(\Gamma)$ clauses either in set theoretic terms or in terms of hypotheticals involving set membership (e.g., "if $\Box\varphi \in \Gamma$ then $\varphi \in \Delta$".)
4. Make the appropriate assumptions (including assumptions for reductio).
5. Invoke the characteristic axiom or a substitution instance of it.
6. Reach desired outcome.

Not every proof will involve every one of these steps, and not every step in a proof will be one of these, but this reflects the overall structure that many of these types of proofs have, and provides a bit of a guide when developing your own proofs.

11.5.3 Logics which are not canonical

Not every modal logic which is complete can be proven to be complete via the canonical models method. A logic is said to be 'canonical' if the frame of the canonical model is a frame for the logic (for instance, we saw in Theorem 11.5.33 that T is canonical, because the frame of $\mathfrak{M}^{\mathrm{T}}$ is reflexive). But while every canonical model will always be a *model* for the relevant system, it doesn't follow that the *frame* of the canonical model will always be a

frame for the relevant system. In this section, we investigate logics which are *not* canonical, that is, where the frame of the canonical model for the system is not a frame for the system itself.

How can this be? It is sometimes easy to forget that one can have many different models built on the same frame; in much of what we have discussed in the preceding sections, we have been interested in *arbitrary* models on frames that meet some particular constraint (are a member of some class of frames). When we work with canonical models, however, the focus is less on the frame and more on the valuation: In particular, we define the canonical valuation precisely so that truth and membership coincide (cf. Lemma 11.5.30). But every canonical *model* is based on a *frame*, and once we have that frame, we can ask "What would happen if I changed up the valuation?" Indeed, for any given frame, there will be infinitely many models that we could define on that frame—the canonical model; the "anti-"canonical model that makes an atom *false* at some world if it is a member of the world; the model that orders all of the atoms and makes all of the even numbered atoms true at every world and the odd numbered ones false at every world; the model that orders all of the worlds and makes every atom false at every even numbered world and every atom true at every odd numbered world; etc., etc., etc. In all cases except for the first two, knowing something about which wffs are members of a given world would say nothing about what wffs are *true* at that world. This gives us a general question to ask: If we keep the frame of a canonical model fixed, and changed the valuation of the atoms, can we falsify the modal axioms of the system?

In the case of canonical systems, because the frame of the canonical model is a frame for the system, changing the valuation won't affect the validity of the modal axiom—our correspondence results from Section 11.3.1 prove this for a number of axiom/condition pairs. But in non-canonical systems, because the frame of the canonical model is not necessarily a frame for the logic, then since the frame condition corresponding to the axiom fails to obtain, we know that it will always be possible to define a model that makes the axiom false. (We just need *one* model for this; the presence of other models, such as the canonical model, that validate the axiom is not a contradiction!)

In the remainder of this section, we look at a variety of different canonical models for various systems, some of which will turn out to be non-canonical systems, that is, systems where the frame of the canonical model is not a frame for the logic. Some of these systems are non-canonical because they will turn out to be incomplete; but at least one of these systems is, however, complete—and we will show how completeness can be proven via methods other than the canonical-models method of the previous section.

The first is the system G.[11] Consider the following wff:

$$\Box(\Box p \to p) \to \Box p \tag{G}$$

If we read $\Box$ here as generic necessity, then it reads: "If it is necessary that necessarily-p implies p, then p is necessary". As a generic modal principle, it is not generally going to be appropriate to adopt. However, there is a way of reading the $\Box$—in terms of provability in a formal system—that renders the principle plausible:

> If you can prove that provability implies truth, then you can prove anything.

This wff is called 'G' in reference to Gödel, because under this provability interpretation of $\Box$, the wff expresses his incompleteness theorem for Peano Arithmetic (cf. Chapter 10). For more on logics of provability, see [Japaridze and de Jongh, 1998, Verbrugge, 2017] and §12.5.

Adding this wff to K (to give us the logic G) results in a consistent system, because G can be validated on a K-frame (and hence is not contradictory), but unlike the systems we saw in the previous section, the canonical model for G is not based on a frame for the logic, and hence G is non-canonical. Once we have demonstrated this, we will then show that, nevertheless, G is complete with respect to $\mathfrak{C}^{\mathrm{G}}$.

We begin with some definitions. First, we introduce the notion of an R-chain:

Definition 11.5.45 (R-chain)**.** Let $\mathfrak{F} = \langle W, R\rangle$ be a frame. An R-chain in $\mathfrak{F}$ is defined inductively:

- For all worlds $w, w' \in W$, wR^0w' iff $w = w'$, and we say that "w is "0 R-steps from itself" and that "there is a 0 R-chain from w to w". (Note that wR^0w even if $\neg wRw$.)
- For all worlds $w, w' \in W$, $wR^{n+1}w'$ iff there is a w'' such that wR^nw'' and $w''Rw'$, and we say that "there is an $n+1$ R-chain from w to w'".

Different frames will have R-chains of different lengths, depending on how R is defined in the frame. However, we can note a few general facts about R-chains, correlated to different frame properties.

Corollary 11.5.46. *Every reflexive frame contains an R^n-chain for every finite n.*

Proof. This is because every world is 0 R-steps away from itself, and then you can simply loop around the reflexive arrow as many times as you want, until you reach n R-steps. □

[11] The system is called W by Segerberg [Segerberg, 1971, p. 84].

Corollary 11.5.47. *In every transitive frame, if there is an R^n-chain between two worlds, there is an R^1-chain between them.*

Proof. Follows directly from the definition of transitivity. □

R-chains are useful because they allow us to be able to talk about paths through frames. If we expand our notion of R-chain to include "backward" paths (e.g., by changing the second clause of the definition to be "$w''Rw'$ or $w'Rw''$") as well as "forward" ones, we can then identify frames that are *cohesive*, in the sense that every world is connected to every other world by an R-chain of some length.

Definition 11.5.48 (Cohesive)**.** A frame $\mathfrak{F} = \langle W, R\rangle$ is *cohesive* if for every $w, w' \in W$, there is an R^n-chain between w and w', for some n.

We've already seen examples of noncohesive frames above; in the frame depicted in Figure 11.14, the ws are connected by an R-chain, and so are the xs, but there is no R-chain between the xs and the ws.

The notion of cohesiveness is important to our investigations here because many frames of canonical models are *not* cohesive [Hughes and Cresswell, 1996, pp. 137–138]. But equally there are canonical models whose frames *are* cohesive—and not only that, but some canonical models with cohesive frames are cohesive because of a single world, a "root" world which sees every world. In fact, we can precisely characterise which canonical models will have frames that have a root world like this.

Lemma 11.5.49. *Let* S *be a consistent modal system extending* K*. If $\{\neg \Box \varphi : \nvdash_{\mathrm{S}} \varphi\}$ is* S*-consistent, then the canonical model $\mathfrak{M}^{\mathrm{S}}$ contains a world Γ^* such that Γ^* sees every world in the model.*

Proof. If $\{\neg \Box \varphi : \nvdash \varphi\}$ is S-consistent, then it can be extended to a maximal S-consistent set Γ^* where $\Box\varphi \in \Gamma^*$ iff $\vdash_{\mathrm{S}} \varphi$. But then $\Box^-(\Gamma^*) = \{\varphi : \vdash_{\mathrm{S}} \varphi\}$, and hence $\Box^-(\Gamma^*) \subseteq \Delta$ for every maximal S-consistent set Δ (including Γ^* itself), so $\Gamma^* R \Delta$. □

Before we prove that the frame of the canonical model for G contains such a world (a world that sees every world), we prove one final interim result.

Lemma 11.5.50. *Let* S *be a consistent modal system extending* K*, and $\mathfrak{M}^{\mathrm{S}} = \langle W^{\mathrm{S}}, R^{\mathrm{S}}, V^{\mathrm{S}}\rangle$ be the canonical model of* S*. Define a new model for* S*, call it $\mathfrak{M}^+ = \langle W^+, R^+, V^+\rangle$ containing a world w^+, as follows:*

- *Choose some $w^+ \notin W^{\mathrm{S}}$ and let $W^+ = W^{\mathrm{S}} \cup \{w^+\}$.*

- *Let* $R^+ = R^S \cup \{\langle w^+, w\rangle : w \in W^S\}$.
- *For* $w \in W^S$, *let* $V^+(p, w) = V^S(p, w)$, *and for* w^+, *let* $V^+(p, w^+)$ *be arbitrary.*

Then, if $V^+(\Box\varphi, w^+) = T$, *then* $\vdash_S \varphi$.

Proof. Assume that $V^+(\Box\varphi, w^+) = T$. Then $V^+(\varphi, w) = T$ for every w such that w^+Rw, i.e., for every $w \in W^S$. Since V^S and V^+ agree on all worlds in W^S, it follows that $V^S(\varphi, w) = T$ for all $w \in W^S$, and hence $\mathfrak{M}^S \vDash \varphi$. By Corollary 11.5.31, it follows that $\vdash_S \varphi$, which is what we needed to show. □

Furthermore, since $\mathfrak{M}^+$ is such that if $V^+(\Box\varphi, w^+) = T$, then $\vdash_S \varphi$, it follows that $\{\neg\Box\varphi : \nvdash_S \varphi\}$ is S-consistent. Since it is S-consistent, per Theorem 11.5.24, it is contained in a maximal S-consistent set Λ, and hence this world is already a part of the original canonical model, and by Lemma 11.5.49, it sees every world in the canonical model, including itself.

We now apply the construction in the previous lemma, and the concommittent results, to system G.

Lemma 11.5.51. *Let* $\mathfrak{M}^G = \langle W^G, R^G, V^G\rangle$ *be the canonical model of* G. *Let* $\mathfrak{M}^{G+} = \langle W^{G+}, R^{G+}, V^{G+}\rangle$ *be defined as in Lemma 11.5.50. Then,* $\mathfrak{M}^{G+}$ *is a model of* G.

Proof. By definition, $V^G(\Box(\Box p \to p) \to \Box p, w) = T$ for all $w \in W^G$, and hence $V^+(\Box(\Box p \to p) \to \Box p, w) = T$ for all $w \in W^G$, so all we need to show is that $V^+(\Box(\Box p \to p) \to \Box p, w^+) = T$ in order to prove the desired result.

What we will show is that if $V^+(\Box(\Box p \to p), w^+) = T$, then $V^+(\Box p, w^+) = T$. First, assume that $V^+(\Box(\Box p \to p), w^+) = T$. By Lemma 11.5.50, it follows that $\vdash_G \Box p \to p$. By Nec, we have $\vdash_G \Box(\Box p \to p)$; by MP and G, it follows that $\vdash_G \Box p$. Since we've already shown that $\vdash_G \Box p \to p$, by MP again we have $\vdash_G p$, and hence $V^G(p, w) = T$ for all $w \in W^G$. By the definition of V^+, it then follows that $V^+(\Box p, w^+) = T$, which is what we needed to show. □

Corollary 11.5.52. *There is a world in* $\mathfrak{M}^G$ *that sees every world, including itself, and hence there is a reflexive world in* $\mathfrak{M}^G$.

We are now in a position to prove that G is not canonical. Per the immediately preceding corollary, the canonical model for G contains a reflexive world. We now show that no frame with a reflexive world in it is a frame for G, and hence that the frame of the canonical model is a not a frame for the logic.

Lemma 11.5.53. *Let* $\mathfrak{F}$ *be a frame with a reflexive world. Then,* $\mathfrak{F} \nvDash \mathsf{G}$.

Proof. Let $\mathfrak{F}$ be such a frame and w^* such a world, that is (1) w^*Rw^*. Define a valuation V on $\mathfrak{F}$ in which (2) $V(p, w^*) = F$ and (3) $V(p, w) = T$ for every $w \neq w^*$. By (1) and (2), it follows that (4) $V(\Box p, w^*) = F$, and hence (5) $V(\Box p \to p, w^*) = T$. It follows from (3) that (6) $V(\Box p \to p, w) = T$ for all $w \neq w^*$, and from (5) and (6) that (7) $V(\Box p \to p, w) = T$ for all $w \in W$. Consequently, (8) $V(\Box(\Box p \to p), w^*) = T$. But then the antecedent of G is true at w^* by (8), and the consequent of G is false at w^* by (2), so the entire axiom is false. □

Theorem 11.5.54. G *is not canonical.*

Proof. Immediate from the previous result; the frame of the canonical model $\mathfrak{M}^{\mathsf{G}}$ is not a frame for the logic. □

Note that we have shown that $\mathfrak{M}^{\mathsf{G}}$ is not a frame for the logic *without ever having yet said what class of frames (if any) characterizes* G. At this point, the keen reader will hopefully be wondering, just what class of frames (if any) characterizes G? What we've seen so far is that whatever the class is, it cannot contain any reflexive world. In fact, if we generalize on this, by looking at what the presence of a reflexive world means in terms of R-chains in the frame, we can completely characterise class of frames that corresponds to G.

Lemma 11.5.55. *The logic* G *is characterized by the class of frames where R contains no infinitely descending chains.*

Proof. I suspect that we won't prove this until Section 12.5. □

Lemma 11.5.56. *The logic* G *is complete with respect to the class of frames where R contains no infinitely descending chains.*

Proof. ditto. □

11.6 Strange modal logics

A common question, upon being introduced to S4 and S5, is "where are the other S systems?" In this section, we briefly discuss some modal logic oddities, including the logic S0.5.

S0.5 comprises all propositional tautologies, and modal axioms K and T, the rule MP, and the rule:

Rule 11.6.1 (Rule of Tautological Necessitation (TN))**.** If φ is a propositional tautology, then $\Box\varphi$ is a theorem.

A rather strange logic, no doubt; but Lemmon argues that it "can very suitably be interpreted as a formalized metalogic of the propositional calculus" [Lemmon, 1957, p. 181], where $\Box$ is interpreted as "it is tautologous by truth-table that". The modal axioms then capture the facts that "being a tautology" is preserved under modus ponens, and that all tautologies are true.

Another interesting modal logic is Carnap's modal logic C, which contains the axiom $\diamond\varphi$ where φ is a consistent propositional formula. Like many of the other logics discussed in this section, it is a non-normal logic.

Chapter 12

Modal logic: applications [last modified 19 Jan 21]

In the previous chapter, we discussed the *theory* of modal logic, as the logic of abstract relational structures, and showed how to prove soundness and completeness results for different modal systems. In this chapter, we look at how these systems can be applied to particular philosophical problems, such as knowledge, action, and obligation. (Another application, that of time and tense, is given its own separate chapter, Chapter 13.)

12.1 Epistemic logic

Epistemic logic is the logic of epistemic notions such as knowledge and belief. There are two ways that $\Box$ can be given an epistemic interpretation, and in both cases, we often use K rather than $\Box$. Either $K\varphi$ can be read 'it is known that φ' or 'φ is known', without specifying *who* it is that is doing the knowing; or, since knowledge is always known *by someone*, the modality can be indexed to the name of a person, and hence the knowledge can be indexed to the name of a person: $K_a\varphi$ is read 'a knows that φ'. This latter route allows us to express complex statements such as 'a knows that b knows φ': $K_aK_b\varphi$.

The field owes its origin to [Hintikka, 1962], who was the first to systematically exploit the grammatical and semantic parallels between the operators 'it is necessary that' and 'it is known that'. If necessity is truth in all relevant possible worlds (e.g., all logically possible worlds, all physically possible worlds, etc.), knowledge, then, is truth in all epistemically possible worlds.

What is an 'epistemically-possible world'? They are all the worlds that are consistent with my knowledge; for example, if I know that $2 + 2 = 4$, then any world where $2 + 2 = 5$ is

not epistemically possible. Or, if I know that it is raining today, then any world where it is not raining today is not epistemically possible. Or, to look at it another way, knowledge is simply *defined* to be everything that is true in any world that I cannot distinguish from the actual world, that is, any world which, "for all I know", could be the actual world. This relation between worlds, of 'epistemic indistinguishability', is the relation we represent by R in our models, and it is an *equivalence relation* (cf. Definition 11.3.24.) That is: no world can be distinguished from itself; if x cannot be distinguished from y, then y cannot be distinguished from x; and if x cannot be distinguished from y, and y cannot be distinguished from z, then x cannot be distinguished from z either. From the fact that epistemic indistinguishability is an equivalence relation, we know that epistemic logic will coincide with the logic S5 (cf. Definition 11.4.72). Thus, we know a number of facts about the modal operator K:

- K is *factive*: That is, $K\varphi \to \varphi$. This follows from the fact that every equivalence relation is reflexive, and implies that whatever is known is true.

- K distributes over conjunction: That is, $K(\varphi \wedge \psi) \to (K\varphi \wedge K\psi)$. This is because K is a normal modal operator. (Using Fitch's terminology, K is closed with respect to conjunction elimination [Fitch, 1963, p. 136].)

12.1.1 Fitch's paradox

Fitch's (epistemic) paradox arises from the combination of two rather innocuous theses and the above principles of epistemic logic. The two theses are:

1. "Every truth is (in principle) knowable" (the *Knowability Thesis*)

2. "Not all truths are known".

If knowledge is factive and closed under conjunction, then these two theses cannot both be maintained (and that is the paradox).

Proof. Suppose that φ is true, but not known to be true. That is, assume:

$$\varphi \wedge \neg K\varphi \tag{12.1}$$

This sentence is true, and by the Knowability Thesis, it is possible that it is known:

$$\Diamond K(\varphi \wedge \neg K\varphi) \tag{12.2}$$

Because K distributes over conjunction, this implies:

$$\Diamond(K\varphi \wedge K\neg K\varphi) \tag{12.3}$$

Since K is factive, this implies:

$$\Diamond(K\varphi \wedge \neg K\varphi) \tag{12.4}$$

But no contradiction is possible:

$$\Box\neg(K\varphi \wedge \neg K\varphi) \tag{12.5}$$

And this is a contradiction. Thus, either the Knowability Thesis must be rejected; or we must reject the idea of there being truths we don't know; or we must reject factivity; or we must deny that knowledge distributes over conjunction. □

For more information on the paradox, and succeeding attempts to mitigate it, see [Brogaard and Salerno, 2013].

12.1.2 Epistemic uncertainty

One of the crucial ways that epistemically-possible worlds differ from, e.g., logically-possible worlds is that the latter are fixed. We do not get to dictate which worlds are logically possible.

However, because different people have different knowledge, the set of worlds that are epistemically possible worlds differs relative to the knowledge of each individual.[1] If there are multiple epistemically-possible worlds for a given agent, given their current information, then the agent is said to be in a state of *epistemic uncertainty*.

Further, this set of epistemically-possible worlds can change as the agent is given new information. For example, Walking home from nursery, a child asks, "What are we having for supper?" Her mother does not know if her husband has been to the store or not, so she does not know if they are having leftovers or if he has cooked. Each of these worlds—the world in which he went to the store and the world in which he didn't—is consistent with her knowledge, so each of these worlds is epistemically possible, and she don't know which is the actual world. Both are in a state of epistemic uncertainty until they arrive home, at which time he announces "I didn't have time to go to the store today". Now, with this new information that wasn't available after before, her epistemic uncertainty is removed. She knows they're in the "we're going to have leftovers for supper world". (For more discussion, see §16.)

[1] We will make the agent-dependence of knowledge more explicit earlier in this chapter in a later revision.

12.1.3 Justification

One question you might have is what the relationship between epistemic logic and epistemology is—in particular, where has the notion of 'justification' gone? This is a question that has bothered many epistemic logicians too, such that they have developed so-called Evidence Logics or Justification Logics. If you want references to these, email Sara.

12.2 Doxastic logic

Doxastic logic (the logic of belief) crucially differs from epistemic logic in that belief is not generally factive, and so does not validate T—not only that, but if we admit that it is possible to believe inconsistent propositions[2] then belief doesn't even validate D! What principles, then, are validated? We shall see, in the course of this section.

Doxastic logic can be built directly from propositional logic, or the language of epistemic logic can be extended with the addition of a new doxastic operator, B, which may or may not be indexed by agents in a set of agents. We shall, throughout, consider the index version of this operator, since anything that is said of the nonindexed operator will be true of the indexed operator when the index set contains only one agent.

Definition 12.2.1. A *language of belief* $\mathcal{L}_B$ is any propositional or epistemic language extended with a family of operators B_a for $a \in$ index set A.

We denote the dual of B_a with $\hat{B}_a$.

If we stipulate, following fn. 2, that no one can believe a contradiction, then this is equivalent to accepting the following principle:

$$\neg B_a \bot \tag{Bel}$$

(Bel) is itself equivalent to the doxastic version of D:

$$B_a p \to \hat{B}_a p \tag{D-Bel}$$

[2]Note that this is not taken generally as possible; quite often, it is assumed that even though people's belief sets might include false beliefs, they'll never include contradictory ones [Rendsvig and Symons, 2019, §2.6].

12.3 Agentive and deontic logic

12.3.1 Agency as a modal notion

Agency is a rather new addition to the "cannon" of modal notions — modernly at least: It was only in the 1980s that the idea of agency or doing or 'seeing to it that' was recognized as modal in nature.

What makes something modal?

- Previous answer: Modalities are propositional operators.

But: Are all propositional operators modal?

- Another answer: Koslow (structuralist approach to logic [Koslow, 1992, Koslow, 1999]) offers the idea that maybe it is satisfaction of the Square of Opposition (cf. Figure 12.1).

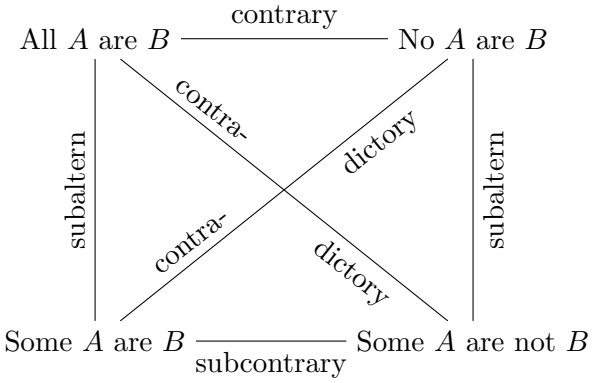


Figure 12.1: Aristotelian square of opposition

Alethic modalities can be fit naturally into a Square of Opposition (cf. Figure 12.2): Note, this assumes that every necessity is possible, so only systems which validate at least axiom D (cf. D), i.e., including any system which validates axiom T, since T implies D. Is satisfying the Square of Opposition is a sufficient condition for being a modality? A necessary condition? If it is either of these, then Saint Anselm of Canterbury is one of the first, if not the first, to interpret agency as a modal notion (cf. Figure 12.3.)

This is an interesting realization for two reasons. The first, and most straightforward, is that if agency is a modal notion, we can use the tools and techniques of modal logic to explore it.

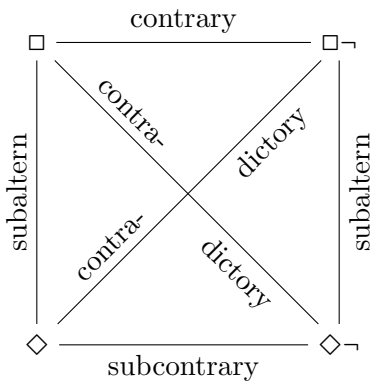


Figure 12.2: Modal square of opposition

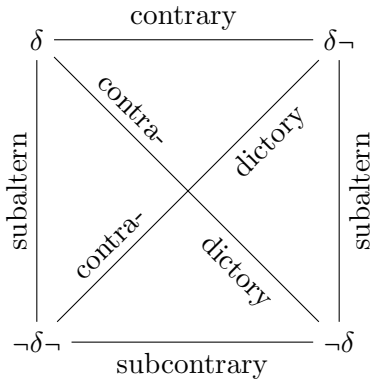


Figure 12.3: Agentive square of opposition

The second stems from the way in which this realization is embedded in Anselm's larger methodological programme. In his philosophical works, Anselm distinguishes between the *usus loquendi* or *usus communi* of words, and their *usus proprie* [Uckelman, 2009, §2.1]. The common way of speaking—how people actually speak—is the purview of grammarians, while the proper way to use words—how people *should* speak—is the purview of logicians. For Anselm, grammar is descriptive while logic is prescriptive. But there's more to it than just that: If logic and grammar diverge, i.e., if there are ways that language should be used in which it is not, or ways it is used in which it should not be, then logic must be able to give an explanation of where those discrepancies come from. That is, the discrepancies must

somehow be grounded, and logic should be able to tell us in what. Thus, the prescriptive rules of logic are only as good as their ability to explain the descriptive uses of words.

12.3.2 Obligation

Why is it relevant that we can use modal logics for agency? Because it allows us to be more precise about another modal notion: obligation.

In ordinary language, obligations are expressed in terms of agents and actions: "You are obliged/permitted/required to do X". But standard deontic logic has propositions of the form $O\varphi$: "It ought to be the case that φ is true"—that is, it is tied to propositions. One way in which formalizing obligations in this fashion is problematic is that it seems to remove a lot of the personal responsibility that is related to action: "It ought to be the case that no one starves" absolves individuals of the responsibility to ensure that this is the case. And not only that, it is unlikely that any single individual could fulfill this obligation! (Collective action is covered in §12.3.3 below.)

By combining deontic and agentive operators we can get a step closer: $O\delta_a\varphi$ is "a ought to do such that φ" indicates that obligations are tied to agents and their actions, but it leaves the actions underspecified. This brings us back to two ways that agentive logic can be approached:

- "see to it that", "does": "do whatever action is sufficient to guarantee φ." (Anselm, Belnap et al.)

- explicit actions: "Do this action, whatever the consequences" (cf. [Uckelman, 2014] for an example of this type of approach).

These two approaches to agency correspond to two different types of obligations:

- STIT: Consequentialist views (e.g., utilitarianism)

- Explicit actions: Deontological (e.g., Kant)

Now, what about the logic side of things? Ordinary basic deontic logic has two axioms, axiom K (see Lemma 11.3.10), and axiom D (see Lemma 11.3.40). That is, deontic logic is the logic which is characterized by the class of serial frames.

12.3.3 Collective action

12.4 Modal term logic

When we introduces the Aristotelian syllogistic in Chapter 6, we restricted our attention to *assertoric* sentences: Simple present-tense statements without any modal force. Aristotle himself considered extensions of the syllogistic to modal propositions (ones involving necessity, possibility, impossibility, and contingency), although he struggled to develop a coherent, consistent, and complete system (the pursuit of rehabilitating what Aristotle has said about modal propositions and modal syllogisms has given rise to an active subfield of modal logic for the last 2500 years). In this section, we build on the language introduced in §6.1 to define a modal term language, and then discuss the variety of ways people have attempted to give semantics to this language in such a way as to validate as many of Aristotle's claims as possible.

- $\mathbf{A}p \triangleq$ "p" (no modality, "assertoric").
- $\mathbf{N}p \triangleq$ "necessarily p".
- $\mathbf{P}p \triangleq$ "possibly p" (equivalently, "not necessarily not p").
- $\mathbf{C}p \triangleq$ "contingently p" (equivalently, "not necessarily not p and not necessarily not p").

Every (assertoric) mood $p, q : r$ represents a modal mood $\mathbf{A}p, \mathbf{A}q : \mathbf{A}r$. For each mood, we combinatorially have $4^3 = 64$ modalizations, i.e., $256 \times 64 = 16384$ modal moods.

As with assertoric syllogisms, we develop our theory of modal syllogisms by identifying the perfect syllogisms, which we take as axioms, along with a set of valid conversion rules. With modal syllogisms, it is easiest to first give the valid conversions. We note that simple and accidental conversions hold in all modalities:

- **Simple.**
  - $\mathbf{N}X\mathsf{e}Y \rightsquigarrow \mathbf{N}Y\mathsf{e}X$
  - $\mathbf{N}X\mathsf{i}Y \rightsquigarrow \mathbf{N}Y\mathsf{i}X$
  - $\mathbf{C}X\mathsf{e}Y \rightsquigarrow \mathbf{C}Y\mathsf{e}X$
  - $\mathbf{C}X\mathsf{i}Y \rightsquigarrow \mathbf{C}Y\mathsf{i}X$
  - $\mathbf{P}X\mathsf{e}Y \rightsquigarrow \mathbf{P}Y\mathsf{e}X$

- $\mathbf{P}X\mathsf{i}Y \rightsquigarrow \mathbf{P}Y\mathsf{i}X$

- **Accidental.**
  - $\mathbf{N}X\mathsf{a}Y \rightsquigarrow \mathbf{N}X\mathsf{i}Y$
  - $\mathbf{N}X\mathsf{e}Y \rightsquigarrow \mathbf{N}X\mathsf{o}Y$
  - $\mathbf{C}X\mathsf{a}Y \rightsquigarrow \mathbf{C}X\mathsf{i}Y$
  - $\mathbf{C}X\mathsf{e}Y \rightsquigarrow \mathbf{C}X\mathsf{o}Y$
  - $\mathbf{P}X\mathsf{e}Y \rightsquigarrow \mathbf{P}X\mathsf{o}Y$
  - $\mathbf{P}X\mathsf{a}Y \rightsquigarrow \mathbf{P}X\mathsf{i}Y$

There are also four conversion rules which reflect the symmetric nature of contingency:

- $\mathbf{C}X\mathsf{i}Y \rightsquigarrow \mathbf{C}X\mathsf{e}Y$
- $\mathbf{C}X\mathsf{e}Y \rightsquigarrow \mathbf{C}X\mathsf{i}Y$
- $\mathbf{C}X\mathsf{a}Y \rightsquigarrow \mathbf{C}X\mathsf{o}Y$
- $\mathbf{C}X\mathsf{o}Y \rightsquigarrow \mathbf{C}X\mathsf{a}Y$

And finally we have an implication between modes expressing the fact that whatever is necessary is in fact true:

- $\mathbf{N}X\mathsf{x}Y \rightsquigarrow \mathbf{A}X\mathsf{x}Y$ (this corresponds to Axiom $\mathbf{T}$:$\Box\varphi \rightarrow \varphi$ in modal logic.)

The next question we must answer is "What are the 'perfect' modal syllogisms?" This turns out to be very diffcult. We know that valid assertoric syllogisms remain valid if **N** is added to all three propositions: **Barbara** ($A\mathsf{a}B, B\mathsf{a}C$:$A\mathsf{a}C$) $\rightsquigarrow$ **NNN Barbara** ($\mathbf{N}A\mathsf{a}B, \mathbf{N}B\mathsf{a}C$:$\mathbf{N}A\mathsf{a}C$). By our conversion rules, we can also add the following to our list of valid assertoric syllogisms:

- **NNA**
- **NAA**
- **ANA**

Everything else is problematic. In particular we are faced with what is nowadays called the "two Barbaras" problem. Consider the following two modal syllogistic moods:

NAN Barbara	**ANN Barbara**
$\mathbf{N}A\mathtt{a}B$	$\mathbf{A}A\mathtt{a}B$
$\mathbf{A}B\mathtt{a}C$	$\mathbf{N}B\mathtt{a}C$
$\mathbf{N}A\mathtt{a}C$	$\mathbf{N}A\mathtt{a}C$

From the modern point of view, both modal syllogisms are invalid. To show that the first is invalid, let A: "white", B: "white bird", and C: "swan". The syllogism is thus:

Necessarily, every white bird is white.
Every swan is a white bird.

Ergo: Necessarily, every swan is white.

To show that the second is invalid, let A: "less than 300 years old", B: "animal", and C: "human". The syllogism is thus:

Every animal is less than 300 years old.
Necessarily, every human is an animal.

Ergo: Necessarily, every human is less than 300 years old.

In both cases we have true premises and intuitively false conclusions. Yet Aristotle claims that **NAN Barbara** is valid, but **ANN Barbara** is not.

We now look at a medieval solution to the two Barbaras problem, first offered by Peter Abelard. [Get something to cite for this.]

In the above, we interpreted $\mathbf{N}A\mathtt{a}B$ as

"The statement '$A\mathtt{a}B$' is necessarily true."

This is called the *de dicto* interpretation of necessity.

Alternatively, we could interpret $\mathbf{N}A\mathtt{a}B$ *de re* (Becker 1933):

"Every B happens to be something which is necessarily an A."

The remainder of this section is still to be written.

12.5 Provability logic and the logic of provability

Chapter 13

Temporal logic [last modified 01 Mar 21]

The intuition underlying the present chapter is a simple fact about ordinary language: That one and the same series of words can be uttered at different times and receive different truth values. For instance, the sentence "There is a rocking chair in my office" is currently true, and has been true for a number of years, and will (presumably) be true for a number of years to come. Yet, nothing prevents me from taking the rocking chair out of my office, at which point it is natural say that "There is a rocking chair in my office" is no longer true.

How are we to represent this logically? "There is a rocking chair in my office" is, on the face of it, a statement of the existence and location of a particular object, a rocking chair, so we might first attempt to render it in predicate logic.

Example 13.0.1. Let $\mathcal{L}_q$ be a quantified language containing the constant o, the unary predicate C and the binary relation I. Informally interpreting o as "(my) office", a fixed location, C as "is a chair", and I as "in", "There is a rocking chair in my office" is:

$$\exists x(Cx \wedge Ix\mathsf{o}) \tag{13.1}$$

Given a suitable model (cf. Definition 8.3.1) and an interpretation that respects the informal interpretation in the example, we can then determine whether (13.1) is true on the model.

But... that is it. Once we have fixed a language and a model, every sentence in that language has a determine, fixed truth value in that model. There is no way to represent the possibility of the changing truth value of the sentence, because the model does not change. In our informal motivation, we pointed out that it the chair could be removed from the

office, making the sentence false. But this movement—this change in the model—cannot be represented in a fixed model. Thus, we cannot express what we wanted to express in predicate logic.

"But," I hear you say, "surely it is *possible* that the chair not be in the office—it isn't necessary that it be in the office; so just because it is true on a given model it doesn't follow that it is true on *every* model." And indeed, you would be correct; as we saw in Chapter 11, we can represent the varying of the truth value of contingent sentences across models—truth-value assignments, possible worlds—with the tools of modal logic. But it is one thing to say that in some possible world the rocking chair is in my office and in other possible worlds it is not, and another thing altogether to say that sometimes the rocking chair is in my office and sometimes it is not—statements about the *actual* world, not about other possible worlds. And given that the chair *is* in my office in the actual world, it is *impossible* that it also not be in my office in the actual world; for unlike quantum particles, chairs cannot be both in and not in a given location. Therefore, modal logic will not do the trick either.

13.1 What is temporal logic?

By "temporal logic" we mean a family of logics and logical techniques which can be applied to a wide range of problems, both abstract and concrete, in philosophy and computer science, which are unified by the fact that they all deal with time in some fashion or other. Thus, we use "temporal logic" as an umbrella term covering a large number of different theories, logical systems, and formal tools that are variously called "temporal logic", "logic of time", and "tense logic". The divide between "logic of time" and "tense logic" is grey and amorphous; many of the problems and issues that we will place under one header can be described in terms of specifications under the other. Many of the tools and techniques which can be used in one can be used in the other. We will call all of these various facets "temporal logic", and call anything "temporal logic" that has some connection with **Time**.

If temporal logic is the logic of time, whether this manifests itself in tensed or untensed statements, we should at least comment on the question **What is time?**

Aristotle defines time as the "number of motion in respect to before and after" (Physics bk. vi 11 (219 b 2), where "number" here is in the sense of that which is countable, rather than that which is counted [Corish, 1976, p. 241]. [McTaggart, 1908] argues, on the basis of a similar definition, that time does not exist in reality. His argument, or whether his conclusion is correct or not, needn't occupy us here, but his article is important historically for two concepts that he introduces which have become entrenched in the philosophical

study of temporal logic: The distinction between the A-series and the B-series views of time.

A-series A series of positions which runs from the past, to the present, to the future. [This gives rise to tense logic]

B-series A series of positions which are ordered with respect to earlier and later. [This gives rise to logic of time]

The A-series is dynamic: A single position can be viewed, from different viewpoints, as being either past, present, or future. The B-series position, however, is static; if position a is earlier than b it can never be other than earlier than b. (McTaggart's argument is essentially that one cannot express time by a B-series alone, because time involves change, and the positions as described in the B-series never change; however, time involving change presupposes the existence of an A-series, but the A-series presupposes the existence of time). He also introduces a third type of series, the C-series, which is simply an ordering on positions/events. This ordering does not become temporal (i.e., a B-series), unless time/change is added.

These two series correspond to the different approaches to temporal logic, which we can broadly classify into two types: Logics of Time and Tense Logics:

- Logic of time.
  - Linear time, branching time. Discrete time, dense time, continuous time.
  - Point-based vs. interval-based.
  - Connections to statistical interpretations of modality.
  - Timelessly true or time-indexed propositions.
    - Hybrid logic
  - Problems in computer science.
    - Specifying program behavior, program properties.
    - Modeling computational processes.
- Tense logic.
  - Analysis of tensed propositions.
    - Reichenbach's (1947) tripartite analysis (building on Jesperson), in terms of S (point of speech), R (point of reference), and E (point of event). See Figure 13.1.

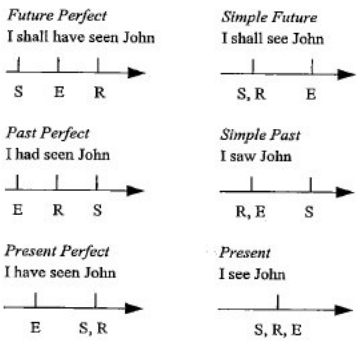


Figure 13.1: The six English tenses [Øhrstrøm and Hasle, 2006b, p. 451]

  - ∗ Problems of complicated embedded tenses: "I shall have been going to see John".
- – Semantics of natural language.
- – Reasoning from and to tensed statements.
- – Metric tense logic.
- – Problems in philosophy.
  - ∗ The problem of future contingents.
  - ∗ Quantified modal/tense logic: divided and composite readings.
  - ∗ Modeling space-time.
  - ∗ Time travel

These two types correspond to the "first-order approach" and the "modal approach" of Galton [Galton, 1987, p. 2ff], and some specific features of both are given in Figure 13.2. (Temporally definite propositions never change their truth value. Temporally indefinite ones change their value depending on their time of evaluation.)

What is interesting is that each of these ways of approaching temporal logic can be slotted in to one of Prior's four grades of tense-logical involvement discussed below. So, we are justified in lumping all of this together and calling it "temporal logic" in the singular.

Where does temporal logic fit in the scheme of things?

Is temporal logic prior to or posterior to "ordinary" logic?

Logic of Time	Tense logic
B-series	A-series
Static	Dynamic
Fixed earlier than/later than relation	"Is present, was future, will be past"
Temporally definite propositions:	Temporally indefinite propositions:
"It is raining at 12:05am on Feb. 25, 2009"	"It is raining".

Figure 13.2: Two approaches to temporal logic

> It was Prior's view that, properly understood, all of logic is really temporal, and that logical languages without some kind of temporal operators were really devoted just to a proper subset of logic [Øhrstrøm and Hasle, 2006b, p. 448].

Though not said explicitly in so many words, this view is very similar to that of the medieval logicians, who took the study and analysis of tensed statements as a basic requirement of logic. The belief that tensed propositions are the basic/fundamental building blocks of logic implies that you cannot divorce a proposition from its tense when you analyze it.

Medieval logicians (and perhaps also the Stoics) took seriously the idea that one and the same proposition could have different truth values at different times. This is in contrast to the modern view of propositions as abstract objects which exist necessarily and have their properties necessarily (hence do not change their properties). The first modern logician to try to formalize this medieval idea was Arthur Prior, who developed temporal logic as we know it today during the 1950s and 1960s. Prior was motivated to study the problem not only because of what he'd come across in the ancient and medieval logicians, but also because of the problem of free will and determinism, a problem which occupied him because of his theological background [Øhrstrøm and Hasle, 2006a, §2].

Prior, however, did not see himself as the father of temporal logic: instead, he gave that honor to his teacher J.N. Findlay.

Findlay's footnote.

In [Findlay, 1941], where Findlay considers the problem of why we sometimes find time to be mysterious, puzzlesome, or even paradoxical, given that we can (and regularly do) successfully make temporal statements whose truth conditions and generally agreed upon[1], the following footnote appears:

[1] Findlay believes that temporal problems can be difficult even if we have all the other philosophical problems; in that sense, problems relating to time are rather nice to study because they are, in a sense, self-contained.

The calculus of tenses should have been included in the modern development of modal logics. It includes such obvious propositions as that

$$\begin{aligned} x \text{ present} &= (x \text{ present}) \text{ present}; \\ x \text{ future} &= (x \text{ future}) \text{ present} = (x \text{ present}) \text{ future}; \end{aligned}$$

also such comparatively recondite propositions as that

$$(x).(x \text{ past}) \text{ future}$$

i.e., all events past, present and future, *will* be past [Findlay, 1941, fn. 17].

It is in this proposition that Prior finds the germ of temporal logic. (Note here that the "temporal logic" is of the latter kind—a calculus of tenses, not of times.) More recent authors have found Prior rather generous to attribute the fatherhood of temporal logic to Findlay on the basis of this note, but given that it was this footnote that inspired Prior, and it was Prior who developed the field as we know it, perhaps there is nothing wrong in recognizing Findlay as the inspirational father.

Prior's approach to logics of time and tense was almost exclusively syntactic (but see below); however, he rejected the view that logic was symbol manipulation without meaning [Copeland, 2008]. His choice of logical problems to solve was always motivated by specific temporal problems (such as the correct characterization of Diodorean modality, see below). In particular, much of his work dealt with the relationship between time and necessity and on problems related to quantified modal and temporal logic. Most of his work on time and tense has been collected in the books *Time and Modality* [Prior, 1957], *Past, Present, and Future* [Prior, 1967], *Papers on Time and Tense* [Prior, 1968], and *Worlds, Times, and Selves* [Prior and Fine, 1977].

13.2 Minimal propositional tense logic

In this section, we introduce the minimal propositional tense logic $\mathrm{K_t}$. It is *minimal* in the sense that we do not make any assumptions about the nature of time, e.g., whether it is discrete, dense, or continuous, whether it is linear or branching, whether it has a beginning or an end, whether it contains any cycles, etc.—though we will discuss how we can model these constraints through the addition of certain axioms to the minimal tense logic. In this, much of what we introduce will resemble the minimal modal logic K introduced in Chapter 11. We therefore recommend the reader refresh themself with the basic semantics and proof theory of K and its extensions, as seeing the analogies between modality and time will make our account of the temporal logic more natural.

Throughout this chapter, we are going to focus on the use of temporal logic as a modeling tool to clarify particular philosophical or computational problems; we will be less interested in providing formal proofs within the temporal systems. Nevertheless, we will begin as we have done when introducing other logical systems: With the language (§13.2.1), the semantics (§13.2.2), and the proof theory (§13.2.3).

13.2.1 Language

In this section, we define a minimal language for propositional tense logic and define rules for the construction of well-formed formulas.

Definition 13.2.1. A *propositional tense language* $\mathcal{L}_{T_p}$ consists in:

- A countably infinite set of atomic proposition letters $p, q, r, \ldots$
- Eight logical connectives: Five unary $\neg$, F, P, G, H, three binary: $\vee$, $\wedge$, $\rightarrow$.
- Punctuation: (,).

This language is identical to the language of propositional logic (cf. Definition 7.3.2) with the addition of four unary operators: F, P, G, and H.

The sett of wffs of $\mathcal{L}_{T_p}$ is defined recursively:

Definition 13.2.2 (Propositional tense logic wffs)**.**

- Every atomic proposition is an $\mathcal{L}_{T_p}$-wff.
- If φ and ψ are $\mathcal{L}_{T_p}$-wffs, then so are $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, and $(\varphi \rightarrow \psi)$.
- If φ is an $\mathcal{L}_{T_p}$-wff, then so are $\neg\varphi$, $F\varphi$, $P\varphi$, $H\varphi$, and $G\varphi$.

When no confusion results, we will omit the outermost parentheses.

13.2.2 Semantics

Intuitively, we will read the new operators as follows:

$F\varphi$ "It will sometime be the case that φ".

$P\varphi$ "It was sometime the case that φ".

$G\varphi$ "It will always be the case that φ".

$H\varphi$ "It was always the case that φ".

That is, we can understand F and P as the tense analogues of $\Diamond$, expressing future and past possibility, and G and H as the tense analogues of $\Box$, expressing future and past necessity—that is, tense operators are a type of modal operator.

However, tense logic differs from other modal logics in that instead of having two modalities corresponding to the relevant necessity (universal) and possibility (existential) notions, it has four: Two look forward to the future and two look backward to the past. Thus, tense logic is a *multi-modal* system.

Like their modal counterparts $\Box$ and $\Diamond$, the semantics of the tense operators can be defined via Kripke frames. However, in addition to G and H being the duals of F and P, just as $\Box$ is the dual of $\Diamond$ in mono-modal logic, F and G are also the converses of P and H. The truth conditions of these operators are defined as follows:

Definition 13.2.3 (Truth conditions for tense operators)**.** Let $\mathfrak{M}$ be a Kripke model as defined in Definition 11.3.2. Then:

$$\begin{array}{lll}
\mathfrak{M}, w \vDash P\varphi & \text{iff} & \text{there is } w' \text{ such that } w'Rw \text{ and } \mathfrak{M}, w' \vDash \varphi \\
\mathfrak{M}, w \vDash F\varphi & \text{iff} & \text{there is } w' \text{ such that } wRw' \text{ and } \mathfrak{M}, w' \vDash \varphi \\
\mathfrak{M}, w \vDash H\varphi & \text{iff} & \text{for all } w' \text{ such that } w'Rw, \mathfrak{M}, w' \vDash \varphi \\
\mathfrak{M}, w \vDash G\varphi & \text{iff} & \text{for all } w' \text{ such that } wRw', \mathfrak{M}, w' \vDash \varphi
\end{array}$$

Properties of time

All of the Kripke models above are point-based. In a point-based time-model there are a number of different properties of time that we might want to express via constraints on the model, such as:

$\forall x \in W$	$\neg xRx$	irreflexivity
$\forall xy' \in W$	$xRy \to \neg yRx$	antisymmetry
$\forall x \forall y \forall z \in W$	$xRy \wedge yRz \to xRz$	transitivity
$\forall x \exists y \in W$	xRy	no last point
$\forall x \exists y \in W$	yRx	no first point
$\forall xy \exists z \in W$	$xRy \to xRz \wedge zRy$	density
$\forall x \forall y \forall z \in W$	$xRz \wedge yRz \to xRy \vee yRx \vee x = y$	backwards linearity
$\forall x \forall y \forall z \in W$	$zRx \wedge zRy \to xRy \vee yRx \vee x = y$	forwards linearity

Many of these properties correspond to tense-logic formulas (in the sense that the class of frames satisfying the property is exactly the class of frames which validate the formula),

though not all of them do. Some formulas which correspond to frame properties are:

$G\varphi \to GG\varphi$	transitivity
$F(\varphi \vee \neg\varphi)$	no last point
$P(\varphi \vee \neg\varphi)$	no first point

Not all frame properties correspond to valid formulas. For example, there is no formula whose validity on a class of frames implies that the frames are irreflexive. Having formulas which correspond to frame properties is useful for using the canonical model method of proving completeness of modal or temporal systems: See §11.5.2.

We give a straightforward example showing how correspondence between a formula and a frame property can be proven:

Lemma 13.2.4. *A frame $\mathfrak{F}$ is reflexive iff every model $\mathfrak{M}$ based on $\mathfrak{F}$ validates $G\varphi \to \varphi$ (or $H\varphi \to \varphi$).*

Proof. We do just the case of $G\varphi \to \varphi$.

($\Rightarrow$) Assume $\mathfrak{F}$ is reflexive, i.e., $\forall w \in W, wRw$. Now, suppose that there is $w \in W$ such that $\mathfrak{M}, w \vDash Gp \wedge \neg p$. If $\mathfrak{M}, w \vDash Gp$, then by the fact that wRw, $\mathfrak{M}, w \vDash p$. But then $\mathfrak{M}, w \vDash p \wedge \neg p$, which is a contradiction.

($\Leftarrow$) We prove contrapositively. Suppose that $\mathfrak{F}$ is not reflexive, i.e., $\exists w \in W, \neg wRw$. Then let $\mathfrak{M}$ be a model on $\mathfrak{F}$ such that V makes p is true everywhere except for w. Either $\exists w', wRw'$ or not. If not, then $\mathfrak{M}, w \vDash Gp$, trivially, and hence $\mathfrak{M}, w \nvDash Gp \to p$. If so, then by our definition of V, $\mathfrak{M}, w' \vDash p$, and this holds for any w' such that wRw'. So $\mathfrak{M}, w \vDash Gp$ again, and the implication is false. □

13.2.3 Proof theory

The minimal (propositional) tense logic $\mathrm{K_t}$ has the following axioms and rules of inference:

A1 φ, where φ is a tautology of classical propositional logic.

A2 $G(p \to q) \to (Gp \to Gq)$.

A3 $H(p \to q) \to (Hp \to Hq)$.

A4 $p \to HFp$ (or its contrapositive $FHp \to p$).

A5 $p \to GPp$ (or its contrapositive $PGp \to p$).

Rule 13.2.5 (Modus ponens (MP))**.** If $\vdash \varphi$ and $\vdash \varphi \to \psi$ then $\vdash \psi$.

Rule 13.2.6 (Uniform Substitution (US))**.** The result of uniformly replacing atoms $p_1, \ldots, p_n$ by wffs $\varphi_1, \ldots, \varphi_n$ in a theorem is also a theorem.

Rule 13.2.7 (RG)**.** If $\vdash \varphi$ then $\vdash G\varphi$.

Rule 13.2.8 (RH)**.** If $\vdash \varphi$ then $\vdash H\varphi$.

Readers familiar with the minimal (propositional) modal logic K will recognize axioms A2 and A3 as the tense-logical analogues of axiom K (the distributivity of necessity over implication), and rules RG and RH as the tense-logical analogues of the Rule of Necessitation.

This logic corresponds to the class of Kripke frames which have no constraints on R. But just as in the modal case, we may want to put constraints on the class of temporal models that we consider, for example, we might think that it is infinite in both directions, continuous, and linear. Various potential constraints are listed in §11.3.1, and others are listed below, along with the axioms that correspond to the constraints (in the technical sense of correspondence discussed in that section). All of the results in §11.3.1 carry over to their temporal analogues. For example, the temporal analogue of Lemma 11.3.40:

Lemma 13.2.9. *The temporal formula $Gp \to Fp$ is valid on a temporal frame $\mathfrak{T} = \langle T, < \rangle$ iff $<$ is (forwards) serial.*

can be proven in a completely analogous fashion:

Proof.

$\Leftarrow$ Let $\mathfrak{T} = \langle T, < \rangle$ be such that $<$ is serial (that is, $\forall x \exists y (x < y)$), but suppose that $Gp \to Fp$ is not valid on $\mathfrak{T}$. Then there is a model $\mathfrak{M}$ on $\mathfrak{T}$ and some $t \in T$ such that $t \not\models Gp \to Fp$; that is, (1) $t \models Gp$ and $t \not\models Fp$, that is, (2) $t \models \neg Fp$. If $t \models \neg Fp$, then (3) $t \models G\neg p$. Since $<$ is serial, there is a t' such that $t < t'$. By (1), $t' \models p$; by (3), $t' \models \neg p$. But this is a contradiction.

$\Rightarrow$ Proved contrapositively. Suppose that $\mathfrak{T} = \langle T, < \rangle$ is such that $<$ is not serial. Then, there is some $t \in T$ that is a dead end (that is, it cannot see anything). Let $\mathfrak{M}$ be any model on $\mathfrak{T}$, with the valuation arbitrary. Either (1) $t \models p$ or (2) $t \models \neg p$. Either way, (3) $t \models Gp$ and (4) $t \models G\neg p$. By (4), (5) $t \models \neg Fp$. But then by (3) and (5), (6) $t \not\models Gp \to Fp$, showing that $Gp \to Fp$ is not valid on this frame.

□

Other corresponding axioms include:

- No first point:

$$Hp \to Pp$$

- No last point (see Lemma 11.3.40):

$$Gp \to Fp$$

- Backward linearity:

$$FPp \to (Pp \vee p \vee Fp)$$

- Forward linearity:

$$PFp \to (Pp \vee p \vee Fp)$$

- (Forward) convergence (see Lemma 11.3.44):

$$FGp \to GFp$$

- (Forward) density:

$$Fp \to FFp$$

- (Forward) discreteness (in conjunction with reflexivity)

$$G(G(p \to Gp) \to p) \to (FGp \to p)$$

Backward versions of the forward properties can be constructed by simply replacing G with H and F with P.

Sometimes we can express properties of time only under the assumption of other properties. E.g., assuming reflexivity, the following two correspondences hold:

- There is a last point:

$$(GFp \wedge GFq) \to F(p \wedge q)$$

- There is a first point:

$$(HPp \wedge HPq) \to P(p \wedge q)$$

If we make certain natural assumptions about the properties of time (for example, that it is, two-way infinite, continuous, and linear), it is possible to ask how iterated modalities collapse (as we did in the modal case). As an example, considered by Prior and Hamblin, if the following axioms are added to $\mathrm{K_t}$, we can prove some nice things about the number of distinct tenses:

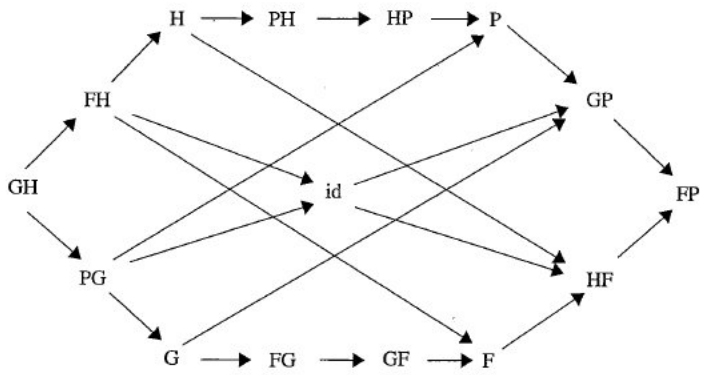


Figure 13.3: Hamblin's bird's nest[Øhrstrøm and Hasle, 2006a, p. 429]

Ax1 $F(p \vee q) \leftrightarrow (Fp \vee Fq)$

Ax2 $Gp \rightarrow Fp$

Ax3 $FFp \leftrightarrow Fp$

Ax4 $FPp \leftrightarrow (p \vee Fp \vee Pp)$

Ax5 $GPp \leftrightarrow (p \vee Pp)$

along with the following rules:

Rule 13.2.10 (R2)**.** If $\vdash \varphi \leftrightarrow \psi$ then $\vdash F\varphi \leftrightarrow F\psi$.

Rule 13.2.11 (R3)**.** If $\vdash \varphi$, and φ' is the result of simultaneously replacing all F's in φ by P, and vice versa, then $\vdash \varphi'$.

then it can be proven that there are only 30 distinct tenses (see Figure 13.3).

We can also prove a result that connects $\mathrm{K_t}$ with the language and logic that we'll introduce in Section 13.9:

Theorem 13.2.12. *The theorems of* $\mathrm{K_t}$ *are exactly those which are provable in* U *when preceded by* T_a.

Proof. to be proven. □

13.3 Time and necessity

One question that immediately arises, after the introduction of a tense logic like $\mathrm{K_t}$, is:

What is the relationship between time and necessity?

Or, to put the question more abstractly and less philosophically, can we recover the uni-modal system, with $\Box$ and $\Diamond$, from the bi-modal temporal system?

There are two natural ways that we can define $\Box$ in terms of the temporal operators:

$$\Box_A \varphi \quad := \quad \varphi \wedge G\varphi \wedge H\varphi \tag{13.2}$$

$$\Box_D \varphi \quad := \quad \varphi \wedge G\varphi \tag{13.3}$$

In the first way, φ is necessary if it is true now, has always been true in the past, and will always be true in the future. This captures many of our natural intuitions about necessity, including that what is always true is necessary, and that, e.g., laws of nature and mathematics because they are necessary are always true.

In the second way, φ is necessary if it is true now, and will always be true in the future; on this account, what has been the case in the past is not relevant for determining necessity. This account of necessity adheres to an implicit principle that governed all of the systems we introduced in Chapter 11: Namely, that what is true in worlds that can see w is irrelevant to what is necessary at w—only what is true at worlds that w can see is relevant.

Both of these approaches have historical antecedents; the first is the Aristotelian-Megarian interpretation of the relationship between time and necessity. (We will say more about this interpretation here later.) Apart from the adherence to an implicit principle that is satisfied in the uni-modal case (and which when made explicit may perhaps no longer seem all that plausible), why might we think the second account is plausible? We discuss one historical argument, due to Diodorus Cronus, for this definition of necessity in §13.6, but we will for the remainder of this section set aside these historical and motivational questions, and focus on technical matters.

Each of the two accounts corresponds to a well-known modal system: If the underlying tense logic is $\mathrm{K_t}$, then the logic of $\Box_D$ is T, while the logic of $\Box_A$ is B + T.

However, if we adopt a stronger underlying tense logic, such as one including the temporal analogues of the 4 axiom, stating that the future of the future is also the future of now, and that the past of the past is also the past of now:

$$FFp \to Fp \tag{F4}$$

$$PPp \to Pp \tag{P4}$$

then the logics of $\Box_D$ and $\Box_A$ must be at least S4 and S5, respectively, from which it follows that the logic of $\Box_D$ *cannot* validate the B axiom, or the two definitions of necessity would collapse. (The correspondence of the logic of $\Box_D$ and S4 was first demonstrated by Prior in [Prior, 1955]).

In [Prior, 1967], Prior revisited the logic of Diodorean necessity, defined as above, and asked what system of logic is generated when time is taken to be an infinite matrix of points where each proposition receives either the value 0 or the value 1. Though he does not put it in such terms, it is clear that he is adopting Kripke-style semantics; we can view these infinite matrices as defining frames where every point in the frame is related to itself (since $\Box_D p$ implies p, by definition) and to every point to its right (cf. Figure 13.4). Because these matrices are both reflexive and transitive, the logic of Diodorean necessity over such a matrix will be *at least* S4. Prior himself originally thought that it was *exactly* S4[2], though later investigate proved this original hunch incorrect.

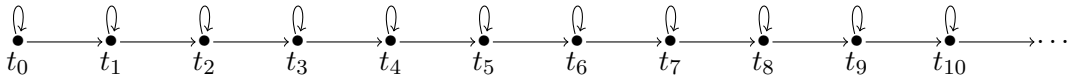


Figure 13.4: A Diodorean matrix

Why might Prior have thought that the logic of infinite matrices like the one in Figure 13.4 was S4? At the time, only one between S4 and S5 was known, and that was the system S4.5 due to W.T. Parry.

Definition 13.3.1. The logic S4.5 is the logic S4 + the axiom $\Box \Diamond \Box p \to \Box p$.

But this logic is too strong; the Parry axiom is not validated by the Diodorean matrices, as Figure 13.5 shows. However, there are in fact other systems between S4 and S5 which are not S4.5, as we shall see by considering specific axioms which are validated by the Diodorean matrices but which are not theorems of S4.

Lemma 13.3.2. *The formula $\Diamond \Box p \to \Box \Diamond \Box p$ is valid on Diodorean frames.*

Proof. Suppose it isn't. Then there is a Diodoream frame $\mathfrak{F}$ and a model $\mathfrak{M}$ on $\mathfrak{F}$ with a time t where

$$\begin{array}{lll} 1 & \mathfrak{M}, t & \vDash \Diamond \Box p \\ 2 & \mathfrak{M}, t & \vDash \neg \Box \Diamond \Box p \end{array}$$

[2] I will provide a citation for this.

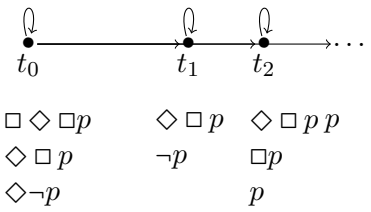


Figure 13.5: Parry's axiom is falsified on a Diodorean matrix

that is,

$$3 \quad \mathfrak{M}, t \quad \vDash \quad \Diamond \Box \Diamond \neg p$$

From (1) it follows that there is a future time t' where

$$4 \quad \mathfrak{M}, t' \quad \vDash \quad \Box p$$

that is, p is true at t' and forever after. Because the frame is transitive, it follows that:

$$5 \quad \mathfrak{M}, t' \quad \vDash \quad \Box \Box p$$

From (3) it follows that there is also a future time t'' where

$$6 \quad \mathfrak{M}, t'' \quad \vDash \quad \Box \Diamond \neg p$$

Now, either (a) $t = t''$, (b) $t < t'' < t'$, (c) $t'' = t'$, or (d) $t' < t''$; it is easy to check that in all four of these cases, there is a contradiction with (5). □

But this formula is not a theorem of S4, which we can show with a simple argument. Recall that S4 has seven distinct positive modalities, whose implication relations are given in Figure 11.12. In this figure, $\Box \Diamond \Box p$ implies $\Diamond \Box p$, but not vice versa. *IF* $\Diamond \Box p \rightarrow \Box \Diamond \Box p$ were added to S4, then this part of the implicational structure of the modalities would collapse, and there would be fewer than seven distinct positive modalities.

As it turns out, $\Diamond \Box p \rightarrow \Box \Diamond \Box p$ is in fact derivable in S4 with the addition of the Geach axiom:

$$\Diamond \Box p \rightarrow \Box \Diamond p \qquad (\mathsf{Geach})$$

In Chapter 11, we showed that this axiom corresponds to the convergence or 'diamond' property (Lemma 11.3.44). The Geach axiom is also not a thesis of S4, since S4 has frames which are forever divergently branching.

Another formula which is valid on Diodorean frames but which is *not* a thesis of S4 is the Lemmon axiom:

$$\Box(\Box p \to \Box q) \vee \Box(\Box q \to \Box p) \qquad \text{(Lemmon)}$$

We will prove that this formula can be falsified in a reflexive and transitive model, and leave the proof of its validity on Diodorean frames as an exercise for the reader.

Lemma 13.3.3. Lemmon *is not a theorem of* S4.

Proof. We give a reflexive and transitive from on which it is falsefied at t. Let $W = \{t, t', t''\}$, with $t < t$, $t' < t'$, $t'' < t''$, $t < t'$, and $t < t''$; then $\mathfrak{T} = \langle T, < \rangle$ is a reflexive and transitive frame. Define a model on $\mathfrak{T}$ where p is false at t'' and true everywhere else; and q is false at t' and true everywhere else. Then:

$$\begin{array}{rlll} 1 & \mathfrak{M}, t' & \vDash & \Diamond\neg q \\ 2 & \mathfrak{M}, t'' & \vDash & \Diamond\neg p \end{array}$$

but

$$\begin{array}{rlll} 3 & \mathfrak{M}, t' & \vDash & \Box p \\ 4 & \mathfrak{M}, t'' & \vDash & \Box q \end{array}$$

so

$$\begin{array}{rlll} 5 & \mathfrak{M}, t' & \vDash & \neg(\Box p \to \Box q) \\ 6 & \mathfrak{M}, t'' & \vDash & \neg(\Box q \to \Box p) \end{array}$$

and hence

$$7 \quad \mathfrak{M}, t \quad \vDash \quad \neg\,\Box\,(\Box p \to \Box q) \wedge \neg\,\Box\,(\Box q \to \Box p)$$

So Lemmon is false at t. □

The frame condition that Lemmon corresponds to is *linearity*.

Definition 13.3.4. The logic S4.2 is S4 + Geach.

Definition 13.3.5. The logic S4.3 is S4 + Lemmon.

It is clear that Diodorean frames are reflexive, transitive, and linear (and hence have the diamond property); however, there is one further property that they have, and that is *discreteness*. We introduce one new axiom, which is not a theorem of S4, S4.2, or S4.3, Dummett's axiom as simplified by Geach:

$$\Box(\Box(p \to \Box p) \to p) \to (\Diamond\,\Box\,p \to p) \qquad \text{(Dum)}$$

This axiom corresponds to discreteness in the presence of reflexivity.

Theorem 13.3.6. S4.3Dum *is the logic of* $\Box_D$ *over Diodorean frames.*

13.4 Diodorus's master argument

Diodorus Cronus (c340–284BC) was a Greek philosopher of the Dialectical school (contemporary with, but distinct from, the Megarian school [Sedley, 1977, p. 75]) who taught in Athens and in Alexandria between 315 and his death [Sedley, 1977][3] He was a student of Apollonius Cronus, and a teacher of Zeno of Citium. No writings of his remain. All that we know of his views comes from philosophers (often from schools that opposed him) writing many years later, including the *Discourses* of Epictetus (written down by Arrian c108AD), ii.19.1, Cicero's *De Fato* (45AD), and (indirectly) Boëthius. He is best known for his so-called "Master Argument", in which he proves that the possible is that which is or will be true. Only the premises and the conclusion of the argument are known; we do not know how the premises were ordered or how he reasoned from them to the conclusion. The clearest statement of the argument is in the *Discourses*:

> The Master Argument seems to have been developed from the following starting points: There is a general conflict between the following three statements: (I) every past true proposition is necessary; and (II) the impossible does not follow from the possible; and (III) something is possible which neither is true nor will be true. Being aware of this conflict, Diodorus used the plausibility of the first two statements in order to show that (IV) nothing is possible that neither is nor will be true. (Epictetus, Dissertations 2.19.1)

We can give a simplified reconstruction of the argument as follows:

1. Assume that p is not the case.
2. In the past, "It will be the case that p is not the case" was true.
3. In the past, "It will be the case that p is not the case" was necessarily true.
4. Therefore, in the past, "It will be the case that p" was impossible.
5. Therefore, p is not possible.
6. *Ergo:* Everything that is possible is true.

[3] Sedley relates a great anecdote about how he supposedly died. "He was in Alexandria, where both he and Stilpo happened to be guests at a banquet given by Ptolemy Soter. Stilpo presented Diodorus with a set of logical puzzles, which he could not immediately solve. Ptolemy made things worse by scolding Diodorus and calling him by his nickname Cronus. Diodorus left the court, wrote out a solution to the problems, then died in misery" [Sedley, 1977, p. 80].

This argument can be understood as either an argument for determinism, or as a clarification of the relationship between tense/time and modality. Ancient authors were agreed that (1)–(3) were in fact inconsistent, though they did not all agree that (3) should be rejected in favor of (1) and (2). Chrysippus, for example, a contemporary of Diodorus's, rejected (2) in favor of (1) and (3). He believed that an impossible thing may follow from a possible one. His example, discussed by Prior in [Prior, 1967, pp. 148–149], is "If Dion is dead, this man [indicating Dion] is dead". Chrysippus believes that this is a sound conditional, but while the antecedent could be true (and in fact at some point *will* be true), the consequent will never be true, because any time that the antecedent is true, there will no longer be any referent for 'this man' in the consequent (because the man Dion will no longer exist, and so cannot be pointed at), so "this man is dead" no longer expresses a proposition. Since it does not express a proposition, it cannot have a truth value.

The correct translation of the premises from the original Greek is a complicated question. [Foster, 2008] gives a detailed discussion of the possible translations of the premises that have been considered in the literature. One standardly-used translation is that given by Mates in [Mates, 1961, p. 38]:

M1 Every proposition true about the past is necessary.

M2 An impossible proposition cannot follow a possible one.

M3 There is a proposition which is possible, but which neither is nor will be true.

Sedley uses the following translations [Sedley, 1977, p. 97]:

S1 'Every proposition true about the past is necessary' — a true statement about the past can never become false.

S2 'An impossible proposition does not follow from a possible one' — a sound rule of modal logic, already well known to Aristotle.

S3 'There is something possible which neither is nor will be true' — the view which Diodorus wanted to reject in order to prove his definition.

13.4.1 Reconstructing the argument

The simplified reconstruction given above is *too* simple: This needs to be converted into something formal.

Sedley reconstructs the argument as follows:

Temporally definite propositions, follows temporally	Temporally definite propositions, follows logically
Temporally indefinite propositions, follows temporally	Temporally indefinite propositions, follows logically

Figure 13.6: Four ways to reconstruct the Master Argument

> What neither is nor will be true is impossible. For if (p) it were or were going to be true, then (q) it would already in the past have been the case that it would be true. But not-q — and necessarily not-q, since every proposition true about the past is necessary. Therefore it is impossible that-q. Therefore it is impossible that-p, since an impossible proposition does not follow from a possible one. Therefore nothing is possible which neither is nor will be true [Sedley, 1977, p. 97].

Translations of the premises such as these raise two interpretational questions that must be answered when trying to reconstruct the Master Argument:

1. How do we understand 'proposition'? Are propositions temporally definite or temporally indefinite?

2. How do we understand 'follow'? Is it "follows temporally" or "follows logically"?

Correspondingly, there are four different ways to formally reconstruct the argument, as depicted in Figure 13.6.

There are examples of each of the four interpretations in recent literature on the Master Argument [Foster, 2008, Table 1]. Michael's reconstruction is discussed in [Øhrstrøm and Hasle, 1995], where it is pointed out [p. 17] that it over-generates, since it ends up proving that 'true' and 'necessary' are equivalent, and the same with 'false' and 'impossible', which is clearly not the case for Diodorus, Chrysippus, or anyone else who considered the argument or related arguments. On the other hand, that this collapse could follow from the argument was already recognized by Diodorus's followers; Sedley says that "the object of the Master Argument was to establish the definitions of 'possible' and 'necessary', and we know that [Diodorus] kept these two terms carefully distinguished. He did not take the step of identifying the actual, the possible and the necessary. But it is clear from Cicero's *De fato* that this step was foisted on him, no doubt by his pupils" [Sedley, 1977, p. 99].

Probably the most well-known reconstruction of the argument is given by Prior. Prior was inspired by Mates's "Diodorean Implication" [Mates, 1949] as well as *Stoic Logic* [Mates, 1961]. Prior discusses the argument first in [Prior, 1955], further in [Prior, 1957,

chs. II, III], and most fully in [Prior, 1967, ch. II]. Prior interprets the propositions as temporally indefinite, and the "follows" relation as (logical) implication. His reconstruction, however, requires the addition of two further premises [Prior, 1955, p. 210]:

P4 When anything is the case, it has always been the case that it will be the case.

P5 When anything neither is nor will be the case, it has been the case that it will not be the case.

Both of these premises are not explicitly discussed in any of the sources (or, I think, but am not sure, in Stoic logic generally), but they can be found in Aristotle. On the basis of this, Prior thinks it's OK to add them. However, given that the Master Argument can be understood as a reaction *against* Aristotelian temporal philosophy, this type of justification for the premises may be problematic. (P4) in particular is not innocuous: It imposes linearity. We'll see in §13.6 that on forward-branching structures, both Peircean and Ockhamist, this premises is not validated. Sedley argues that this premise actually falls out of a belief in bivalence, that "true and false are the only possible truth-values for any proposition, including those about the future" [Sedley, 1977, p. 98]. But you can have bivalence without having linearity.

In the language of modal-tense logic, Prior's reconstruction of the premises and the conclusion is:

1. $P\varphi \rightarrow \neg \Diamond \neg P\varphi$ (M1).
2. $\neg \Diamond \psi \rightarrow (\Box(\varphi \rightarrow \psi) \rightarrow \neg \Diamond \varphi)$ (M2).
3. $\varphi \rightarrow HF\varphi$ (P4).
4. $\neg\varphi \wedge \neg F\varphi \rightarrow P\neg F\varphi$ (P5).
5. $\neg\varphi \wedge \neg F\varphi \rightarrow \neg \Diamond \varphi$ (neg of M3).

where the φs and ψs are temporally indefinite propositions, whose truth values can vary over time, and $\varphi \rightarrow \psi$ is Philonean (material) implication, and $\Box(\varphi \rightarrow \psi)$ is Diodorean (strict) implication. Here is how the argument works out:

1. $P\varphi \rightarrow \Box P\varphi$ (Diod. prem. 1)
2. $((\varphi \rightarrow \psi) \wedge \Diamond\varphi) \rightarrow \Diamond\psi$ (Diod. prem. 2)
3. $\neg p \wedge \neg Fp \wedge \Diamond p$ (Diod. prem. 3)

4. $\varphi \rightarrow HF\varphi$ (Prior prem. 1)

5. $(\neg\varphi \wedge \neg F\varphi) \rightarrow P\neg F\varphi$ (Prior prem. 2)

6. $P\neg Fp$ (from 3, 5, by *modus ponens*)

7. $\Box P\neg Fp$ (from 1, 6, by *modus ponens*)

8. $\neg \Diamond \neg P\neg Fp$ (from 7, by def. of $\Box$).

9. $\neg \Diamond HFp$ (from 8, by def. of H).

10. $\neg((p \rightarrow HFp) \wedge \Diamond p)$ (from 2, 9, by modus tollens)

11. $\neg(p \rightarrow HFp) \vee \neg \Diamond p$ (from 10, by de Morgan's)

12. $\neg \Diamond p$ (4, 11, by disjunctive syllogism)

And (12) is a contradiction with (3).

13.4.2 Temporal interpretations of necessity

Can we define necessity in terms of truth at times? One way to read Diodorus's argument is as given a *definition* of possibility:

$$\Diamond\varphi := \varphi \vee F\varphi$$

Correspondingly, necessity becomes, for Diodorus:

$$\Box\varphi := \varphi \wedge G\varphi$$

This definition of necessity should be contrasted with the Megarian-Aristotelian of necessity which takes into account the past as well as the present and the future:

$$\Box\varphi := \varphi \wedge G\varphi \wedge H\varphi$$

In each of his considerations of the Master Argument, Prior was interested in which, if any, of the standard classical modal calculi corresponds to necessity defined this way. In [Prior, 1955], Prior argued that the Diodorean logic Dio was at least as strong as S4, satisfying transitivity, reflexivity, and $\Diamond(\varphi \wedge \psi) \leftrightarrow (\Diamond\varphi \wedge \Diamond\psi)$, but weaker than S5. Reflexivity follows directly from the definition of $\Diamond$; transitivity is a consequence of the tense logic law $FF\varphi \rightarrow F\varphi$, which in turn can be derived from $F(\varphi \wedge \psi) \leftrightarrow (F\varphi \wedge F\psi)$. The final property also follows from the tense logic law $F(\varphi \wedge \psi) \leftrightarrow (F\varphi \wedge F\psi)$.

In [Prior, 1957, Prior, 1967] Prior gave infinite matrices for Diodorean possibility and necessity. Propositions are assigned infinite sequences of 1s and 0s. Then:

- $\neg\varphi$ is 1 where φ is 0 and vice versa.
- $\varphi \wedge \psi$ is 1 where φ and ψ are both 1 and 0 otherwise.
- $\varphi \vee \psi$ is 1 where φ is 1 or ψ is one, and 0 otherwise.
- $\Diamond\varphi$ is 1 anywhere that there is a 1 for φ either there or further to the right, and 0 otherwise.
- $\Box\varphi$ is 0 anywhere that there is a 0 for φ either there or further to the right, and 1 otherwise.

A formula is a theorem if all its sequences always have 1 everywhere. It should be clear that this is simply another way to stipulate a Kripke model with the standard truth conditions for the operators.

This matrix verifies S4 and rejects S5, and is linear, discrete, and reflexive. At the time of publication of [Prior, 1955], it was an open question whether Dio = S4 or not. In [Prior, 1957, p. 23; p. 121, fn. 1], Prior stated (without proof) that it was; however, this is wrong, as is shown in [Prior, 1967, pp. 23ff]. The Diodorean matrix validates $\Diamond\Box\varphi \rightarrow \Box\Diamond\Box\varphi$, which is derivable from S4 + $\Diamond \Box \varphi \rightarrow \Box \Diamond \varphi$, but $\Diamond \Box \varphi \rightarrow \Box \Diamond \varphi$ is not a theorem of S4. That Dio validates $\Diamond \Box \varphi \rightarrow \Box \Diamond \varphi$ follows from the fact that it validates the tense-logical equivalent, $FG\varphi \rightarrow GF\varphi$ (but not the converse, obviously). So, if the system is not S4, and it is not S5, but is between it, what is it? Lemmon suggested adding the following axiom, to obtain the system S4.3:

$$\Box(\Box\varphi \rightarrow \Box\psi) \vee \Box(\Box\psi \rightarrow \Box\varphi)$$

Geach showed that this axiom can be simplified to

$$\Box(\Box\varphi \rightarrow \psi) \vee \Box(\Box\psi \rightarrow \varphi)$$

In turn, this is equivalent to

$$\Diamond\varphi \wedge \Diamond\psi \rightarrow ((\Diamond\varphi \wedge \psi) \vee (\Diamond\psi \wedge \varphi))$$

which follows from the tense-logical axiom

$$F\varphi \wedge F\psi \rightarrow (F(\varphi \wedge \psi) \vee F(\varphi \wedge F\psi) \vee F(F\varphi \wedge \psi))$$

which corresponds to the linearity of the underlying frame.

But S4.3 is still not enough. The last required axiom was discovered by Dummett:

$$\Box(\Box(\varphi \rightarrow \Box\varphi) \rightarrow \Box\varphi) \rightarrow (\Diamond \Box \varphi \rightarrow \varphi)$$

which corresponds to the discreteness of the underlying frame.

S4.3Dum exactly corresponds to the logic validated by the infinite matrices.

Connections between time and modality

Remember that at the time that Prior was investigating this, there weren't really any good semantics for modal logics, so determining axioms and non-theorems had to be done by intuition mostly. In [Prior, 1955], Prior argued that the Diodorean logic Dio was at least as strong as S4, satisfying (W3) $\Diamond\Diamond\varphi \rightarrow \Diamond\varphi$, transitivity, (W2) $\varphi \rightarrow \Diamond\varphi$, reflexivity, and (W2) $\Diamond(\varphi \wedge \psi) \leftrightarrow (\Diamond\varphi \wedge \Diamond\psi)$, and weaker than S5. Reflexivity follows directly from the definition of $\Diamond$. (W2) follows from the tense logic law $F(\varphi \wedge \psi) \leftrightarrow (F\varphi \wedge F\psi)$, which is derivable if we add to our tense logic the rule from $\vdash \varphi$ infer $\vdash G\varphi$. Transitivity is derivable from the tense logic law $FF\varphi \rightarrow F\varphi$, which in turn can be derived from $F(\varphi \wedge \psi) \leftrightarrow (F\varphi \wedge F\psi)$.

This matrix verifies S4 and rejects S5, and is linear, discrete, and irreflexive. (It should be clear that we can interpret these matrices the same way we did the ones given for his logic of contingency (discussed earlier). That is, each point in the matrix represents a point in time, where time is considered to be beginning, linear, discrete, irreflextive, and non-ending. On this view, if we define $\Diamond$ as above, then the possibility relation we recover from this temporal relation *is* reflexive.) If (W3) is replaced with (W4) $\Diamond\neg\Diamond\varphi \rightarrow \neg\Diamond\varphi$, then the resulting system is S5, and Prior argues [Prior, 1955, p. 208] that (W4) is clearly not a theorem of the Diodorean modal logic. At the time of publication of [Prior, 1955], it was an open question whether $D = S4$ or not. In [Prior, 1957], Prior stated (without proof), that it was [p. 23; p. 121, fn. 1]. This is wrong, as is shown in [Prior, 1967, pp. 23ff]: The Diodorean matrix validates $\Diamond\Box\varphi \rightarrow \Box\Diamond\Box\varphi$, which is derivable from S4 + $\Diamond\Box\varphi \rightarrow \Box\Diamond\varphi$, but $\Diamond\Box\varphi \rightarrow \Box\Diamond\varphi$ is not a theorem of S4. That D validates $\Diamond\Box\varphi \rightarrow \Box\Diamond\varphi$ follows from the fact that it validates the tense-logical equivalent, $FG\varphi \rightarrow GF\varphi$ (but not the converse, obviously). So, if the system is not S4, and it is not S5, but is between it, what is it? (It is easy to show that it is not S4.5, which is S4 + $\Box\Diamond\Box\varphi \rightarrow \Box\varphi$, as it turns out that S4.5 = S5.) Lemmon suggested adding the following axiom, to obtain the system S4.3:

$$\Box(\Box\varphi \rightarrow \Box\psi) \vee \Box(\Box\psi \rightarrow \Box\varphi)$$

Geach showed that this axiom can be simplified to

$$\Box(\Box\varphi \rightarrow \psi) \vee \Box(\Box\psi \rightarrow \varphi)$$

In turn, this is equivalent to

$$\Diamond\varphi \wedge \Diamond\psi \rightarrow ((\Diamond\varphi \wedge \psi) \vee (\Diamond\psi \wedge \varphi))$$

which follows from the tense-logical axiom

$$F\varphi \wedge F\psi \rightarrow (F(\varphi \wedge \psi) \vee F(\varphi \wedge F\psi) \vee F(F\varphi \wedge \psi))$$

which corresponds to the linearity of the underlying frame.

But S4.3 is still not enough. The last required axiom was discovered by Dummett:

$$\Box(\Box(\varphi \to \Box\varphi) \to \Box\varphi) \to (\Diamond \Box \varphi \to \varphi)$$

which corresponds to the discreteness of the underlying frame.

The Dummett axiom is equivalent to

$$\Diamond \Box \varphi \wedge \Box(\neg\varphi \to \Diamond(\varphi \wedge \Diamond\neg\varphi))) \to \varphi$$

(which is perhaps easier to see corresponding to discreteness.)

Proof.

- $\Box(\Box(\varphi \to \Box\varphi) \to \Box\varphi) \to (\Diamond \Box \varphi \to \varphi)$
- $\Diamond \Box \varphi \to (\Box(\Box(\varphi \to \Box\varphi) \to \varphi) \to \varphi)$ (since $p \to (q \to r) \leftrightarrow (q \to (p \to r))$.
- $\Diamond \Box \varphi \to (\Box(\neg\varphi \to \neg \Box (\varphi \to \Box\varphi)) \to \varphi)$ (contraposition).
- $\Diamond \Box \varphi \to (\Box(\neg\varphi \to \Diamond(\varphi \wedge \neg \Box \varphi)) \to \varphi)$ (since $\neg \Box (p \to q) \leftrightarrow \Diamond(p \wedge \neg q)$).
- $\Diamond \Box \varphi \to (\Box(\neg\varphi \to \Diamond(\varphi \wedge \Diamond\neg\varphi)) \to \varphi)$ (since $\neg \Box p \leftrightarrow \Diamond\neg p$).
- $(\Diamond \Box \varphi \wedge \Box(\neg\varphi \to \Diamond(\varphi \wedge \Diamond\neg\varphi))) \to \varphi$ (since $p \to (q \to r \leftrightarrow (p \wedge q \to r))$.

□

Diodorean modalities have interest beyond purely historical interest generated from the question of how to interpret Diodorus's argument. They are of interest in modeling the logic of Minkowski space-time, where an event is 'possible' if it can be reached in the future light-cone of the present time, and the future-light cone of the present time includes the present. This is discussed in [Goldblatt, 1992], with an overview in [Uckelman and Uckelman, 2007].

What if we get rid of discreteness? What about linearity?

White raised the question of whether the assumption of discreteness (that is, *Dum*) is in fact necessary to reconstruct the Master Argument [White, 1984]. Trzęsicki proves that any syntactic proof of the Master Argument requiring irreflexivity will also require discreteness [Trzęsicki, 1987]. Trzęsicki uses the version of the argument in [Rescher and Urquhart, 1971, ch. XVII].

Trzęsicki's models are Kripke frames with accessibility relations for both the temporal and the modal operators:

$$\mathfrak{F}^T = \langle T, R_t, R_m \rangle$$

where T is a set of time-points, R_t is the accessibility relation for the temporal operators and R_m for the modal operators, and the logic, the truth conditions, and the definition of model and validity on a model are exactly as expected. He then uses Prior's reconstruction of the argument in modal-tense logic:

1. $P\alpha \to \Box P\alpha$.
2. $\Box(\alpha \to \beta) \to (\neg \Diamond \beta \to \neg \Diamond \alpha)$.
3. $(\neg\alpha \wedge \neg F\alpha) \to \neg \Diamond \alpha$.
4. $\Box(\alpha \to HF\alpha)$.
5. $(\neg\alpha \wedge F\neg\alpha) \to PG\neg\alpha$.

where (3) is the conclusion of the argument, that is, the negation of the third premise. Trzęsicki says that $(\neg p \wedge F\neg p) \to PGp$ corresponds to the class of temporal frames satisfying the condition:

$$\forall t \exists t_1 (t_1 R_t t \wedge \forall t_2 (t_1 R_t t_2 \to (t = t_2 \vee t R_t t_2))) \tag{13.4}$$

and that this is yields (i) backwards seriality. However, this axiom is clearly a typo for $(\neg p \wedge F\neg p) \to PG\neg p$, i.e., substituting p for α in (5). Forwards-linearity and irreflexivity do not follow from this frame condition. However, the assumption of (ii) irreflexivity entails both (iii) forward-linearity ($\forall x \forall y \forall z (xRy \wedge xRz \to yRz \vee zRy \vee y = z)$) and (iv) backwards-discreteness [Trzęsicki, 1987, p. 127].

He does not give a proof that (ii) and (1) imply (iii). I'm not entirely convinced that it does unless we also assume (v) transitivity. A proposed counterexample is in Figure 13.7 (assume that the frame is not transitive, and that the ellipsis stands for an infinite descending R-chain).

At this point it is important to note how what Trzęsicki is doing differs from what Prior did. Prior was interested in the question "If we define necessity in the Diodorean way, what modal logic does this correspond to?" Trzęsicki is asking the question "what temporal-modal logic do we need to validate (Prior's reconstruction of) the Master Argument?". One place where they diverge is the assumption of backwards seriality. This assumption is equivalent to requiring that Prior's matrices be two-way infinite, and he never specifies that this is the case, and in fact, in [Prior, 1957, p. 22], he says that the truth values can

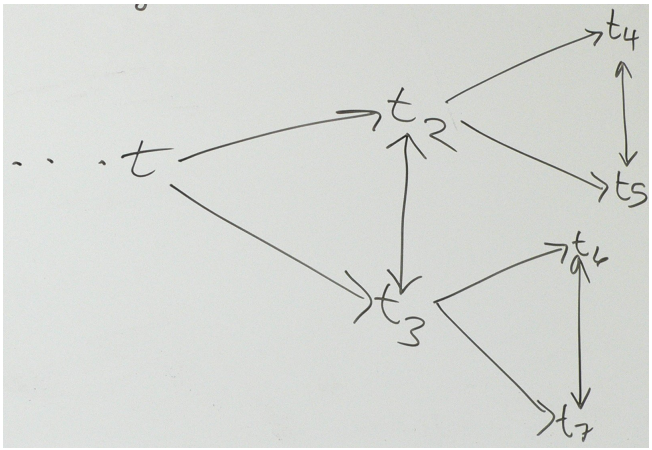


Figure 13.7: Counterexample

be represented as infinite sequences, or "if you like, by non-terminating decimals", which implies that they are not two-way infinite. One interesting question that arises from this is: If the temporal-modal logic needed for validating the Master Argument is *not* S4.3Dum, then what logic is it, and what does this result say about Prior's reconstruction? It seems weird that the logic which arises from the definition of the modalities is not the logic needed to define them in the first place.

The question of how branching future time affects things is discussed in the next section.

Exercise 1. Show that Lemmon (cf. (Lemmon)) is valid on Diodorean frames.

Exercise 2. Prove that S4.3 is a proper extension of S4.2.

Exercise 3 (Hintikka)**.** Prove that Lemmon is equivalent to

$$(\Diamond p \wedge \Diamond q) \rightarrow (\Diamond(p \wedge \Diamond q) \vee \Diamond(q \wedge \Diamond p))$$

Exercise 4. Prove that Lemmon corresponds to linearity of $<$.

13.5 Metric propositional tense logic

Let's return to P1 of the Master Argument: "Every past truth is necessary" / "Every truth about the past is necessary.

- If something happens, then it was true that it would happen: $p \to HFp$. $\vdash p \to HFp$ is a theorem of every minimal tense logic. $HFp \to PFp \to \Box Fp$.
- How can we reconcile this with the contingency of the future: that some things happen that could not have happened? Or: that something will happen that could possibly not happen.
- Recall the Diodorean/A-M model: possibility requires a temporal witness. But if there will be a sea-battle tomorrow, then it's not possible that there not be a sea-battle tomorrow: IF time is linear?
- What do we need to represent what's going on here?
  - Drop linearity, of course: Branching time.
  - "Sea battle tomorrow". *Still* can't do entirely: $F\varphi \wedge F\neg\varphi$ are not contradictory, even in linear time.
- $\Rightarrow$ Extend the minimal tense logic. How?
- "Tomorrow" / "Yesterday" operators. *Indexical.*
- *IF* time is discrete, then $\forall t \in T \exists t'$ such that $t < t'$ and $\forall t'' \neg(t < t'' < t')$. If time is dense, there is no unique "next" moment, so you'd need to augment $\mathfrak{T}$ with something that picks out "next": $\mathfrak{T} = \langle T, <, \mathsf{next}, \mathsf{prev}\rangle$:

$$\begin{array}{lll} t \vDash T\varphi & \text{iff} & \mathsf{next}(t) = t' \text{ and } t' \vDash \varphi \\ t \vDash Y\varphi & \text{iff} & \mathsf{prev}(t) = t' \text{ and } t' \vDash \varphi \end{array}$$

  OR expand the tense operators to take indices as well as props.

$F_n\varphi$: "it will be the case n 'units' from now that φ'.

In the minimal propositional tense logic presented in the previous section, the operators F, G, H, and P are unary, taking as their input a single well-formed formula or proposition. They are not very expressive. We may, however, we be interested in being able to speak more precisely than "will be the case in the future"; we can do so by making the operators dyadic: They take both a proposition and a measure. The measure is used for counting times along the accessibility relation. Thus, the measurement unit is interpreted as whatever is appropriate for the interpretation of T. Thus, if T is a set of days, then $F_n\varphi$ would mean "It will be the case in n days that φ". Formally:

Definition 13.5.1 (Truth conditions for metric tense operators)**.**

$$\begin{array}{lll}
\mathfrak{M}, t \vDash F_n\varphi & \text{iff} & \text{there is a sequence } t, t_1, \ldots, t_n \\
 & & \text{where } t < t_1 < \cdots < t_{n-1} < t_n \text{ and } \mathfrak{M}, t_n \vDash \varphi \\
\mathfrak{M}, t \vDash P_n\varphi & \text{iff} & \text{there is a sequence } t, t_1, \ldots, t_n \\
 & & \text{where } t_n < t_{n-1} < \cdots < t_1 < t \text{ and } \mathfrak{M}, t_n \vDash \varphi \\
\mathfrak{M}, t \vDash G_n\varphi & \text{iff} & \text{for every sequence } t, t_1, \ldots, t_n, \\
 & & \text{if } t < t_1 < \cdots < t_{n-1} < t_n \text{ then } \mathfrak{M}, t_n \vDash \varphi \\
\mathfrak{M}, t \vDash H_n\varphi & \text{iff} & \text{for every sequence } t, t_1, \ldots, t_n, \\
 & & \text{if } t_n < t_{n-1} < \cdots < t_1 < t \text{ then } \mathfrak{M}, t_n \vDash \varphi
\end{array}$$

Note 13.5.2. These operators coincide in any structure where there is a unique path between any two points (i.e., ones that are discrete, linear, non-transitive, ...)

n-chains: Two ways to think of metric tenses: over n-chains of instants *or* as quantifying over *intervals* of a certain size. (We'll come back to interval tense logic in §13.7.

SO we can say what it means for a sea battle to happen or not tomorrow. This new language is *a lot* more expressive.

Often it is convenient to let $T = \mathbb{N}, \mathbb{Z}, \mathbb{Q}$, or $\mathbb{R}$. Then we have a two-sorted language with the quantifiers ranging over T, and the logic being otherwise propositional. We can recover the monadic operators from the dyadic ones straightforwardly: $F\varphi \coloneqq \bigvee_n F_n\varphi$ (that is, if $F\varphi$ is true now, then either φ will be the case in 1 day (i.e., tomorrow), or 2 days, or 3 days, etc.) Given the definability of G in terms of F it follows that $G\varphi \coloneqq \bigwedge_n F_n\varphi$. (If we don't want to allow infinite conjunctions and disjunctions, then we need to add quantification over n). That is, we have the following definitions:

$$\begin{array}{lllll}
F_0\varphi & \coloneqq & \varphi & & \\
F\varphi & \coloneqq & \bigvee_n F_n\varphi & \coloneqq & \exists n F_n\varphi \\
G\varphi & \coloneqq & \bigwedge_n F_n\varphi & \coloneqq & \forall n F_n\varphi \\
P_n\varphi & \coloneqq & F_{-n}\varphi & & \text{if } W \text{ contains negative numbers.}
\end{array}$$

Definition 13.5.3 (Axioms of linear metric tense logic)**.** Let n range over positive and negative integers; then $P_n p$ can be defined as $F_{-n}p$. The following rule and axioms characterise the minimal linear metric tense logic:

$$\text{If } \vdash \varphi \text{ then } \vdash F_n\varphi \qquad (\mathsf{RF})$$

$\mathsf{F0}$ $F_o p \to p$

$\mathsf{FN1}$ $F_n \neg p \to \neg F_n p$

FN2 $\neg F_n p \to F_n \neg p$

FC $F_n(p \to q) \to (F_n p \to F_n q)$

FF $F_m F_n p \to F_{m+n} p$

F$\forall$ $\forall n F_m F_n p \leftrightarrow F_m \forall n F_n p$

These axioms are due to [Prior, 1967, p. 97], with the exception that his F$\forall$ is only a single implication (from left to right) rather than an equivalence; however, assuming time is linear, the converse is also valid.

But now we have have a question specifically having to do with branching futures.

Let us *not* fix a date. Let us consider: "The UK will leave the EU." March 29? June 30? Never?

In everyday English, 'will' without further qualification is generally taken to indicate future possibility: If I say that "It will rain next week", I am not saying that it will rain continuously the entire week, but that it will rain at least once. Similarly, if I say that "I will take a nap tomorrow", this doesn't mean that I will be sleeping day and night.

As soon as we consider multiple possible futures, there arises a question about the interpretation of 'will'. Recall the truth conditions for $F\varphi$ given in Definition 13.2.3:

$$\mathfrak{M}, w \vDash F\varphi \quad \text{iff} \quad \text{there is } w' \text{ such that } wRw' \text{ and } \mathfrak{M}, w' \vDash \varphi$$

In the presence of multiple futures, this corresponds to a very weak sense of 'will-possibly': It simply says that in *one* of the future time-lines of now, there is *one* time where φ is true. This very weak sense of 'will' captures at least part of the sense of English 'might' (used non-epistemically), e.g., "It *might* rain next week" vs. "It *will* rain next week."

Do we want there to be a date in every possible future? Strong will vs. weak might.

HISTORIES

Histories are maximal linear subsets of trees. The set of histories of T is $H(T)$. If $t \in T$, then $H(t)$ is all histories containing t.

(1) $t \vDash F\varphi$ iff $\forall h$ in $H(t)$ there is a t' in h such that $t < t'$ and $t' \vDash \varphi$

(2) $t \vDash F\varphi$ iff $\exists h$ in $H(t)$ and there is a t' in h such that $t < t'$ and $t' \vDash \varphi$

(3) $t \vDash F\varphi$ iff "thin red line"

Also t/h pairs.

Metric tenses/branching time $\Rightarrow$ $F_n\varphi$ and $G_n\varphi$ are quantifying over histories.

Once we branch futures, we need to be able to talk about what happens "at the same time". $\Rightarrow$ clock ticks/ equivalence classes of instants give rise to "moments". Simultaneity in branching time; sorting temporal moments into instants via equivalence classes.

Exercise 5. Define truth conditions for Gp where G is interpreted as:

1. Weak will-always.
2. Strong will-always.
3. 'Thin red line' will-always.

Exercise 6. Define truth conditions for metric tenses on branching time models with histories.

Exercise 7. Instead of defining truth conditions for 'Yesterday' and 'Tomorrow' by adding prev and next functions to the model, for a given temporal frame $\mathfrak{T} = \langle T, < \rangle$, define a second relation R on T that holds between t and t' if t is yesterday of t' and t' is tomorrow of t.

Exercise 8. What other temporal operators can you define? E.g., "every week".

Exercise 9. Consider a specific temporal issue (e.g., free action; divine foreknowledge, determinism; the sea battle; Brexit; time-travel; Bede) and define a language and constraints on models that will allow you to model this issue.

Exercise 10. What (if anything) can you say with intervals that you cannot say with instants?

13.6 Future contingents

According to the square of oppositions, exactly one of "it is the case that p" and "it is not the case that p" is true. That is, *either* "it is the case that there will be a sea battle tomorrow" *or* "it is not the case that there will be a sea battle tomorrow".

This is problematic for existence of free will, and for Aristotelian metaphysics. The issues here can also be found in the **Master Argumnet** of Diodorus Cronus, as not all past-tensed statements are about the past (ref. Reichenbach above).

Adding an omniscient God

Adding an omniscient God is, by itself, not sufficient to draw the problem. For one could say that future contingents cannot be known until their truth value is settled (i.e., they are no longer contingent), and thus there is no problem with saying that God does not know future contingents, since there is nothing to know. The problem arises when you add that God, on traditional definitions, must also be unchanging. This means that his knowledge cannot

change, either. This means that he cannot learn new propositions; if he doesn't know a future contingent statement, then he can never know it, even if it becomes non-contingent, because that would violate his unchangingness. But then there are propositions which he does not know, and that is a violation of his omniscience.

13.6.1 Metric tenses and modality

If we allow that $\neg F_n\varphi \leftrightarrow F_n\neg\varphi$, then if we define the modal operators as follows:

$$\begin{aligned} \Box\varphi &:= \varphi \wedge G\varphi \\ \Diamond\varphi &:= \varphi \vee F\varphi \end{aligned}$$

Then the system is genuinely modal, and at least S4.

The future-only case has the rule of inference RF: if $\vdash \alpha$ then $\vdash F_n\alpha$, and the following axioms:

FN1 $F_n\neg\varphi \to \neg F_n\varphi$.

FN2 $\neg F_n\varphi \to F_n\neg\varphi$.

FC $F_n(\varphi \to \psi) \to (F_n\varphi \to F_n\psi)$.

FO $F_0\varphi \to \varphi$.

FF $F_m F_n\varphi \to F_{m+n}\varphi$.

F$\forall$ $F_m \forall n F_n\varphi \to \forall n F_m F_n\varphi$.

It is easy to extend these postulates to the case with negative numbers (e.g., replace FO with $\forall n(F_n\varphi \to \varphi)$.)

Question 13.6.1. What type of logic do you get if you drop FN1 or FN2 or both?

13.7 Interval semantics

1. Two ways to think about metrics: as counting discrete steps or as quantifying over intervals.

2. Two ways to think about truth at at time. For every tensed proper there is a tenseless one.

3. Talk about time structures directly.

(2) Gives rise to hybrid logic and closely linked to (3): "true at t", "t' before t"
(1) Approaches to intervals.

- Why consider intervals instead of instants?
  - Events have duration.
  - Medieval tokens
  - "While X, Y", "When X, Y" $\Rightarrow$ not to be confused with the "When X, *do* Y" of programming logics like PDL.
- Are intervals made up of instants?
- Is there anything you can say with only intervals you wouldn't say with instants?
- Are we interested in describing interval *structure* or in describing interval *truth*?

We'll look at both. LINEAR interval temp logic.

Definition 13.7.1. Let $\mathbb{D} = \langle D, < \rangle$ where D is $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{Z}$, or $\mathbb{R}$ and $<$ the usual ordering. An *interval* on $\mathbb{D}$ is a pair $[d, d']$, where $d, d' \in D$ and $d \leq d'$. $\rightarrow$ for now, *closed* intervals only.

We also have $d \leq d^* \leq d'$ iff $d^* \in [d, d']$
We sometimes write $I(d, d')$ for $[d, d']$. The set of intervals of $\mathbb{D}$ is $\mathbb{I}(\mathbb{D})$.

Definition 13.7.2. An interval is *strict* if $d < d'$ and a *point* or *isntant* (or *nonstrict*) if $d = d'$ (so: can always recapture point-based semantics!). Points are degenerate intervals.

Some intervals are subintervals of each other:

$$[d_m, d_n] \subseteq [d_k, d_l] \text{ iff for all } d \in [d_m, d_n], d_k \leq d \leq d_l$$

We can also order intervals:

$$[d_k, d_l] \lhd [d_m, d_n] \text{ iff } d_l < d_m$$

$$[d_k, d_l] + [d_m, d_n]$$

Definition 13.7.3. An *interval structure* is a pair $\mathcal{D} = \langle \mathbb{D}, \mathbb{I}(\mathbb{D}) \rangle$s.

Valuations on *interval structures*. What do we want with truth?

V makes atoms to sets of intervals.

$$\begin{aligned} V(\neg\varphi) &= \{a : \forall b \subseteq a(b \notin V(\varphi))\} \\ V(\varphi \wedge \psi) &= V(\varphi) \cap V(\psi) \\ V(G\varphi) &= \{a : \forall b, c((b \subseteq a) \wedge b \lhd c) \rightarrow c \in V(\varphi)\} \\ V(H\varphi) &= \{a : \forall b, c((b \subseteq a) \wedge c \lhd b) \rightarrow c \in V(\varphi)\} \end{aligned}$$

($F\varphi$ defined $\neg G\neg\varphi$.)

What do we want with truth?

Two possibilities:

Definition 13.7.4. φ is *persistent* over an interval $[d_m, d_n]$ if $[d_m, d_n] \vDash \varphi$ implies $[d_k, d_l] \vDash \varphi$ for every subinterval of $[d_m, d_n]$.

Definition 13.7.5. φ is *cumulative* over two overlapping or abutting intervals $[d_m, d_n]$ and $[d_k, d_l]$ if $[d_m, d_n] \vDash \varphi$ and $[d_k, d_l] \vDash \varphi$ implies $[d_m, d_n] + [d_k, d_l] \vDash \varphi$.

Definition 13.7.6. If φ is both persistent and cumulative then it is *homogeneous*.

Any truth value assignment where every sentence is homogeneous is called homogeneous.

What contexts would non-homogeneous valuations be suitable for? Buridan's *Sophismata*.

Even if V is not homogeneous, we might want to have something to pickout homogeneous sentences:

$$[d_m, d_n] \vDash O\varphi \quad \text{iff} \quad \forall a \subseteq [d_m, d_n] \exists c \subseteq a, c \vDash \varphi$$

"$O\varphi$ is true over an interval x iff φ is true for some subinterval z of every subinterval y of x." $O\varphi$ is homogeneous.

NOTE: $O\varphi \wedge O\neg\varphi \Rightarrow$ not a contradiction; but shouldn't these be contraries?

Two types of negation:

- Sentence negation: "It is not the case that Socrates is sitting."
- Term/atomic negation: "Socrates is not sitting."

$V(\bar{p}) := \{[d_m, d_m] : [d_m, d_m] \notin V(p)\}$

- Linguistics
  - Analyses of natural languages are generally interval-based (Reichenbach, Buridan).

- AI
  - Interval-based approaches are used in *expert systems*, *planning systems*, *theories of actions and change*, *natural language analysis and processing*: Use for temporal representation and reasoning in artificial/formalized agents.
- Computer science
  - Used in specification and design of hardware components. *Duration calculi* are used for specification and verification of real-time processes; connections to Φ and Ψ.

Definition 13.7.7. Given a strict partial (transitive and asymmetric) ordering $\langle \mathbb{D}, < \rangle$, an *interval* in $\mathbb{D}$ is a pair $[q_0, q_1]$ such that $q_0, q_1 \in \mathbb{D}$ and $q_0 \leq q_1$. It is a *strict interval* if $q_0 < q_1$. $I(\mathbb{D})$ is the set of intervals of $\mathbb{D}$; $I(\mathbb{D})^+$ is the set of non-strict intervals, and $I(\mathbb{D})^-$ the set of strict intervals.

We may be interested in interval frames with the following properties:
linear interval property:

$$\forall xy(x < y \to \forall z_1, z_2(x < z_1 < y \wedge x < z_2 < y \to z_1 < z_2 \vee z_2 < z_1 \vee z_1 = z_2)$$

linear: $\forall xy(x < y \vee y < x \vee y = x)$
discrete:

$$\forall xy(x < y \to \exists z(x < z \wedge z \leq y \wedge \forall w(x < w \wedge w \leq y \to z \leq w)))$$

dense: $\forall xy(x < y \to \exists z(x < z \wedge z < y))$
unbounded above: $\forall x \exists y(x < y)$
Dedekind complete: Every non-empty and bounded above set of points has a l.u.b.

Example 13.7.8. $\mathfrak{I}(\mathbb{Q}) = \langle \mathbb{Q}, O_+(\mathbb{Q}) \rangle$ where $O_+(\mathbb{Q})$ is the set of all non-empty open rational intervals.

Example 13.7.9. $\mathfrak{I}(\mathbb{Z}) = \langle \mathbb{Z}, C_+(\mathbb{Z}) \rangle$ where $C_+(\mathbb{Z})$ is the set of all non-empty closed integer intervals.

We can also identify a number of relations on intervals. For a partial order $\mathbb{D}$,

- Define the precedence relation $\prec$ on $I(\mathbb{D})$:

$$[d_0, d_1] \prec [e_0, e_1] \quad \text{iff} \quad d_1 < e_0$$

The non-strict version $\preceq$ is defined as expected.

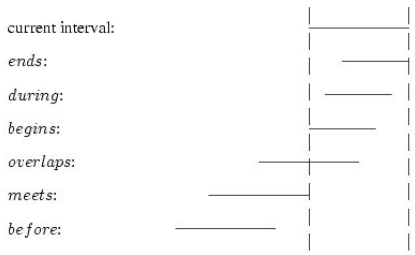


Figure 13.8: Allen's 13 relations [Goranko et al., 2004, p. 16]

- Define an inclusion relation $\sqsubseteq$ on $I(\mathbb{D})$ satisfying the following properties:

$$\begin{aligned} &\forall x \forall y \forall z (x \sqsubseteq y \sqsubseteq z \rightarrow x \sqsubseteq z) && \text{(transitivity)} \\ &\forall x (x \sqsubseteq x) && \text{(reflexivity)} \\ &\forall xy (x \sqsubseteq y \subseteq x \rightarrow x = y) && \text{(anti-symmetry)} \end{aligned}$$

  That is, $\sqsubseteq$ will generally induce a partial order on $I(\mathbb{D})$.

Definition 13.7.10. $[s_0, s_1]$ is a *sub-interval* of $[d_0, d_1]$ if $d_0 \leq s_0$ and $s_1 \leq d_1$.

Definition 13.7.11. $[s_0, s_1]$ is a *proper sub-interval* of $[d_0, d_1]$, denoted $[s_0, s_1] \sqsubset [d_0, d_1]$ if $[s_0, s_1] \sqsubseteq [d_0, d_1]$ and $[s_0, s_1] \neq [d_0, d_1]$.

Definition 13.7.12. $[s_0, s_1]$ is a *strict sub-interval* of $[d_0, d_1]$, denoted $[s_0, s_1] \sqsubset [d_0, d_1]$ if $d_0 < s_0$ and $s_1 < d_1$.

There are 13 different binary relations between any two intervals i, j on a linear ordering: *equals*, *ends*, *during*, *begin*, *overlaps*, *meets*, *before* and their converses. $\Rightarrow$ connections to Interval Algebras (cf. [Chittaro and Montanari, 2000]). These relations are represented graphically in Figure 13.8 (only the first seven are pictured; the converses can be extrapolated).

And we introduce one ternary relation:

Definition 13.7.13. For intervals i, j, k, $Aijk$ holds if $i_0 = k_0$, $j_1 = k_1$, and $\neg \exists d$ s.t. $i_1 < d < j_0$.

$Aijk$ is to be read 'k is the concatenation of i and j'.

Interval models

Note 13.7.14. There are two approaches to interval models:

- Take intervals as primitive, and evaluate formulas against intervals.
- Take intervals as defined by their endpoints.

Definition 13.7.15. An *interval Kripke model* (or *interval model*) is a tuple $\mathfrak{M} = \langle \mathfrak{I}, V \rangle$, where $\mathfrak{I}$ is an interval frame, and V is a valuation function.

Definition 13.7.16. An interval semantics is *strict* if it does not allow point-intervals, and *non-strict* if it does.

Definition 13.7.17. A *strict interval model* is an interval model $\mathfrak{M}^+$ where $V : I(\mathbb{D})^+ \to \mathcal{P}(\text{PROP})$, that is, each strict interval is assigned a set of atomic propositions considered true in the model.

Definition 13.7.18. A *non-strict interval model* is an interval model $\mathfrak{M}^-$ where $V : I(\mathbb{D})^- \to \mathcal{P}(\text{PROP})$.

There are two natural restrictions on valuations:

Definition 13.7.19. An interval semantics is *local* if all atomic propositions are point-wise and truth at an interval is defined as truth at its initial point.

Definition 13.7.20 (Homogeneity)**.** An interval semantics is *homogeneous* if the truth of a formula in an interval implies the truth of the formula in every sub-interval of that interval.

Note 13.7.21. Homogeneity is concerned with *downwards* truth, so it would be better to call it "downwards homogeneity". There is very little consideration in the literature of what might be called "upwards homogeneity".

> Should the truth of φ extend downward, to subperiods—or maybe upward, to superperiods?. . . [T]he upward direction seems less relevant; and hence attention will be restricted to the [downward] case [van Benthem 1983, p. 193–194].

As we'll see later, there are good reasons to not just dismiss the upward case.

Allen's relations can be defined in terms of two primitive operators $\langle B\rangle\varphi$ 'begins φ' and $\langle E\rangle\varphi$ 'ends φ' and their inverses.

$$\begin{array}{lll}
\mathfrak{M}^+, [d_0, d_1] \vDash \langle B\rangle\varphi & \text{iff} & \mathfrak{M}^+, [d_0, d_2] \vDash \varphi \text{ for some } d_2 \text{ s.t. } d_0 \leq d_2 < d_1 \\
\mathfrak{M}^+, [d_0, d_1] \vDash \langle E\rangle\varphi & \text{iff} & \mathfrak{M}^+, [d_2, d_1] \vDash \varphi \text{ for some } d_2 \text{ s.t. } d_0 < d_2 \leq d_1 \\
\mathfrak{M}^+, [d_0, d_1] \vDash \overline{\langle B\rangle}\varphi & \text{iff} & \mathfrak{M}^+, [d_0, d_2] \vDash \varphi \text{ for some } d_2 \text{ s.t. } d_1 < d_2 \\
\mathfrak{M}^+, [d_0, d_1] \vDash \overline{\langle E\rangle}\varphi & \text{iff} & \mathfrak{M}^+, [d_2, d_1] \vDash \varphi \text{ for some } d_2 \text{ s.t. } d_2 < d_0
\end{array}$$

$[B], [E], \overline{[B]}, \overline{[E]}$ are defined as expected.

Note 13.7.22. How would you define Φ and Ψ given $\langle B\rangle, \langle E\rangle$ and their converses?

Relating to the ternary relation A we can define three binary modalities C, D, T [Venema 1991]. For intervals i, j, k,

$$\begin{array}{lll}
\mathfrak{M}^+, k \vDash \varphi C \psi & \text{iff} & \exists i, j, Aijk, \mathfrak{M}^+, i \vDash \varphi \text{ and } \mathfrak{M}^+, j \vDash \psi \\
\mathfrak{M}^+, j \vDash \varphi D \psi & \text{iff} & \exists i, k, Aijk, \mathfrak{M}^+, i \vDash \varphi \text{ and } \mathfrak{M}^+, k \vDash \psi \\
\mathfrak{M}^+, i \vDash \varphi T \psi & \text{iff} & \exists j, k, Aijk, \mathfrak{M}^+, j \vDash \varphi \text{ and } \mathfrak{M}^+, k \vDash \psi
\end{array}$$

In non-strict semantics, it is sometimes useful to be able to refer to point-intervals specifically:

$$\mathfrak{M}^+, [d_0, d_1] \vDash \pi \quad \text{iff} \quad d_0 = d_1$$

If our language includes $[B]$ or $[E]$, then we can define π as $[B]\,\bot$ or $[E]\,\bot$.

If our language includes π (either as a primitive or derived), then we can interpret the strict semantics in the non-strict one:

Definition 13.7.23. The translation τ is defined recursively:

- $\tau(p) = p$ for $p \in \mathrm{PROP}$.
- $\tau(\neg\varphi) = \neg\tau\varphi$.
- $\tau(\varphi \wedge \psi) = \tau\varphi \wedge \tau\psi$.
- $\tau(\langle *\rangle\varphi) = \langle *\rangle(\neg\pi \wedge \tau\varphi)$ for any (unary) interval diamond-modality $\langle *\rangle$.

Proposition 13.7.24. *For every interval model $\mathfrak{M}$, proper interval $[d_0, d_1] \in \mathfrak{M}$, and formula φ,*

$$\mathfrak{M}^-, [d_0, d_1] \vDash \varphi \quad \textit{iff} \quad \mathfrak{M}^+, [d_0, d_1] \vDash \tau(\varphi)$$

Proof. By induction on φ. □

Some monadic interval logics

The weakest monadic interval logic is the sub-interval logic D.

Syntax:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid \langle D\rangle\varphi \mid \pi$$

Semantics:

$$\mathfrak{M}, [d_0, d_1] \vDash \langle D\rangle\varphi \quad \text{iff} \quad \exists [d_2, d_3] \prec [d_0, d_1], \mathfrak{M}, [d_2, d_3] \vDash \varphi$$

where $\prec$ is any of $\sqsubset, \sqsubseteq, \sqsupset$.

Note 13.7.25. As of [§3.1.1]GMS, it is not known how D is axiomatized, what its expressivity is, or whether any fragments of it are decidable.

The logic B$\overline{\text{B}}$:

Syntax

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid \langle B\rangle\varphi \mid \overline{\langle B\rangle}\varphi$$

This logic can be translated into linear tense logic by any mapping τ that replaces $\langle B\rangle$ with F and $\overline{\langle B\rangle}$ with P.

Definition 13.7.26. For any interval model $\mathfrak{M}$ and $d \in D$, the model $\mathfrak{M}[d) = \langle [d), V'\rangle$ where $[d) = \{d' \in D | d \leq d'\}$, and V' is defined as follows: $\forall d' \in [d)$ and $p \in \text{PROP}$, $p \in V'((d')$ iff $p \in V([d, d'])$.

Every point-based tense logic model which is linear and has a least element can be defined from an interval model in this fashion.

Lemma 13.7.27. *For every interval model $\mathfrak{M} = \langle \mathbb{D}, V\rangle$ of* B$\overline{\text{B}}$*, $d \in D$, and $\varphi \in$* B$\overline{\text{B}}$*,*

$$\mathfrak{M}, [d, d'] \vDash \varphi \quad \textit{iff} \quad \mathfrak{M}[d), d' \vDash \tau\varphi$$

for any $d' \geq d$.

Proof. By induction on the complexity of φ. □

13.7.1 Buridan and intervals

The seventh chapter of Jean Buridan's *Sophismata* is a collection of 9 sophisms relating to time, tense, and truth value of propositions. Each sophism establishes a claim, which builds upon the claims established by the previous sophism, with all 9 working together to create a sophisticated approach to tensed propositions in an interval-based setting.

Relative vs. *Absolute* past and future.

Sophism 1

No spoken proposition is true.

This is argued for by noting that propositions are only true or false when they exist, and by the time the predicate of the proposition has been spoken the subject no longer exists, so there is never any proposition the parts of which all exist at the same time which can be true or false.

Buridan rejects this by noting that we don't have to evaluate truth against an instant, but rather we should use an interval:

> I say that it is not determined for us how much time we ought to use as the present, but we may use as much as we want. For we call this year the present, and this day the present, and this hour the present, and if we use this day as the present, then the first hour is and the noon hour is and the vesper hour is, but successively. If, however, we use only the noon hour as the present, then we say that the first hour is past and is no longer, and the vesper hour is future and not yet is [Buridan and Klima, trans., 2001, p. 942]

(It is interesting to compare this sophism with sophism 1 of chapter 1, which argues that "everything spoken proposition is true").

Sophism 2

Nobody can contradict my proposition.

Contradictions must be related to the same time. This requires the intention of the speaker.

Sophism 3

The proposition 'Socrates is sitting' is true at a time throughout which Socrates is not sitting.

This sophism introduces the distinction between "true at" and "true of" or "true for" a time. A propositional letter will be true at instants, but its truth is of or for intervals. Something which is true at some time will be true for any interval containing that time.

Sophism 4

This conjunction is true. 'Socrates is sitting and Socrates is not sitting'.

Negation distributes the verb, so it applies to the whole of the interval; thus, "Socrates is sitting" does not imply "Socrates is not standing". Also, "Socrates is non-sitting", an affirmative with an infinite predicate, does not imply "Socrates is not sitting", a negative

with a finite predicate. Thus, this sophism is an argument for upwards homogeneity of truth and downwards homogeneity of falsity, in contrast with usual modern approaches (cf. Def. 13.7.20).

Sophism 5

This conjunction is true: 'Aristotle argues and Antichrist preaches'.

This sophism is linguistically based: It takes seriously the (linguistic) tense of a proposition. Modern philosophy often interprets present-tensed statements as being timeless, but medievals took the tenses in propositions at face value.

Sophism 6

At every time Socrates is running.

If the verb has no ampliating force, then "at every time" is equivalent to "at every present time", since it's only with ampliation that we can talk about something other than the present. By convention, we often use "always" and "every time" *as if* they ampliate to all times, but strictly speaking this is incorrect. Only the verb ampliates. (This is different from Lambert of Lagny's approach, where an adverb like "always" does ampliate to all times.)

Sophism 7

The same written or spoken prposition is true and false for people of the same language and without a new imposition or obligation.

This sophism indicates that the duration of the present is conventional, and person-indexed. See Sophism 2.

Sophism 8

Whatever moves moved earlier.

This introduces the distinction between relative and absolute tenses, i.e., between the A-series and the B-series notions. Something can be absolutely present, but relatively past within an interval.

Sophism 9

No change is instantaneous.

Since and Until	Metric Tense Logic	U-calculus
$\varphi\Phi\psi$	$\exists n P_n\varphi \wedge \forall m(m < n \to P_m\psi)$	$T_a\varphi\Phi\psi \leftrightarrow \exists b U b a T_b\varphi \wedge$ $\forall c(Ubc \wedge Uca \to T_c\psi)$.
$\varphi\Psi\psi$	$\exists n F_n\varphi \wedge \forall m(m < n \to F_m\psi)$	$T_a\varphi\Psi\psi \leftrightarrow \exists b U a b T_b\varphi \wedge$ $\forall c(Uac \wedge Ucb \to T_c\psi)$.

Table 13.9: Ways to define 'Since' and 'Until'

13.7.2 Branching interval structures

Exercise 11. What happens if p is true at an interval without it being true in any subinterval?

13.8 *Since*, *Until*, and *While*

Hans Kamp (1966/1968) introduced the operators Φ 'since' and Ψ 'until'. Informally, $\varphi\Phi\psi$ 'Since φ, ψ' is true iff ψ has been true since φ was true, i.e., there is some point in the past where φ was true, ψ was true at that point and has been true ever since. (Ψ is just the future tense version.)

- The concepts "since" and "until" are used a lot in computer science, e.g., for program specification and verification.
- Φ and Ψ cannot be defined in terms of F, P, G, H, though the opposite is true: $P\varphi \equiv \varphi\Phi(\varphi \to \varphi)$.
- But Φ and Ψ *can* be defined in terms of the metric tenses, and, more generally in the U calculus, as demonstrated in Table 13.9

13.8.1 Language

In this section, we look at the minimal postulates for 'since' and 'until', as well as for another natural natural-language temporal connective, 'while'. Instead of using Kamp's Φ and Ψ (which could introduce confusion given that we use capital Greek letters for sets of formulas), we will define new notation.

Definition 13.8.1. We introduce three new binary operators, $\mathcal{S}$ 'since', $\mathcal{U}$ 'until', and $\mathcal{Q}$ 'while'.[4]

[4]For reasons that will become clear below, we pick $\mathcal{Q}$ as it is the first letter of Latin *quando* 'while'.

13.8.2 Proof theory

The minimal postulates for $\mathcal{S}$ and $\mathcal{U}$ are:

R1 if $\vdash \alpha$ then $\vdash \neg(\neg\alpha\mathcal{S}\beta)$.

R2 if $\vdash \alpha \to \beta$ then $\vdash \gamma\mathcal{S}\alpha \to \gamma\mathcal{S}\beta$.

A1 $\neg(\neg(\varphi \to \psi)\mathcal{S}\chi) \to (\varphi\mathcal{S}\chi \to \psi\mathcal{S}\chi)$.

> "If it's not the case that since $\neg(\varphi \to \psi)$ was true χ has been true, then if χ has been true since φ, it's also been true since ψ."

A2 $(\neg(\neg\varphi\mathcal{U}\psi)\mathcal{S}\psi) \to \varphi$.

> "If since it's not been the case that ψ is true until φ is false, ψ has been true, then φ is true."

It is possible to prove the U counterparts of these postulates without putting any conditions on U, and the results, along with their mirror images, give rise to $\mathrm{K_t}$. In order to do so, we'll need the following theorems:

T1 $(\varphi \to \varphi) \to (\psi \to \psi)$ [PL].

T2 $\chi\mathcal{S}(\varphi \to \varphi) \to \chi\mathcal{S}(\psi \to \psi)$ [T1, R2].

T3 $\neg(\neg(\varphi \to \psi)\mathcal{S}(\chi \to \chi)) \to ((\varphi\mathcal{S}(\chi \to \chi)) \to \psi\mathcal{S}(\chi \to \chi))$ [A1].

T4 $\neg(\neg\varphi \to \psi\mathcal{S}\neg(\varphi \to \varphi)) \to ((\varphi\mathcal{S}(\varphi \to \varphi)) \to \psi\mathcal{S}(\psi \to \psi))$ [T3,T2].

T5 $\neg P\neg(\varphi \to \psi) \to (P\varphi \to P\psi)$ [T4, dfn. P].

T6 $((\neg\varphi\mathcal{S}(\psi \to \psi))\mathcal{S}(\psi \to \psi)) \to \varphi$ [A2].

T7 $(\neg(\neg\varphi\mathcal{U}(\neg\varphi \to \neg\varphi))\mathcal{S}(\psi \to \psi)) \to \varphi$ [T6, T2].

T8 $(\neg(\neg\varphi\mathcal{U}(\neg\varphi \to \neg\varphi))\mathcal{S}\neg(\neg\varphi\mathcal{U}(\neg\varphi \to \neg\varphi)) \to \neg(\neg\varphi\mathcal{U}(\neg\varphi \to \neg\varphi))) \to \varphi$ [T8, T2].

T9 $P\neg F\neg\varphi \to \varphi$ [T8, dfn. F, P].

RH $\vdash \alpha$ then $\vdash \neg(\neg\alpha\mathcal{S}(\neg\alpha \to \neg\alpha))$ and $\vdash \neg P\neg\alpha$ [dfn. P].

Note 13.8.2.

$$\begin{aligned} T_a\alpha\mathcal{S}\beta &\equiv \exists b(Uba \wedge T_b\alpha \wedge \forall c(Ubc \to (Uca \to T_c\beta))) \\ T_a\neg\alpha\mathcal{S}\beta &\equiv \exists b(Uba \wedge \neg T_b\alpha \wedge \forall c(Ubc \to (Uca \to T_c\beta)) \end{aligned}$$

The U-counterparts of the rules are as follows:

UR1 for $a \notin \alpha$, if $\vdash T_a\alpha$ then $\vdash T_b\alpha$, that is,

$$\vdash Uba \to (\forall c(Ubc \to (Uca \to T_c\beta)) \to T_b\alpha).$$

UR2 if $\vdash T_a(\alpha \to \beta)$ then $\vdash T_a(\gamma \mathcal{S} \alpha \to \gamma \mathcal{S} \beta)$, that is,

$$\vdash \forall b(Uba \to (T_b\gamma \to (\forall c(Ubc \to (Uca \to T_c\alpha))))) \to \exists d(Uda \wedge T_d\gamma \wedge \forall e(Ude \to (Uea \to T_e\beta))).$$

And the U-counterparts of the axioms are as follows:

T_a**A1**

$$\forall b(Uba \to (T_b\varphi \to \forall c(Ubc \to (Uca \to T_c\chi)))) \to (\forall d(Uda \to \forall e(Ude \to (Uea \to T_e\chi)) \to (T_d\varphi \to T_d\psi))) \to \exists f(Ufa \wedge T_f\psi \wedge \forall g(Ufg \to (Uga \to T_g\chi)))$$

T_a**A2**

$$\exists b\big(Uba \wedge \forall d(Ubd \to (\forall e(Ube \to (Ued \to T_e\psi)) \to T_d\varphi)) \wedge \forall c(Ubc \to (Uca \to T_c\psi))\big) \to T_a\varphi$$

Proof of T_aA1. 1. Uba (assumption).

2. Tb_φ (assumption).

3. $\forall e(Ube \to (Uea \to T_e\chi))$ (assumption).

4. $\forall d((Uda \to \forall e(Ude \to (Uea \to T_e\chi))) \to (T_d\varphi \to T_d\psi))$ (assumption).

5. $Uba \to \forall e(Ube \to (Uea \to T_e\chi))) \to (T_b\varphi \to T_b\psi))$ (UI, 4).

6. $T_b\psi$ (1 and 3 as antecedent of 5; MP; MP with 2).

7. $\forall g(Ubg \to (Uga \to T_g\chi)$ (3, variables rewritten).

8. $Uba \wedge T_b\psi \wedge \forall g(Ubg \to (Uga \to T_g\chi))$ ($\wedge$I, 1,6,7).

9. $\exists f(Ufa \wedge T_f\psi \wedge \forall g(Ufg \rightarrow (Uga \rightarrow T_g\chi)))$ (EI, 8).

□

Proof of T_aA2. 1. Uba (assumption).

2. $\forall d(Ubd \rightarrow (\forall e(Ube \rightarrow (Ued \rightarrow T_e\psi)) \rightarrow T_d\varphi))$ (assumption).

3. $\forall e(Ube \rightarrow (Uea \rightarrow T_e\psi))$ (assumption).

4. $Uba \rightarrow (\forall e(Ube \rightarrow (Ued \rightarrow T_e\psi)) \rightarrow T_a\varphi$ (UI, 2).

5. $T_a\varphi$ (MP, 1,3,4).

□

13.9 Pure temporal logic

Pure temporal logic is the logic of earlier and later, the logic of specific time points, and the logic of what are called "temporally definite" propositions.

We have two operators: $T_a\varphi$ 'It is the case at instant a that φ" and Uab, "Instant a is earlier than instant b". Notice that with these operators we are able to talk about two levels at the same time: the "logical" level of truths of formulas, and the "structural" level of properties of instants. Uab is essentially a predicate logic statement about the accessibility relation R on a frame. In the minimal earlier-later calculus, which we call U following Prior, we have the following postulates [Prior, 1968, p. 117]:

T1 $T_a(\varphi \rightarrow \psi) \rightarrow (T_a\varphi \rightarrow T_a\psi)$ (i.e., *modus ponens* holds).

T2 $T_a\neg\varphi \leftrightarrow \neg T_a\varphi$ (i.e., the logic is two-valued).

UT1 $T_aG\varphi \leftrightarrow \forall b(Uab \rightarrow T_b\varphi)$.

UT2 $T_aH\varphi \leftrightarrow \forall b(Uba \rightarrow T_b\varphi)$.

Some theorems:

- $(T_a\varphi \rightarrow T_a\psi) \rightarrow T_a(\varphi \rightarrow \psi)$. (The converse of (T1) above).
- $T_aF\varphi \leftrightarrow \exists b(Uab \wedge T_b\varphi)$.
- $T_aP\varphi \leftrightarrow \exists b(Uba \wedge T_b\varphi)$.

Formulas preceded by T_a or U are temporally definite; they will never change their truth value no matter where in the model they're evaluated. It was partly out of the development of this calculus that Prior was led towards developing hybrid logic (cf. Chapter 15); it can often be convenient to view the earlier-later calculi as hybrid logics (this is essentially the fourth grade of tense-logical involvement, discussed below): For a given frame $\mathcal{F}$, augment your set of proposition letters with another set of proposition letters the same size as W; extend V so that each of the new proposition letters is true in exactly one world in W. We can use the propositions, called nominals, to label or "name" the worlds, or instants, and then we can add the hybrid operator $@_i$: $\mathcal{M} \vDash @_i\varphi$ iff $\mathcal{M}, i \vDash \varphi$. We can also use the binder operator $\downarrow$ which is used to "bind" nominals to worlds.

Note: If you distinguish between propositions and what they express, then you can use hybrid logic to express the distinction between "true of" a world or a time and "true at" a world or a time.

13.10 Four grades of tense logical involvement

Above we noted that McTaggart took the A-series as basic and the B-series as derivative. Prior actually argues the opposite, that we can take the B-series and derive the A-series from it. The different ways that the A- and B-series concepts can be related form Prior "four grades of tense logical involvement" (like Quine's three grades of modal involvement). These grades are discussed in [Prior, 1968, ch. 11]. There are, quite naturally, four different approaches [Øhrstrøm and Hasle, 2006b, p. 456]:

1. The B-series is more fundamental than the A-series; the A-series needs to be defined in terms of the B-series.

2. The two series are equally fundamental; one cannot be derived from the other.

3. The A-series is more fundamental than the B-series; the B-series needs to be defined in terms of the A-series, but this can only be done with the help of another fundamental notion, of temporal possibility.

4. The A-series is more fundamental than the B-series; the B-series needs to be defined in terms of the A-series. The notion of temporal possibility can be defined in the A-series. (McTaggart's view).

Each of these ways gives rise to a different foundation to temporal logic.

On the first grade, we view $\mathrm{K_t}$ (that is, "tensed") formulas as 'predicates' which can be true or false of instants. Then,

> Tense logic, we might say, is a logic of pure predicates which are artificially torn away from their subjects and given a spurious independence. Its theorems only make sense if we understand them to be implicitly preceded by a T_a, as they are explicitly in the U-calculus [Prior, 1968, p. 117].

On this grade, the T_a operator is not embeddable, and the U-calculus is conceptually prior to the $\mathrm{K_t}$ temporal logic. This approach is problematic for those who dislike the two-sorted logic; there are ontological questions about what exactly these a's and b's are. Perhaps we don't want them to be primitive, but to be *defined* from tensed statements. This leads us to the second grade.

In the second grade, propositions of the form $T_a\varphi$ are considered to be well-formed (Prior: "genuinely propositional"), meaning that we can embed the T_a operator, e.g., $T_bT_a\varphi$ is a wff. We ignore the Uab formulas (except in so far as we retain them as used in the previous grade). Then we must add the following axioms [Prior, 1968, p. 119]:

RT From $\vdash \varphi$ infer $\vdash T_a\varphi$.

T3 $\forall aT_a\varphi \rightarrow \varphi$.

T4 $\forall aT_a\varphi \rightarrow T_b\forall aT_a\varphi$.

T5 $T_a\varphi \rightarrow T_bT_a\varphi$.

(This is equivalent to Rescher's system SI [Rescher, 1966]). For a not free in φ, let $\Box\varphi :=$ $\forall aT_a\varphi$. This system has the following theorems:

RN From $\vdash \varphi$ infer $\vdash \Box\varphi$.

$\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$.

$\Box\varphi \rightarrow \varphi$.

$\neg\,\Box\,\varphi \rightarrow \Box\neg\,\Box\,\varphi$.

These are Gödel's postulates for S5. Thus, we've now omitted the operator T_a from the earlier-later calculus. Instead, if we reintroduce the Uab's, we're back to tense logic, since we can prove such theorems as $\Box\varphi \rightarrow G\varphi$ and $\Box\varphi \rightarrow H\varphi$. Note that while $FT_a\varphi \rightarrow T_a\varphi$ and $T_a\varphi \rightarrow GT_a\varphi$ hold, their converses don't unless time has no beginning or ending.

The third grade allows for the formation of instant predicates not just from other predicates but from propositions as well. Let α be any chronologically definite proposition (whereas in the first grade we only had the indefinite, i.e., tense-logical propositions). We add two axiom schemata:

TS1 $\vdash T_a(\varphi \to \alpha) \leftrightarrow (T_a\varphi \to \alpha)$.

TS2 $\vdash T_a(\alpha \to \varphi) \leftrightarrow (\alpha \to T_a\varphi)$.

The fourth grade turns instants into propositions, and is essentially hybrid logic.

13.11 The Event Calculus

Even when we consider continuous temporal structures (which we haven't, really), we are still evaluating propositions at static points—atomic evaluation. We tried to fix this when we moved to interval-based structures, but we struggled with both downard and upward closure of truth.

we are still in a position where we can't take into account *aspect*: "John wins" vs. "John is winning".

Fundamentally requires a shift from logic of time to logic of action—connection because actions happen during time. Also relevant to obligations/deontic logic: Are we interested in obligation of action or of outcome? (STIT: outcome focused; dynamic logic: action focused).

Accounting for actions via events—[Davidson, 1967]:

$$\exists e(\text{stab}(e, \mathsf{Brutus}, \mathsf{Caesar}) \wedge \text{violent}(e))$$

$\Rightarrow$ common in linguistics where all verbs denote events.

$\Rightarrow$ we can add tenses directly to this, since as we will see in Chapter14 that this isn't so scary (might need to get into metaphysical questions of what we're quantifying over).

Event calculus: An approach to formal representation of actiosn via events that avoids the *frame problem*: how to completely specify not only the effects of actions but also non-effects—cf. sneezing doesn't unload a gun.

Three flavors: simple, full, and extended.

A logical mechanism for inferring what's true when, given what happens when and what actions do.

1. *narrative of events*: what happens when.

2. *effects of actions*: what actions do.

3. what is true when.

Three types of tasks: (1) + (2) and get (3) = deductive; (2) + (3) and get (1) = abductive; (3) + (1) and get (2) = inductive.

deductive: temporal projection or prediction. abductive: explanation/postdiction; diagnosis; planning inductive: learning; scientific discovery; theory formation

(Q) What goes into the black box? Modal logic (dynamic logic) or first-order logic? Event calculus = FOL. What would happen if you did modal logic instead?

EC: many-sorted: quantification over events/actions; fluents (things that change truth value over time); time points.

13.11.1 The Simple Event Calculus

- what happens when
- describing the initial situation
- effects of actions
- what fluents hold when

give axioms The frame problem—uniqueness and predicate completion work; but can we do it automatically?

13.11.2 The Full Event Calculus

- new axioms saying when a fluent *doesn't* hold
- Three arg version of "happens", actions with duration; actions that are compound.
- new predicates

Full EC cannot handle well: non-determinism (cf. Russian turkey shoot) (also want to handle indirect effects).

13.11.3 The Extended Event Calculus

- concurrent action and continuous change
- new predicates
- new axioms

compounding actions: &

13.12 Further reading

Because temporal logic intersects with so many different disciplines, there is a wide range of relevant secondary literature, which we list (non-exhaustively) here. Some of the books are more geared towards logic, some towards philosophy, and some more towards computer science. (Some of these are also included in the bibliography, q.v., which contains other references not mentioned here.) [van Benthem, 1983, Freeman and Sellars, 1971, Gabbay, 1976, Galton, 1987, Hintikka, 1973, Kilwardby, 1993, Kröger, 1987, Kröger and Merz, 2008, Manna and Pnueli, 1 Reichenbach, 1971, Rescher and Urquhart, 1971, Øhrstrøm and Hasle, 1995].

Chapter 14

Quantified modal and temporal logic [last modified 01 Mar 21]

14.1 Introduction

Everything we've done so far has been propositional. What about adding modalities or tenses to predicate logic?

Quine, per Fitting and Mendelsohn, says quantified modal logic cannot be done coherently. Why? This is because quantifying into modal or temporal contexts is incoherent (Quine says).

de re vs. *de dicto* modality:

$$\exists x \Diamond \varphi x \text{ vs. } \Diamond \exists x \varphi x$$

Similar problems arise — you might think — with quantified temporal logic:

$$\exists x F \varphi x \text{ vs. } F \exists x \varphi x$$

Are we talking of propositions or of people?

Yet, in natural language we make statements that mix quantification and modality or tense all the time!

- Something white could be black.
- What is white will be black.
- What is alive will be dead.

- Sometime I won't exist.
- No man will be a boy.
- Before my daughter was born, was it true to say that I will have a child?

In this chapter, we combine the languages of quantified logic (Chapter 8) and modal logic (Chapter 11) or temporal logic (Chapter 13) in order to express these types of sentences, and philosophical problems arising from them. A quantified modal language is built by taking a quantified language (cf. Definition 8.2.1) and extending it with the operators $\Box$ and φ in the way that we defined propositional modal languages by extending propositional languages; and a quantified temporal language is built in the same way by adding the operators F, P, G, and H.

14.2 Philosophical issues

In [Fitting and Mendelsohn, 1998, §4.4], Fitting and Mendelsohn address the question "is quantified modal logic possible?" They note that

> for much of the latter half of the twentieth century, there has been considerable antipathy toward the development of modal logic in certain quarters. Many of the philosophical objectors find their inspiration in the work of W.V.O. Quine, who as early as (Quine, 1943), expressed doubts about the coherence of the project... Quine does not believe that quantified modal logic can be done coherently... [Fitting and Mendelsohn, 1998, p. 89]

And Garson in his introduction says

> The problem is that quantified modal logic is not as well developed... Philosophical worries about whether quantification is coherent or advisable in certain modal settings partly explains this lack of attention [Garson, 2006, p. xiii]

Problems in quantified modal logic will often transpose into problems in quantified tense or temporal logic. In this chapter, we look at quantified modal and temporal logic together.

de re vs. *de dicto*

Many natural language sentences involving bothe modality and quantification are ambiguous:

All men are necessarily running

has two readings:

- "It is necessary that all men are running", i.e., "The proposition that all men are running expresses a necessary truth". ⇒ *de dicto* 'of the *dictum* (statement)' reading.
- "Every man is such that he is necessarily running", i.e., "Each man exemplifies the property "necessarily running". ⇒ *de re* 'of the *res* (thing)' reading.

For some statements, the *de dicto* reading is simply contradictory:

Some white thing is possibly black.
⇒ "It is possible that some white thing is black" (*de dicto*).

But, on the *de re* reading, what do we mean by "necessarily running"? or "possibly black"?

Fine defines '*de re* skepticism' as "the doctrine that quantification into modal contexts does not, as it stands, make sense" [Fine, 1978, p. 125]. Quine's complaint against quantified modal or temporal logic is a form of *de re* skepticism.

The same distinction arises in tense logic. We distinguish two ways of reading

Some white thing will be black.

- It will be the case that some white thing is black.
- Some white thing is such that it will be the case that it is black.

If we read

- Some white thing will be black.

as

- Some white thing is such that it will be the case that it is black.

then we need to be able to talk about objects which persist through time. In modal logic, where elements of W are interpreted as possible worlds, we are faced with a similar problem, the problem of trans-world identity—when are two objects existing in different worlds the 'same' object?

One solution (due to Lewis 1968):

- Deny the existence of the 'same' object in more than one possible world.
- Introduce *counterparts* for objects.

- The counterpart in w' of object x in w is that object y which is most similar to x.

There are some problems with this approach:

- How do we cash out "similar"?
- There may be no uniquely "most similar" object.
- The "(most) similar" relation(s) is neither transitive nor symmetric (i.e., if $y \in w'$ is the counterpart of $x \in w$, it may be that $z \in w$ is the counterpart of $y \in w'$, with $x \neq z$).

People have attempted to address these problems, and to develop a robust theory of counterparts. But while counterpart theory may work for the modal approach, it is philosophically undesirable for tense logic if anyone thinks objects persist through time.

Desideratum # 1 We want to be able to take seriously the idea of the *same* object(s) existing at different times.

What doesn't exist is nothing at all...

Ryle writes:

> What does not exist... cannot be named, individually indicated or put on a list, and cannot therefore be characterised as having been prevented from existing [Prior, 1967, pp. 142–143].

If this is true, then when something does not exist, you cannot say anything about it, even that it did exist in the past.

Note 14.2.1. Speaking about non-existents is a problem generally, not just in modal or temporal logic. Logics which are able to account for non-existents are called "free logics".

While many of the issues with mixing modality and quantification are shared by both quantified modal logic and quantified temporal logic, quantified temporal logic has some of its own, unique problems, that arise from mixing the past and future operators. For instance, we also have problems with formulas of the following sort: $\exists x \varphi x \rightarrow P \exists x F \varphi x$, despite the fact that we make statements about past existents all the time:

- Caesar crossed the Rubicon.
- Caesar is dead.

And we also make statements about future existents:

- There is a future president of the U.S. who has not yet been born.
- Some man will go to Mars.
- Antichrist will exist.

Note 14.2.2. But these latter statements combine two issues: The truth value of future contingent statements, and the truth value of statements about future existents.

Desideratum # 2 We want to be able to talk about non-existent objects.

Desideratum # 3 We want to be able to allow objects to come in and out of existence.

In the next sections, we consider two approaches to these problems: Fitting & Mendelsohn's quantified modal models, where some objects do not exist (§14.3) and Prior's many-valued matrix models: some propositions do not exist (§14.4).

14.3 Kripke-style quantified modal logic

In this subsection, we extend the Kripke models of Chapter 11 with the quantificational models of Chapter 8, that is, we take Kripke frames and associate with each world a domain, and add to the model an interpretation function that interprets the predicates, functions, and constants.

Quantified Modal Logic comes in two variants: Varying Domain (where each world possibly has a distinct domain) and Constant Domain (where each world has the same domain). As it turns out, it is possible to simulate varying domain models within constant domain ones, so that ultimately it doesn't matter which you pick. Thus, the models we define will be a hybrid between the two.

Definition 14.3.1 (Quantified Kripke Models)**.** A *quantified Kripke model* is a tuple $\mathfrak{M} = \langle W, R, D_w : w \in W, D, I \rangle$ where:

- W is a non-empty set of worlds.
- $R \subseteq W \times W$ is a binary relation between possible worlds (that is, it is a set of ordered pairs $\langle w, w' \rangle$, where w and w' are both elements of W).
- D_w for each $w \in W$ is a set of objects, called the 'domain' of w.
- $D = \bigcup_{w \in W} D_w$ is the union of all the D_w.

- I is an interpretation function, such that
  - For every constant c, $I(\mathsf{c}, w) \in D_w$.
  - For every n-place function f^n, $I(f^n, w)$ is an n-ary function on D_w.

Note that the interpretation function is relativized to each individual world w; when worlds do not have the same domain, we cannot give a uniform interpretation across all the worlds. Furthermore, even if we did have a constant domain, with the same objects "available" at each world, we might still want to allow objects to have different properties at different worlds.

There are two type of quantified Kripke models:

Definition 14.3.2 (Constant Domain)**.** A quantified Kripke model is *constant-domain* if every $D_w = D$.

Definition 14.3.3 (Varying Domain)**.** A quantified Kripke model is *varying-domain* if it is not constant domain, that is, at least one $D_w \subset D$.

Formulas are evaluated on a model with respect to a world and a variable assignment (cf. Definition 8.3.2). For non-modal or -temporal wffs, the conditions of truth are exactly the same as they are in a single-world scenario (cf. Definition 8.3.4).

Just as the R relation can be constrained in many ways (see §11.3.1), so too can the ways in which domains can vary, if you have a varying domain model. We consider here just two of the most important ways:

Definition 14.3.4 (Monotonicity)**.** A varying domain model is *monotonic* iff if wRw', then $D_w \subseteq D_{w'}$.

Definition 14.3.5 (Anti-Monotonicity)**.** A varying domain model is *anti-monotonic* iff if $w'Rw$, then $D_w \subseteq D_{w'}$.

A monotonic model is one in which, as you move from world to world down the R relation, you only ever gain new objects, you never lose them. (If you think of R as a temporal relation, then this is equivalent to saying that nothing ever goes out of existence, but new things could perhaps come into existence.) An anti-monotonic model is one in which, as you move from world to world down the R relation, you only ever lose objects, you never gain then. (On a temporal interpretation, this would be the same as saying that things only go out of existence, they never come into eistence.) A model which is both monotonic and anti-monotonic will be constaint domain: Nothing is ever lost or gained.

Just as we can introduce modal formulas that correspond to constraints on R, so too it is possible to introduce quantified modal formulas that correspond to constraints on how the domain can vary. The formulas that correspond to monotonicity and anti-monotonicity are called Barcan and Converse Barcan formulas, in honor of the first person to develop modal systems which validate them, Ruth Barcan Marcus.[1]

$$\Diamond \exists x \varphi \rightarrow \exists x \Diamond \varphi \qquad \text{(Barcan Formula)}$$

$$\exists x \Diamond \varphi \rightarrow \Diamond \exists x \varphi \qquad \text{(Converse Barcan Formula)}$$

Lemma 14.3.6. *A model validates (all instances of) the Barcan Formula iff it is anti-monotonic.*

Proof.

$\Rightarrow$: Let $\mathfrak{M}$ be a quantified Kripke model which is anti-mononic, but that there is some world w and variable assignment v such that

$$w, v \vDash \neg(\Diamond \exists x Bx \rightarrow \exists x \Diamond Bx) \tag{14.1}$$

That is:

$$w, v \vDash \Diamond \exists x Bx \tag{14.2}$$

and

$$w, v \vDash \neg \exists x \Diamond Bx \tag{14.3}$$

which is the same as saying:

$$w, v \vDash \forall x \Box \neg Bx \tag{14.4}$$

By (14.2), there is a w' such that wRw' and

$$w', v \vDash \exists x Bx \tag{14.5}$$

This means there is some $a \in D_{w'}$ such that $v(x, w') = a$ and

$$w', v \vDash Ba \tag{14.6}$$

Now, because the model is anti-monotonic, we know that since wRw', if $a \in D_{w'}$, also $a \in D_w$. This means we can instantiate $\forall x \Box \neg Bx$ to

$$w, v \vDash \Box \neg Ba \tag{14.7}$$

[1] Strictly speaking, these are not formulas but formula schemes: the result of replacing φ with another formula will still be a Barcan formula.

Since wRw', it follows that

$$w', v \vDash \neg Ba \tag{14.8}$$

which is a contradiction with (14.6).

$\Leftarrow$: Exercise for the reader. □

Lemma 14.3.7. *A model validates (all instances of) the Converse Barcan Formula iff it is monotonic.*

Proof. Exercise for the reader. □

14.3.1 Applying this approach to temporal logic

Take ordinary quantification theory, and add your choice of tense operators (metric or non). We can express the *de re*/*de dicto* distinction by means of scope of the modal operators and quantifiers:

- *de dicto*: $F\forall x\varphi x$.
- *de re*: $\forall xF\varphi x$.

This approach only works for quantified sentences; we'll discuss how to deal with open formulas and sentences containing constants at the end.

Definition 14.3.8. A *quantified temporal frame* is tuple $\mathfrak{F}^Q = \langle W, R, D, V\rangle$ where W, R are in ordinary temporal frames, D is a set of objects (the domain), and $V : W \to \mathcal{P}(D)$ is a function specifying for each world the set of objects which 'exist' at that world.

Definition 14.3.9. A *quantified temporal model* is a tuple $\mathfrak{M}^Q = \langle \mathfrak{F}^Q, I\rangle$, where $\mathfrak{F}^Q$ is a quantified temporal frame and I is an interpretation assigning to each n-place relation symbol R and each possible world $w \in W$ a set of n-tuples of D.

Definition 14.3.10. A *valuation* on a model is a mapping $v : x \to D$, where x is a free variable.

Definition 14.3.11. v' is an *x-variant* of a valuation function v if v and v' agree on every variable except possibly x.

Definition 14.3.12. v' is an *x-variant at w* of a valuation function v if v and v' agree on every variable except possibly x and $v'(x) \in V(w)$.

Definition 14.3.13. A model is *locally constant domain* iff wRv implies $V(w) = V(v)$.

Note 14.3.14. Constant domains correspond to possibilist quantification, and varying domains to actualist quantification.

For every varying domain model, there is a constant domain model that mimics it, in the presence of an extended language.

Proposition 14.3.15 (Fitting & Mendelsohn Prop. 4.9.10)**.** *A sentence is valid on all locally constant domain models iff it is valid on all constant domain models.*

Barcan formula	$F\exists x\varphi x \to \exists x F\varphi x$	$\forall x G\varphi x \to G\forall x\varphi x$
Converse Barcan formula	$\exists x F\varphi x \to F\exists x\varphi x$	$G\forall x\varphi x \to \forall x G\varphi x$

(Technically, these are formula *schemes*. But we'll call them formulas because it's easier.)

Definition 14.3.16. We call any logic containing the Barcan formula a *Barcan system*.

Barcan systems validate:

1. $G\forall x\varphi x \leftrightarrow \forall x G\varphi x$ (collapse of *de re*/*de dicto* distinction).
2. $G\forall x\varphi x \to F\forall x\varphi x$.
3. $\forall x G\varphi x \to \exists x G\varphi x$.
4. $F\forall x\varphi x \to \forall x F\varphi x$.
5. $\exists x G\varphi x \to G\exists x\varphi x$.
6. $\forall x F\varphi x \to \exists x F\varphi x$.
7. $G\exists x\varphi x \to F\exists x\varphi x$.

Proposition 14.3.17 (Fitting & Mendelsohn Prop. 4.9.6)**.** *A varying-domain model is monotonic iff it validates the Converse Barcan formula.*

Proposition 14.3.18 (Fitting & Mendelsohn Prop. 4.9.8)**.** *A varying-domain model is anti-monotonic iff it validates the Barcan formula.*

Note 14.3.19. In a monotonic model, nothing ever goes out of existence. In an anti-monotonic model, nothing ever comes into existence. Thus, generally we will want to ensure our models are neither monotonic nor anti-monotonic.

If our models are constant domain, then

$$\begin{array}{lll} \mathfrak{M}^Q, w \vDash_v \forall x\varphi & \text{iff} & \text{for every } x\text{-variant } v' \text{ of } v, \mathfrak{M}^Q, w \vDash_{v'} \varphi \\ \mathfrak{M}^Q, w \vDash_v \exists x\varphi & \text{iff} & \text{for some } x\text{-variant } v' \text{ of } v, \mathfrak{M}^Q, w \vDash_{v'} \varphi \end{array}$$

If the model is varying-domain, then

$$\begin{array}{lll} \mathfrak{M}^Q, w \vDash_v \forall x\varphi & \text{iff} & \text{for every } x\text{-variant-at-}w\ v' \text{ of } v, \mathfrak{M}^Q, w \vDash_{v'} \varphi \\ \mathfrak{M}^Q, w \vDash_v \exists x\varphi & \text{iff} & \text{for some } x\text{-variant-at-}w\ v' \text{ of } v, \mathfrak{M}^Q, w \vDash_{v'} \varphi \end{array}$$

That is, we only quantify over actually existing objects.

We can simulate varying domain models from within constant domain models by the addition of an existence predicate E with this semantics:

$$\mathfrak{M}^Q, w \vDash_v E(x) \quad \text{iff} \quad v(x) \in V(w)$$

Definition 14.3.20. The *existence relativization* of a formula φ, denoted φ^E is defined as follows:

- $p^E = p$ for atomic p.
- $(\neg\varphi)^E = \neg(\varphi^E)$.
- For binary connection $\circ$, $(\varphi \circ \psi)^E = (\varphi^E \circ \psi^E)$.
- $(L\varphi)^E = L\varphi^E$.
- $(M\varphi)^E = M\varphi^E$.
- $(\forall x\varphi)^E = \forall x(E(x) \to \varphi^E)$.
- $(\exists x\varphi)^E = \exists x(E(x) \wedge \varphi^E)$.

Proposition 14.3.21. *Let φ being a formula not containing E. Then φ is valid on every varying domain model iff φ^E is valid on every constant domain model.*

Proof. By induction. □

The existence-relativization is very similar to the proposal by Kripke that Prior discusses [Prior, 1967, pp. 161–162]: Identify a set of predicates whose predication entails the existence of the object predicated of (e.g., 'red', 'mortal', 'hard', as opposed to, e.g., 'is thought of', 'is thought to be red'.)

1. $fy \to (\forall x\varphi x \to \varphi y)$.

2. $fy \to (\forall x \neg \varphi x \to \neg \varphi y)$.

3. $fy \to (\varphi y \to \neg \forall x \neg \varphi x)$.

4. $fy \to (\varphi y \to \exists x \varphi x)$.

5. $fy \to (fy \to \exists x fx)$.

6. $fy \to \exists x fx$.

7. $\forall y fy \to (\forall x fx \to fy)$.

8. $\forall x fx \to (\forall x fx \to fy)$.

9. $\forall x fx \to fy$.

We now return to the problem of sentences with constants and open formulas. We can add constants to our language, augmenting our interpretation function in the standard way.

Definition 14.3.22. We say that a constant c *designates* at w if $I(c, w) \in V(w)$. It does not designate otherwise.

If our domains are constant, every constant c will designate. On varying domain models, constants may not always designate at every world.

For constants that do designate, we specify two ways they can do so:

Definition 14.3.23.

- a constant c is a *rigid designator* if $I(c, w) = I(c, w')$ for all w, w'.
- it is a *non-rigid designator* if $I(c, w) \neq I(c, w')$ for some w, w'.

Formulas such as $F\varphi c$ (and $F\varphi x$) are essentially ambiguous between the *de re* or *de dicto* readings:

de re c has the "will-be-φ' property.

de dicto It will be the case that c is φ.

Formally,

de re $\mathfrak{M}^Q, w \vDash F\varphi c$ iff $\exists w'$ such that wRw' and $\mathfrak{M}^Q, w' \vDash_v \varphi x$ where $v(x) = I(c, w)$ and $v(x) \in I(\varphi, w')$.

de dicto $\mathfrak{M}^Q, w \vDash F\varphi c$ iff $\exists w''$ such that wRw'' and $\mathfrak{M}^Q, w'' \vDash_v \varphi x$ where $v(x) = I(c, w'')$ and $v(x) \in I(\varphi, w'')$.

If c is rigid, then the *de re* and *de dicto* readings collapse: $I(c, w) = I(c, w') = I(c, w'')$.

But if c is non-rigid, then even if $w' = w''$, the two readings do not collapse:

- "That very object which c designates here (w) has the property φ there (w')."
- "That very object which c designates there (w') has the property of φ there (w')."

If we want to resolve this ambiguity at the level of syntax, then we need to add new notation, such as $\varphi^F c$ for the *de re* reading, and letting $F\varphi c$ only indicate the *de dicto* reading.

14.4 Prior's system Q

In *Time and Modality* (and again in *Past, Present, and Future* [Prior, 1967, ch. VIII]), Prior introduced a modal "logic of contingent beings". It is given by matrices, rather than by axioms or models.

- Each proposition is associated with an infinite sequence of 0, 1, and 2, where 2 cannot be the first digit.
- Any compound will have 2 where any component has 2.
- Otherwise,
  - $\neg\varphi$ interchanges 0 and 1;
  - $\varphi \wedge \psi$ is 1 where both φ and ψ are 1, and 0 otherwise.
  - $M\varphi$ is 1 everywhere (apart from where there are 2's) if φ is 1 somewhere.
  - $L\varphi$ is 0 everywhere unless φ is 1 everywhere.

A sentence is a theorem iff its sequence has only 1s.

Q can be interpreted as a temporal logic by letting the places in the sequence stand for linearly ordered time points, and reading L as temporal 'always', M as 'sometimes', $\neg M\neg$ for 'never not', and $\neg L\neg$ as 'not always'.

The value 2 can be taken to stand for an 'undesignated' truth value. But this leads to some conceptual oddities, in that if ψ has value 2, then $(\varphi \vee \neg\varphi) \vee \psi$ also has value 2. More interestingly, it can be taken to mean 'this proposition does not exist at this time', or 'this proposition is not assertible'. Then, we do not have the previous conceptual problem. As

Prior notes: "When there is no such proposition as p, there are no functions of p either" [Prior, 1967, p. 154]. This interpretation gives rise to a distinction between "true *at* a time" and "true *of* a time". If there is some time point where every proposition receives the value 2, then it is true *of* that time that no proposition is true, though it is not true *at* that time that no proposition is true, since the proposition "no proposition is true" receives the value 2 as well. ⇒ Connections to Buridan's *sophismata*.

Q can also be 'temporalized' by appending it to the $\mathcal{U}$-calculus. Prior offers the following proposals for axiomatizing the $Q\mathcal{U}$-calculus:

1 $T_a\neg\varphi \to \neg T_a\varphi$ (but NOT the converse).

2 $T_a(\varphi \to \psi) \leftrightarrow (T_a\varphi \to T_a\psi)$.

3 $T_a(\varphi \wedge \psi) \leftrightarrow (T_a\varphi \wedge T_a\psi)$.

TP $T_aP\varphi \leftrightarrow \exists b(Uba \wedge T_b\varphi)$.

TH $T_aH\varphi \leftrightarrow \forall b(Uba \to T_b\varphi)$.

TF $T_aF\varphi \leftrightarrow \exists b(Uab \wedge T_b\varphi)$.

TG $T_aG\varphi \leftrightarrow \forall b(Uab \to T_b\varphi)$.

QUR1 from $\vdash \alpha$ infer $\vdash \neg T_a\neg\alpha$.

QUR2 from $\vdash \alpha \to \beta$ infer $T_a\alpha \to T_b\beta$ if β has no free variable not in α.

This logic was axiomatized in [Bull, 1964].

14.5 Medieval quantified temporal logic

Medieval logicians: "Always some man has existed" does not imply "some man has always existed."

$$G\exists xMx \not\to \exists xGMx$$

The problem of coming to be and ceasing to be.

non-being | being

how does something get from the left to the right?

Varying vs. constant domains.

Barcan is true when nothing comes into existence. Converse Barcan is true when nothing goes out of existence.

If both are valid, constant domain.

Let's consider a restricted language:

1. Predicates, S, P, Q, R
2. Copulae, a, e, i, o

Why? Lambert of Lagny/Auxerre, 13th C, on supposition, ampliation, and restriction

Define signification and supposition.

Define appellation, ampliation, restriction

Connection between appellata and non-existents.

Models: $\langle T, <, O, E, V\rangle$: instants, ordering, objects, existence, valuation. If $a \in V(Q,t)$ then "a is Q at t". If $t \in E(a)$, then a exists at t and $t \vDash \hat{\mathsf{a}}$. $\hat{\mathsf{a}}$ is an existence proposition.

$V'(Q) := \bigcup_{t\in T} V(Q,t)$

Existential import: $V(Q,t) \neq \varnothing$ for every Q,t pair.

$$\begin{aligned} t \vdash S\,a\,Q \quad &\text{iff} \quad V(S,t) \subseteq V(Q,t) \\ t \vdash S\,e\,Q \quad &\text{iff} \quad V(S,t) \cap V(Q,t) = \varnothing \\ t \vdash S\,i\,Q \quad &\text{iff} \quad V(S,t) \cap V(Q,t) \neq \varnothing \\ t \vdash S\,o\,Q \quad &\text{iff} \quad V(S,t) \nsubseteq V(Q,t) \end{aligned}$$

$Ap(Q,t) = V(Q,t) \cap \{a : t \vDash \hat{\mathsf{a}}\}$

Simple predications:

$$\begin{aligned} t \vDash \exists x Q x \quad &\text{iff} \quad Ap(Q,t) \neq \varnothing \\ t \vDash \forall x Q x \quad &\text{iff} \quad Ap(Q,t) = \{a : t \vDash \hat{\mathsf{a}}\} \end{aligned}$$

Possible for $V(Q,t) \neq \varnothing$ but $t \nvDash \exists x Q x$ if none of the things that are Q at t exist. $\forall x/\exists x \Rightarrow$ varying domain. $a, e, i, o \Rightarrow$ constant domain.

Natural restriction

$$Res(Q,S,t) = Ap(Q,t) \cap Ap(S,t)$$

Tense ampliation:

$$Amp_t^F(Q,w) = \bigcup_{t>w} Ap(Q,t)$$

$$Amp_t^G(Q,w) = \bigcap_{t>w} Ap(Q,t)$$

Exercise 12. Define medieval quantified temporal models that make the following sentences true:

1. All roses are flowers (even when there are no roses).
2. A white door will be black.
3. No man will be a boy.

Chapter 15

Hybrid logic [last modified 26 Feb 19]

Note: This section was originally written to be self-contained, so there may be some repeats of material earlier in the notes until the content here gets better integrated.

A hybrid logic is a logic which extends the expressive power of modal logic through the addition of a special class of propositional symbols, called nominals, and operators which act on those nominals.

15.0.1 Basic hybrid tense logic

In this section, we take our base modal logic to be a tense logic with forward- and backward-looking operators, thus the basic propositional hybrid tense language $\mathcal{L}^{@}_{Tp}$ (cf. Definition 11.2.1) is defined as follows:

Definition 15.0.1. The *propositional hybrid tense language* $\mathcal{L}^{@}_{Tp}$ consists in:

- An infinite set $\mathfrak{P}$ of atomic propositional letters p, q, r, ...
- An infinite set $\mathfrak{N}$ of nominals i, j, k, ...
- Four propositional operators: $\neg$ (unary), $\vee, \wedge, \rightarrow$ (binary).
- Two unary modal operators: F, P.
- A family of hybrid operators: $@_i$, one for each $i \in \mathfrak{N}$.
- Punctuation: (,).

The set of well-formed formulas is defined recursively:

Definition 15.0.2 (Propositional wffs)**.**

- Every atomic proposition is an $\mathcal{L}_{Tp}$-wff.
- Every nominal is an $\mathcal{L}_{Tp}$-wff.
- If φ is a $\mathcal{L}_{Tp}$-wff, then so are $\neg\varphi$, $F\varphi$, and $P\varphi$.
- If φ and ψ are $\mathcal{L}_{Tp}$-wffs, then so are $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, and $(\varphi \rightarrow \psi)$.
- If i is a nominal and φ is an $\mathcal{L}_{Tp}$-wff, then so is $@_i\varphi$.

We define $G\varphi$ as $\neg F\neg\varphi$ and $H\varphi$ as $\neg P\neg\varphi$.

Models for hybrid logics are extensions of the standard possible worlds models for modal logic.

Definition 15.0.3 (Hybrid Model)**.** A *hybrid model* is a triple $\langle W, R, n\rangle$ where W is a set of possible worlds, R is a binary relation on $W \times W$, called an accessibility relation, and n is a naming function from $\mathfrak{N}$ to W, such that $n(i) = w$ for every $i \in \mathfrak{N}$ and some $w \in W$. If $n(i) = w$, then we say that i is a name for w.

A nominal can be thought of as a proposition which is true of exactly one world; in that sense it 'names' the world it is true at, and makes it possible for us to pick out individual worlds in a model uniquely *from within the language itself*. This means that from an internal perspective we can say external things, about the structure of the entire model.

The truth conditions for the propositional operators and the tense operators F and P are as defined earlier (cf. §7.4.1 and §13.2.2). The truth conditions for the hybrid extensions are as follows:

Definition 15.0.4 (Truth conditions for hybrid formulas)**.**

$$\begin{array}{lll} w \vDash i & \text{iff} & n(i) = w \\ w \vDash @_i\varphi & \text{iff} & n(i) \vDash \varphi \end{array}$$

A consequence of this definition is that if $w \vDash @_i\varphi$, then $\vDash @_i\varphi$; that is, statements prefaced with $@_i$ are globally true or globally false; their truth value does not depend on the world of evaluation. It is in this way that hybrid logic allows us to take an internal perspective (via the ordinary tense operators, whose evaluation is dynamic and where the truth value changes with the world of evaluation) and an external perspective (via the

hybrid operators, whose evaluation is static and where the truth value does not change with the world of evaluation).

Here are a few hybrid logic validities:

$$(@_i j \wedge @_i \varphi) \rightarrow @_j \varphi \tag{15.1}$$

$$\text{for every } i \in \mathfrak{N},\ @_i i \tag{15.2}$$

$$\text{for every } i, j \in \mathfrak{N},\ @_i j \rightarrow @_j i \tag{15.3}$$

$$\text{for every } i, j, k \in \mathfrak{N},\ (@_i j \wedge @_j k \rightarrow @_i k) \tag{15.4}$$

15.0.2 Indexical hybrid tense logic

For more information, see [Blackburn and Jørgensen, 2012].

The statement "I am here now" is a curious one: No matter when or where or by whom it is uttered, it can never be false. And yet, one would hesitate to call it a logical validity: For surely it is a contingent matter that I am here now; had things gone differently, I could have been here later, or I could have been elsewhere now (or even perhaps I could have been someone else). So it is not that "I am here now" is true *as a matter of logic*: It must be true as a matter of something else. This "something else" is *context*: No matter what context I am in, this sentence cannot fail to be true, and the reason is that indexicals such as 'I', 'here', 'now' are context-dependent, that is, their referents are determined by the context, and context determines them in such a way that "I am here now" can never be uttered falsely.

In this section we augment the basic hybrid tense logic of the previous section with temporal indexicals (in the next section we sketch how a similar extension can be given for geographical and personal indexicals). We introduce four temporal indexicals: N 'now', Y 'yesterday', D 'today', and T 'tomorrow'. N is a special indexical, in that it is nominal in nature (it names a single, unique point). As noted above, 'now' and other indexicals are context-dependent. Thus, in order to interpret statements involving them, we need to add a notion of context to our models:

Definition 15.0.5 (Hybrid Model with Contexts)**.** A *hybrid model with contexts* is a tuple $\langle W, R, n, C, \eta \rangle$ where $\langle W, R, n \rangle$ is an ordinary hybrid model, C is a set of contexts, and η is a context mapping function from C to W. For a context, $c \in C$, $\eta(c) \in W$ is the world that is considered 'now' in that context.

The evaluation of formulas is now relative to both a world and a context; for the propositional, tense, and non-indexical hybrid cases, the addition of the context does not change

how the formulas are evaluated. For the indexical nominal, we have the following:

$$w, c \vDash @_N \varphi \quad \text{iff} \quad c, \eta(c) \vDash \varphi$$

A consequence of this is the following:

$$c, \eta(c) \vDash N \tag{15.5}$$

That is, no matter the context, at whichever world that context says is now, the statement 'It is now' is true. Thus, 'It is now' is a contextual validity.

Y, D, and T are indexical, but they are not nominals: They pick out a span of time, rather than a single point in time. We need axioms which specify the relationships of these spans of time to each other, and their relationship to now. These are the following:

Location of now:

$$N \to D \text{ (If it is now, it is today)} \tag{15.6}$$

$$Y \to FN \text{ (If it is yesterday, then it will be now)} \tag{15.7}$$

$$T \to PN \text{ (If it is tomorrow, then it was now)} \tag{15.8}$$

Disjointness

$$D \to \neg T \text{ (Today is not tomorrow)} \tag{15.9}$$

$$D \to \neg Y \text{ (Today is not yesterday)} \tag{15.10}$$

$$Y \to \neg T \text{ (Yesterday is not tomorrow)} \tag{15.11}$$

Alignment

$$D \to G\neg Y \text{ (If it is today, it will always be the case that it's not yesterday)} \tag{15.12}$$

$$T \to G\neg D \text{ (If it is tomorrow, it will always be the case that it's not today)} \tag{15.13}$$

$$T \to G\neg Y \text{(If it is tomorrow, it will always be the case that it's not yesterday)} \tag{15.14}$$

No gaps

$$(PY \wedge FD) \to (Y \vee D) \text{ (If it was yesterday and will be today, then it's either yesterday or today)} \tag{15.15}$$

$$(PD \wedge FT) \to (D \vee T) \text{ (If it was today and will be tomorrow, then it's either today or tomorrow)} \tag{15.16}$$

Convexity

$$(PY \wedge FY) \rightarrow Y \text{ (If it was yesterday and will be yesterday, it is yesterday)} \quad (15.17)$$

$$(PD \wedge FD) \rightarrow D \text{ (If it was today and will be today, it is today)} \quad (15.18)$$

$$(PT \wedge FT) \rightarrow T \text{ (If it was tomorrow and will be tomorrow, it is tomorrow)} \quad (15.19)$$

And it is around now that I'm always reminded of Alice in Wonderland, and the folks who got 'jam tomorrow, jam yesterday, and never ever jam today'.

But let us consider something a bit less frivolous. Given an ordinary understanding of how tenses and temporal indexicals work, a sentence like

$$\text{Niels will die yesterday} \quad (15.20)$$

is false, if not nonsensical. Can we use the power provided by indexical hybrid temporal logic to explain why? Yes: We can show that the following sentence is unsatisfiable:

$$F(Y \wedge \text{Niels-dies}) \quad (15.21)$$

First, let us pick an arbitrary time for now, call it k, and set that 'now' as the moment of evaluation for this formula:

1 $@_k N$

2 $@_k F(Y \wedge \text{Niels-dies})$

It follows that there is at least one future point, call it i, and i is the witness for the F claim in (2):

3 $@_k Fi$

4 $@_i(Y \wedge \text{Niels-dies})$

$@_i$ distributes over conjunction, so this gets us:

5 $@_i Y$

6 $@_i$Niels-dies

Invoking one of the axioms connecting 'now' and today', we have:

7 $@_k(N \rightarrow D)$

8 $@_k D$

Invoking three more axioms, we have the following:

9 $@_k(D \to G\neg Y)$

10 $@_k G\neg Y$

11 $@_i \neg Y$

And (5) and (11) are contradictory.

15.0.3 Extending to "here" and "I"

It is straightforward to extend the analysis of "now" into one that works for "here" as well; "I" is a bit different, but can in principle be done in an analogous way.

For "here", think of the element of W not as possible worlds, but rather as (actual) places—for example, a London tube map can be interpreted as a Kripke model: Each of the points on the map is one of the elements of W, and if there is a tube route from point x to point y, then we say that y is accessible from x: xRy. It should be clear that if we think in terms of absolute accessibility, the accessibility of London via the tube is reflexive, transitive, and symmetric: You can always get to any station from itself; if you can get from Paddington to Kings Cross, you can get from Kings Cross to Paddington, and if you can get from Paddington to Euston and from Euston to Kings Cross, then you can get from Paddington to Kings Cross.

But, as anyone who has ever travel led London knows, not every point is "possible" with respect to every other point, especially if you have constraints on your available time. This could be modeled by dropping transitivity, and making each step count explicitly. Then, if we label each arrow between stops with the amount of time it takes to get from one to the other, we can say things like "Euston is n-possible from Kings Cross", to indicate that it takes n or fewer minutes to get from Euston to Kings Cross.

To such models, let us again add a set of contexts and a context function, which work exactly like the models defined in Definition 15.0.5, except that here, the function picks out which point in the model counts as "here". It is then possible to speak of which worlds (which stops on the tube map) are accessible from "here" given some fixed amount of time.

You could also combine "here" and "now" to explore the fact that what is possible from here depends on when it is now—for example, from here and now, it is not possible that we go to Czechoslovakia; for Czechoslovakia no longer exists. The relationship between here and now is symmetric: You could fix "here" and then ask when "now" could be, or you

could fix "now" and ask where "here" could be. In this way (and this is all very sketchy) you could have a two-dimensional modal semantics.

Adapting this sort of analysis to "I" requires that we first add the sort of thing that "I" could refer to to our models. "I" is contextual defined, but it doesn't pick out a possible world, rather, it picks out a possible person or argent.

Definition 15.0.6 (Models with Agents)**.** A *model with agents* is a triple $\langle W, R, A\rangle$ where $\langle W, R\rangle$ is an ordinary Kripke model, and A is a set of agents.

Models with agents can be hybridized in two ways: by adding nominals for worlds and nominals for agents. Hybrid models with agents can be contextualized in exactly the same way, but the context will pick out not only which world is "here" or "now", but also which agent is "I" (and similarly for "you" if we wanted). Then, we can express things like the following:

$$K_I \neg K_Y \varphi \wedge \varphi \tag{15.22}$$

This says "I know that you don't know that φ, and φ is true".

Chapter 16

Dynamic logic [last modified 17 Feb 23]

A number of topics that we've discussed so far have implicitly involved dynamics, including the notion of epistemic uncertainty discussed in §12.1.2 and the dynamic *A*-series notion of time in §13.2. In this section, we make the idea of dynamic logic more precise, and show, among other things, how it can be used to solve certain puzzles of reasoning, including Fitch's paradox.

The basic idea is to take a static logic, which begins with a fixed model and has determinate truth conditions for all the formulas, and then augment this with a set of actions. These actions change the models in certain ways, and we extend the logical language so that we can refer to these actions from within the logic. Thus, while an ordinary static language will consist in a set of atomic propositions and operators that combine these atomic propositions into more complex propositions, a dynamic language will have, in addition to the set of atomic propositions, a set of atomic actions (often called 'programs'), as well as operations that combine these programs into more complex programs. The tricky part is that formulas can occur as *as part of programs*, which means that defining what counts as a well-formed formula or a well-formed program is more complex than in the static case. It is this fact—that programs can contain formulas—that gives us greater expressive power. In fact, the power it gives us is great enough that we don't actually need any atomic programs, we only need a way of constructing complex programs out of formulas. We will introduce one program-creating operator which is sufficient to model a wide range of phenomena; this is, of course, only a very small sliver of dynamic logic. We could easily devote an entire module or more to the topic, but since we cannot, for those who want more information, see [Harel et al., 2002].

16.1 Language

A language for dynamic logic comes in two parts: The propositions and the programs. For each, we begin with a set of atomic propositions, Φ, and a set of atomic programs, Π. If we start with such sets, then the set of well-formed formulas and well-formed programs are defined by induction.

Definition 16.1.1 (Well-formed programs and well-formed formulas)**.** The set of well-formed programs (wfps) and well-formed formulas (wffs) is defined recursively as follows:

- If $\varphi \in \Phi$, then φ is a wff.
- If $\alpha \in \Pi$, then α is a wfp.
- Assume that φ and ψ are wffs and that α and β are wfps, then
  - the following are also wffs:
    - $\varphi \to \psi$ (implication)
    - $\bot$ (falsity)
    - $[\alpha]\varphi$ (program necessity)
  - the following are also wfps:
    - $\alpha; \beta$ (sequential composition)
    - $\alpha \cup \beta$ (nondeterministic choice)
    - $\alpha*$ (iteration)
    - $\varphi?$ (test)
- Nothing else is a wff or wfp.

Once we have introduced models for these operators, we can define their semantics to make these intuitive readings formal. With these basic program operations, it is possible to generate more complex programs, such as programs involving the 'while' operator, which means that these four operators are sufficient to compute all partial recursive functions [Harel et al., 2000, p. xiv].

16.2 Modeling epistemic uncertainty

In epistemic logic, the program $\varphi!$ is interpreted as an announcement: $\varphi!$ is the equivalent of the action of someone coming into the room and saying "It is true that φ!" The modal operator $[\varphi!]$ is then read in a universal temporal fashion: "after every announcement of φ".

As we saw above, if we are uncertain about whether or not it is raining, because we are indoors, this is modeled by two worlds which are epistemically equivalent, one where it is raining and one where it is not (Fig. 16.1).

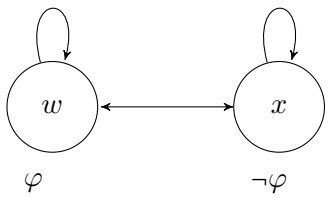


Figure 16.1: Is it raining or not?

The effect of someone coming in and announcing "It is raining" is to remove world x from the model. Formally, this is done by means of *model reduction*.

Definition 16.2.1. The *φ-reduct of* $\mathfrak{M}$ (written $\mathfrak{M} \restriction \varphi$) is:

$$\langle W \restriction \varphi, R \restriction \varphi \rangle$$

where $W \restriction \varphi := \{w \in W : \mathfrak{M}, w \vDash \varphi\}$, that is, it is the set of worlds in the original model where φ is true, and $R \restriction \varphi$ is the restriction of the original to the new set of worlds. (That is, for any arrow in the original model that no longer has an origin or a terminus, we get rid of the arrow in the new model.)

In this setting, we start with a non-empty set of atomic propositions Φ, but our set of atomic programs is empty.

- if φ is a well-formed formula (see Definition 11.2.2), $\varphi!$ is a well-formed program;
- if α is a well-formed program, then $[\alpha]\psi$ is a well-formed program.

The semantics for the modality involving announcements is defined not in terms of the original model, but instead in terms of the relevant reduct:

Definition 16.2.2 (Truth conditions for announcements)**.**

$$\begin{array}{lll} \mathfrak{M}, w \vDash [\varphi!]\psi & \text{iff} & w \vDash \varphi \text{ implies } \mathfrak{M} \upharpoonright \varphi, w \vDash \psi \\ \mathfrak{M}, w \vDash \langle\varphi!\rangle\psi & \text{iff} & w \vDash \varphi \text{ and } \mathfrak{M} \upharpoonright \varphi, w \vDash \psi \end{array}$$

In pictures, announcing that it is raining results in Fig. 16.2, which is the φ-reduct of the model in Fig. 16.1.

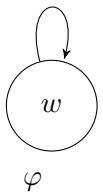


Figure 16.2: After the announcement that it is raining.

Thus, these two models show that $[\varphi!]K\varphi$ is true: After every successful, true announcement of φ, φ is known.

16.2.1 Fitch's paradox

Recall Fitch's paradox, discussed in §12.1.1; one solution to the paradox is to reject the Knowability Thesis, that every truth is in principle knowable. But wholesale rejection of a thesis which has such strong intuitive pull is generally not a desirable move. Instead, it would be useful if we could understand *why* it is that it fails, and thus perhaps identify a non-ad-hoc class of sentences for which the Knowability Thesis still holds true; this is what van Benthem does in [van Benthem, 2004].

One way to reply to Fitch's paradox is to argue that the VT is too strong, and that it is only things which it is consistent to know that are knowable—this is the principle 'CK' introduced by Tennant. This leads to the first restriction of the thesis to: Only what is true and what knowledge of which is consistent can be known. The reason that $\varphi \wedge \neg K\varphi$ violates the original VT is that $K(\varphi \wedge \neg K\varphi)$ is inconsistent. Let us combine this observation with the example of epistemic uncertainty from the previous section. There, we announced φ with the result that afterwards $K\varphi$ was true. Fitch's paradox can be encapsulated by realizing that this doesn't work for $\varphi \wedge \neg K\varphi$. In fact, the announcement of $\varphi \wedge \neg K\varphi$ leads not to knowledge of $\varphi \wedge \neg K\varphi$, but to knowledge of its negation!

This is a rather peculiar phenomenon, and one that we can put a precise label on:

Definition 16.2.3. A sentence is *self-refuting* if its announcement leads to its falsity, that is, if

$$[\psi!]\neg\psi$$

is true.

If we restrict the VT so that it only covers *non-self-refuting* cases, then we can rehabilitate it.

16.2.2 Muddy children/logicians at the bar

> Three logicians walk into a bar. The barman asks, "Do you all want a drink?" The first one says, "I don't know". The second one says, "I don't know". The third says, "Yes!"

How did he know?

Chapter 17

Counterfactuals [last modified 23 Nov 20]

Part IV

Non-classical logics

Chapter 18

Introduction to Part IV [last modified 4 Oct 18]

The logics discussed in this part of the book drop one or both of the two meta-principles underlying classical logic (cf. 5):

LEM The Law of Excluded Middle: Every sentence is, at any given time, either true or false.

PNC The Principle of Non-Contradiction: No sentence is both true and false at the same time.

Chapter 19

Intuitionistic and constructive logics [last modified 01 Mar 21]

19.1 Introduction to constructivity

One of the fundamental axioms of classical propositional logic is that $\varphi \vee \neg\varphi$ is true for every wff φ. It is true *even when we do not know which disjunct is true*, which means that we are able to prove $\varphi \vee \neg\varphi$ even when we cannot prove either disjunct—that is, it is consistent to have $\vdash \varphi \vee \neg\varphi$, $\nvdash \varphi$, $\nvdash \neg\varphi$ in classical propositional logic, and this is because failure to be able to prove a wff is not equivalent to proving its negation.

In certain contexts, this uncertainty about which disjunct is the one that serves as a warrant for the truth of $\varphi \vee \neg\varphi$ is not a problem. But in other contexts, one may want theoremhood in a formal system to correspond to a stronger notion than just "true on every truth-value assignment". For instance, we might wish to say that we can only establish the truth of a mathematical statement by producing a proof of that statement, and not just any old type of proof but a *constructive* one, one that tells us how to construct or obtain the objects that are referred to in the statement. (For instance, a constructive proof of the formal version of statement "for every natural number n there is a larger number m" would give you method fo constructing m in the presence of n.)

This desire for constructivity is what motivates the development of a non-classical (and in particular, a *sub*-classical) logic, intuitionistic logic.

19.2 Propositional intuitionistic logic

19.2.1 Language and semantics

The language of intuitionistic propositional logic $\mathcal{L}_{\text{IL}}$ is the same as the language of classical propositional logic (cf. §7.3) with the addition that instead of taking $\bot$ and $\top$ as defined (cf. p. 7.4.1), we take them as primitive. What differs are the types of truth-value assignments that we allow and the interpretation of the connectives. Informally, we want to correlate the truth of a wff with giving a proof of the wff, and informally, we shall interpret the propositional constants and connectives as follows:

- There is no proof of $\bot$.
- The truth of $\varphi \wedge \psi$ is established by giving a proof of φ and a proof of ψ.
- The truth of $\varphi \vee \psi$ is established by giving a proof of φ or a proof of ψ.
- The truth of $\varphi \to \psi$ is established by giving a construction that takes a proof of φ and turns it into a proof of ψ.
- The truth of $\neg\varphi$ is established by giving a construction which, given a proof of φ would return a proof of $\bot$.

Many different semantics can be given to make this informal account precise (cf. [Chagrov and Zakharyas p. 24]); in this section, we define the semantics using the notion of Kripke frames and models first introduced in §11.3.1.

Definition 19.2.1 (Intuitionistic Kripke model)**.** An *intuitionistic Kripke model* $\mathfrak{M}^{\text{IL}}$ is a pair $\langle \mathfrak{F}, V \rangle$ where:

- $\mathfrak{F} = \langle W, R \rangle$ is a Kripke frame where R is a pre-order (that is, it is reflexive, transitive, and antisymmetric).
- V is a valuation function such that if $V(p, w) = T$ and wRw', then $V(p, w') = T$.

Unlike the Kripke models of Chapter 11, where we allowed arbitrary valuations on frames that meet a certain set of conditions, intuitionistic Kripke models put constraints both on the frame *and* on the valuation. Nevertheless, we can still speak of the frame underlying a given intuitionistic model, by abstracting away from the particular valuation.

Given an intuitionistic Kripke model, we can define the truth conditions for the intuitionistic connectives.

Definition 19.2.2 (Truth conditions)**.** The truth of a wff in $\mathcal{L}_{\mathrm{IL}}$ is evaluated at an individual world w of a given intuitionistic model $\mathfrak{M}$ recursively as follows:

$$
\begin{array}{lll}
\mathfrak{M}, w \not\models \bot & & \\
\mathfrak{M}, w \models p & \text{iff} & V(p, w) = T \\
\mathfrak{M}, w \models \varphi \wedge \psi & \text{iff} & \mathfrak{M}, w \models \varphi \text{ and } \mathfrak{M}, w \models \psi \\
\mathfrak{M}, w \models \varphi \vee \psi & \text{iff} & \mathfrak{M}, w \models \varphi \text{ or } \mathfrak{M}, w \models \psi \\
\mathfrak{M}, w \models \neg\varphi & \text{iff} & \text{for all } w' \text{ such that } wRw', \mathfrak{M}, w' \not\models \varphi \\
\mathfrak{M}, w \models \varphi \rightarrow \psi & \text{iff} & \text{for all } w' \text{ such that } wRw', \text{ if } \mathfrak{M}, w' \models \varphi \text{ then } \mathfrak{M}, w' \models \psi
\end{array}
$$

The definitions of consistency, validity in a model, validity in a frame, and validity on a class of frames carry over from ordinary modal logic. Similarly, just as we can prove correspondence results between modal formulas and classes of modal frames (cf. §11.3.1), we can prove results about the correspondence between $\mathcal{L}_{\mathrm{IL}}$ formulas and constraints on the pre-order.

But because of the extra constraints on the valuation and the stronger requirements for the truth of implications and negations, we can also prove more things about intuitionistic models. In particular, we can prove the following monotonicity condition:

Lemma 19.2.3. *Let $\mathfrak{M}^{\mathrm{IL}}$ be an intuitionistic model. For every wff φ and worlds w, w', if $\mathfrak{M}^{\mathrm{IL}}, w \models \varphi$ and wRw', then $\mathfrak{M}^{\mathrm{IL}}, w' \models \varphi$.*

Proof. By induction on the complexity of φ.

Basis cases: (1) The result holds for $\bot$ trivially. (2) Suppose that $\mathfrak{M}^{\mathrm{IL}}, w \models p$ for some world w, and suppose that wRw'. By the definition of V, it follows that $V(p, w') = T$, and hence by the definition of truth, that $\mathfrak{M}^{\mathrm{IL}}, w' \models p$.

Inductive steps: Assume that the result holds for wffs ψ_1 and ψ_2. We then need to show that it holds for each of $\neg\psi_1$, $\psi_1 \wedge \psi_2$, $\psi_1 \vee \psi_2$, and $\psi_1 \rightarrow \psi_2$. We prove only the case of $\psi_1 \rightarrow \psi_2$, and leave the others as an exercise for the reader. Suppose that wRw'. We will prove the result contrapositively, showing that if $\mathfrak{M}^{\mathrm{IL}}, w' \not\models \psi_1 \rightarrow \psi_2$, then $\mathfrak{M}^{\mathrm{IL}}, w \not\models \psi_1 \rightarrow \psi_2$. Suppose that $\mathfrak{M}^{\mathrm{IL}}, w' \not\models \psi_1 \rightarrow \psi_2$. By the definition of truth, there is a w'' such that $w'Rw''$, and $\mathfrak{M}^{\mathrm{IL}}, w'' \models \psi_1$ and $\mathfrak{M}^{\mathrm{IL}}, w'' \not\models \psi_2$. Because R is transitive, we know that wRw'', and hence $\mathfrak{M}^{\mathrm{IL}}, w \not\models \psi_1 \rightarrow \psi_2$, as required. □

19.2.2 Connections to other logics

Gödel's translation of CL into IL

IL is a subsystem of CL, in the sense that every theorem of IL is also a theorem of CL; but it is a proper subsystem in that there are theorems of CL which are not theorems of IL, since

$\vdash_{\mathrm{CL}} p \vee \neg p$ but $\nvdash_{\mathrm{IL}} p \vee \neg p$. However, despite the fact that CL is a proper extension of IL, it is possible to embed CL theorems in IL. One such embedding is given by Gödel's so-called "negative translation" [Bezhanishvili and de Jongh, 2005, §3.5]. In order to make use of this embedding, we must slightly modify the presentation of CL given above in Chapter 7; instead of taking $\neg\varphi$ as a primitive construction, we take $\bot$ as primitive and define $\neg\varphi$ as $\varphi \to \bot$. This merely allows us to talk about negation in both IL and CL in a uniform way.

Definition 19.2.4. Let φ be a wff in $\mathcal{L}_{\mathrm{CL}}$. We define φ^n recursively as follows:

$$\begin{array}{lcl} p^n & := & \neg\neg p \\ (\varphi \wedge \psi)^n & := & \varphi^n \wedge \psi^n \\ (\varphi \vee \psi)^n & := & \neg\neg(\varphi^n \vee \psi^n) \\ (\varphi \to \psi)^n & := & \varphi^n \to \psi^n \\ \bot^n & := & \bot \end{array}$$

The resulting φ^n is a wff in $\mathcal{L}_{\mathrm{IL}}$. Further, we can prove the following result:

Theorem 19.2.5. $\vdash_{\mathrm{CL}} \varphi$ *iff* $\vdash_{\mathrm{IL}} \varphi^n$*, that is, all* CL *theorems are provable in* IL in their translated form.

Proof.

($\Leftarrow$) Because CL is an extension of IL, if $\vdash_{\mathrm{IL}} \varphi^n$ then $\vdash_{\mathrm{CL}} \varphi^n$. It remains to be shown that $\vdash_{\mathrm{CL}} \varphi^n \leftrightarrow \varphi$, which we leave as an exercise to the reader.

($\Rightarrow$) By induction on the complexity of φ. □

Gödel's translation of IL into S4

The translation discussed in the previous section is not the only translation involving IL that Gödel defined. In [Gödel, 1933], Gödel provided an embedding of intuitionistic propositional logic in S4, that is, a method for translating wffs in $\mathcal{L}_{\mathrm{IL}}$ into wffs in $\mathcal{L}_{\mathrm{S4}}$ in a way that is validity preserving. In this section, we give two such embeddings, due to [Troelstra and Schwichtenberg, 2000, §9.2].

Definition 19.2.6. Let φ be a wff in $\mathcal{L}_{\mathrm{IL}}$. We define φ° and $\varphi^\Box$ recursively as follows:

$$\begin{array}{lcllcl} p^\circ & := & p & p^\Box & := & \Box p \\ \bot^\circ & := & \bot & \bot^\Box & := & \bot \\ (\varphi \wedge \psi)^\circ & := & \varphi^\circ \wedge \psi^\circ & (\varphi \wedge \psi)^\Box & := & \varphi^\Box \wedge \psi^\Box \\ (\varphi \vee \psi)^\circ & := & \Box\varphi^\circ \vee \Box\psi^\circ & (\varphi \vee \psi)^\Box & := & \varphi^\Box \vee \psi^\Box \\ (\varphi \to \psi)^\circ & := & \Box\varphi^\circ \to \psi^\circ & (\varphi \to \psi)^\Box & := & \Box(\varphi^\Box \to \psi^\Box) \end{array}$$

φ° and $\varphi^\Box$ are both wffs in $\mathcal{L}_{\mathrm{S4}}$.

Lemma 19.2.7. *The two embeddings* $^\circ$ *and* $^\Box$ *are equivalent in the following sense:* $\vdash_{\mathrm{S4}} \Box\varphi^\circ \leftrightarrow \varphi^\Box$.

Proof. By induction on the complexity of φ. □

Lemma 19.2.8. *Let* $\mathfrak{M}^{\mathrm{IL}}$ *be an intuitionistic Kripke model. Then,*

$$\mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{IL}} \varphi \quad \textit{iff} \quad \mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{S4}} \varphi^\Box$$

Proof. By induction on the complexity of φ. Note that on the left-hand side we use the truth conditions for the operators and connectives as defined in Definition 19.2.2, while on the right-hand side we use the truth conditions as defined in Definition 11.3.3.

Basis case: Assume that $\mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{IL}} p$. By the definition of $\vDash_{\mathrm{IL}}$, $V(p, w) = T$, because p is an atom. By the definition of $\mathfrak{M}^{\mathrm{IL}}$, if $V(p, w) = T$, $V(p, w') = T$ for every w' such that wRw'. But then by the definition of $\vDash_{\mathrm{S4}}$, $\mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{S4}} \Box p$, which is the same as saying that $\mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{S4}} p^\Box$, as desired. Because all of the moves are definitional, exactly the same argument can be run in the opposite direction.

Inductive step: Assume that $\mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{IL}} \psi$ iff $\mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{S4}} \psi^\Box$ and $\mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{IL}} \gamma$ iff $\mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{S4}} \gamma^\Box$. It remains to be shown that $\mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{IL}} \varphi$ iff $\mathfrak{M}^{\mathrm{IL}}, w \vDash_{\mathrm{S4}} \varphi^\Box$ where $\varphi = \neg\psi$, $= \psi \wedge \gamma$, $= \psi \vee \gamma$, or $= \psi \to \gamma$. We leave these as exercises for the reader. □

Let us think about what this translation means: It means that we can represent this constructive notion of provability via our familiar notions of modal logic. Each of the characteristic axioms of S4 represent a plausible provability principle:

K "If you can prove that p implies q, then if you can prove p then you can prove q."

T "If you can prove p, then p is valid."

4 "If you can prove p, then you can prove that you can prove it.

Similarly, the rule Nec can be interpreted as a plausible principle of provability: If φ is a theorem, then the fact that you can prove it is also a theorem.

Given this, it is not surprising at all that the following result can be proven:

Lemma 19.2.9. *The two embeddings* $^\circ$ *and* $^\Box$ *are sound, that is,* $\vdash_{\mathrm{IL}} \varphi$ *iff* $\vdash_{\mathrm{S4}} \varphi^\circ$ *iff* $\vdash_{\mathrm{S4}} \varphi^\Box$.

Proof. The equivalence of φ° and $\varphi^\Box$ is given in Lemma 19.2.7, so we only need to prove the equivalence between theoremhood in IL and theoremhood in S4 of one of the translations (we will focus on $\varphi^\Box$).[1]

[1] See also the proof of [Troelstra and Schwichtenberg, 2000, Theorem 9.2.3].

($\Rightarrow$) Suppose that $\vdash_{\mathrm{IL}} \varphi$. Since every CL is an extension of IL, it follows that $\vdash_{\mathrm{CL}} \varphi$, and hence φ is a tautology. It can therefore be the start of a proof in S4. We can then generate a proof of $\varphi^{\Box}$ in S4 inductively on the complexity of φ.

($\Leftarrow$) Immediate from Lemma 19.2.8.

□

19.2.3 Proof theory

19.3 Other constructive logics

Chapter 20

Paraconsistent logics [last modified 13 Sep 18]

20.1 A return to conditionals

When we introduced the truth tables for $\rightarrow$ (§7.4.1), we noted that some people object to the material account of the conditional on the basis of the so-called "paradoxes of material implication":

Paradoxes of material implication:

- if p is false, then $p \rightarrow q$ is true (that is, $\neg p \rightarrow (p \rightarrow q)$) is a tautology).
- if q is true, then $p \rightarrow q$ is true (that is, $q \rightarrow (p \rightarrow q)$) is a tautology).

The paradoxes can be avoided by adopting an account of the conditional which is strict, rather than material, but then you get the:

Paradoxes of strict implication

- $(p \wedge \neg p) \rightarrow q$
- $p \rightarrow (q \vee \neg q)$

We now look at two approaches which attempt to mitigate these paradoxes, connexive logics and relevance logics. Both of these share the motivation that the paradoxes arise because there is no connection between the antecedent and the consequence, and that there needs to be some type of connection between them that grounds the truth of the conditional. They differ in how they explicate what this connection is and how to understand it.

20.1.1 Connexive logics

There is no single logic which is *the* connexive logic, but rather "connexive logics" is the name of a family of logics which are united by their agreement on certain principles. Two pairs of principles which are not valid classically but which are valid in connexive logics are colloquially known as "Aristotle's theses" and "Boethius's theses":

Aristotle's theses are derived from a comment he makes in the *Prior Analytics* that "it is impossible that if A, then not A":

- $\neg(\neg\varphi \to \varphi)$
- $\neg(\varphi \to \neg\varphi)$

Both of these principles are plausible if φ is neither a contradiction nor a tautology; where connexive logics go beyond classical logic is by saying that these also hold of contradictions and tautologies.

Boethius's theses are so-called because of a principle that Boethius held, that "the negation of 'if A then B' is 'if A then not B'." However, it is not strictly speaking felicitous to attribute the following theses to Boethius, since his principle is regarding terms, not propositions. Nevertheless, the two theses are:

- $(\varphi \to \psi) \to \neg(\varphi \to \neg\psi)$
- $(\varphi \to \neg\psi) \to \neg(\varphi \to \psi)$

None of these four are theses of classical logic, so connexive logics are all non-classical. They are neither subsystems nor extensions of classical logic.

20.1.2 Relevance logics

Relevance logics also argue that the consequent needs to be 'relevant' to the antecedent in some way in order for the conditional to be true. One way they cash out this idea of relevance is via the Variable Sharing Principle (VSP):

> No formula of the form $\varphi \to \psi$ is provable unless
>
> there is some atom p occurring in both φ and ψ. (VSP)

The VSP is only a necessary condition, not a sufficient one; and it alone doesn't get rid of all the paradoxes—for example, in $p \to (q \to p)$, an instance of one of the paradoxes of material implication, there is an atom which occurs in both the antecedent and the

consequent, namely p. In order to get rid of the paradoxes, the VSP needs to be augmented with semantics.

In relevance semantics, $\wedge$ and $\vee$ are classical; however, $\neg$ and $\rightarrow$ are non-classical (that $\rightarrow$ will be is obvious; that $\neg$ must also be is clear from the fact that you can define $\rightarrow$ in terms of $\neg$ and $\vee$, so if both of the latter were classical, $\rightarrow$ would be as well). The standard semantics for relevance logics are the Routley-Meyer ternary relation semantics. Instead of relating pairs of worlds, we instead relate triples of worlds. Here, we only give the semantics for $\rightarrow$:

Definition 20.1.1 (Relevance truth conditions for $\rightarrow$)**.**

$$\begin{array}{lll} w \vDash \varphi \rightarrow \psi & \text{iff} & \text{for all } xy \text{ such that } Rwxy, \quad \text{either} \\ & & x \vDash \neg\varphi \qquad\qquad\qquad\qquad\quad \text{or} \\ & & y \vDash \psi \end{array}$$

The question then is how to give an intuitive interpretation of these semantics so we can understand what the above definition means. Three ways have been offered:

1. $Rwxy$ means "the combination of the information in w and x is contained in y'.

2. Worlds are 'sites' and arrows are 'channels' of information. $Rwxy$ is true if w is a channel between x and y. $w \vDash \varphi \rightarrow \psi$ iff whenever there is a channel w between x (where φ is true) and y, then $y \vDash \psi$.

3. Situated implication: On this view, worlds are 'situation's, or partial representations of the universe. The information combined in two of these partial representations can warrant the inference to information that goes beyond the combined information. A new type of operator is introduced: $Ixy\varphi$ holds if the information in x and y together licence an inference to a situation where φ is true.

   Let $|\varphi|$ be the set of situations where φ is true. Then:

$$x \vDash \varphi \rightarrow \psi \quad \text{iff} \quad \text{for all } y \text{ such that } y \vDash \varphi, Ixy|\varphi|$$

Here is one way you could understand $Rwxy$: $Rwxy$ is true iff in w I have eggs, flour, and milk, and in x I have sugar, then in y I have cake. Thus, $w \vDash \varphi \rightarrow \psi$ is the same as saying "I have eggs and flour and milk, and if I am given sugar in addition to these, then I can make cake"; what I have in world w combined with what is in world x is all I need to get to world y.

Chapter 21

Many-valued and fuzzy logics [last modified 12 Feb 26]

In this chapter we look at a variety of logics that are all characterisable by the dropping of the Law of Excluded middle, and allowing more than two truth values.

Any of the logics we have discussed in Parts II and III can be turned into a many-valued logic by this method.

21.1 Many-valued modal logic

Traditionally, modal logic, whether propositional or first-order, is based on classical logic, where the only truth values are true and false, and every wff is either true or false and not both. We can extend modal logic by extending the number of truth values, and the definitions for truth, to more than two, including infinitely many.

The simplest way to do this is to simply change the definition of valuation and truth so that instead of building modal logic on classical two-valued logic, we build it on top of a three-valued logic, which instead of having just two values, T and F, has a third, typically U ("undefined").[1] There are a variety of candidate three-valued logics, including Kleene's strong three-valued logic [Kleene, 1950, pp. 332–340], Lewis and Langford's three-valued logic [Lewis and Langford, 1932b, pp. 213ff], and Łukasiewicz's logic as formalized by Wajsberg [Łukasiewicz, 1920, Wajsberg, 1931, Goldberg et al., 1974].

[1] There are a number of motivations for this third value; of especial note in the present context is Łukasiewicz's motivation for his three-valued logic, which was distinctly modal in character. His third value was interpreted as representing "the possible," or more precisely, "the contingent" [Woodruff, 1974, p. 433].

Each of these logics have different applications stemming from different motivations. Because we are interested in how three-valued logics interact with modality, we do not go into the motivations and applications of the propositional versions here. Instead, we simply take the propositional component as given. Kleene's strong three-valued logic is characterized by the following truth-tables:

$\neg\varphi$

φ	
T	F
U	U
F	T

$\varphi \vee \psi$

	ψ	T	U	F
φ	T	T	T	T
	U	T	U	U
	F	T	U	F

$\varphi \wedge \psi$

	ψ	T	U	F
φ	T	T	U	F
	U	U	U	F
	F	F	F	F

$\varphi \rightarrow \psi$

	ψ	T	U	F
φ	T	T	U	F
	U	T	U	U
	F	T	T	T

Lewis and Langford's logics differ from Kleene's in the definition of the conditional: A conditional where both the antecedent and consequent are undefined is *true*, rather than undefined [Lewis and Langford, 1932b, pp. 213ff], [Goldberg et al., 1974, p. 326]. Łukasiewicz's logic uses the LL conditional, but also changes the definition of disjunction: A disjunction where both disjuncts are undefined also comes out to be *true* [Sioson, 1964, p. 63]. A fourth possibility for the propositional connectives is considered by Segerberg in his investigation of Prior's logic Q; in this approach, any wff with input U receives output U [Segerberg, 1967].

But where these logics become truly interesting is how the modal operators are defined. Some many-valued modal logics still require that the modal operators themselves take on

only the values T and F (or 1 and 0)[2]; but others allow modal propositions to take on the new values as well.[3] If modal wffs are allowed to take on the third value, we can then make a distinction between two types of necessity—strong necessity, where the wff is *true* in every accessible world, and weak necessity, where the wff is merely *not false* in every accessible world [Segerberg, 1967, p. 54].

While it is natural to interpret the third value as "undefined" (i.e., neither true nor false) or "meaningless", this is not the only interpretation of the third value that is possible. Another alternative is to consider the three options not as truth values but as *probabilities of truth*. In this case, T is identified with probability 1, F with probability 0, and the third "truth value" is naturally probability "$\frac{1}{2}$". But once one has a probability of $\frac{1}{2}$, why stop there? Probability distributions are generally over the closed interval $[0, 1]$, with infinitely many values between the two borders. Such infinitely valued logics can be used to reason about probabilities or to reason with probabilities (probabilistic modal logic[4]) as well as to reason about uncertainties (fuzzy modal logic [Hájek, 2010, Mironov, 2005]).

[2]Cf. Scotch et al.'s three-valued modal logic $\text{Ł}_3\text{M}_2$ based on Łukasiewicz's semantics [Schotch et al., 1978].

[3]Cf. Scotch et al.'s three-valued modal logic $\text{Ł}_3\text{M}_3$ and Ostermann's logics, all based on Łukasiewicz's semantics [Schotch et al., 1978, Ostermann, 1988], and the logics in [Morikawa, 1989, Segerberg, 1967].

[4]When combining modality and probability there are multiple different ways. One can add modalities to probability logic, instead of to propositional logic, so that instead of associating each possible world with the set of worlds accessible to it, we associated with it instead a probability distribution, probability space, or set of probability distributions [Demey et al., 2019, §4]. Alternatively, one can take probalistic semantics, and use them at each world [Cross, 1993].

Chapter 22

Probabilistic logics [last modified 17 Jun 19]

The natural progression from the many-valued logics of the previous chapter takes us to the topic of this chapter: logics of probability, probability logics, and probabilistic logics. As a supplement to this chapter, we recommend [Adams, 1998].

Chapter 23

Inductive and abductive logics [last modified 13 Sep 18]

Cf. §**??** for discussion of relevant concepts in the Southeast Asian traditions.

Part V

Back matter

Appendix A

Exercises [last modified 12 May 25]

A.1 Exercises for Chapter 1

Exercise 13. Give examples of arguments according to Definition 1.1.2.

Exercise 14.

- Name three logicians who lived before 1000.
- Name three logicians who lived (and died) between 1000 and 2000.
- Name three logicians who lived after 2000.

Pick one from each category and write 100 words (with references) on what is distinctive about their life works/views.

Note: Wikipedia is not an acceptable source to cite.

A.2 Exercises for Chapter 3

Exercise 15. Consider the following $n+1$-player game: There are n players called **agents**, denoted by a_i, and one distinguished player denoted by G. The game has two rounds; in the first round, the agents simultaneously and independently make a decision α_i between two actions, good and bad; in the second round, after seeing the decisions of the agents, G makes n decisions (one for each player) $\gamma_i \in \{\oplus, \ominus\}$. A **run** of the game is a function $F : i \mapsto \langle \alpha_i, \circ \rangle$ where $\circ \in \{\oplus, \ominus\}$. If F is a run of the game, we say that $\circ$ **is the fate of agent** a_i **in** F if $F(i) = \langle \alpha_i, \circ \rangle$.

Use this game model to analyse the following two texts from the *Catholic Encyclopedia*:

> **The Catholic Encyclopedia on Predestination.** The principal question then is: Does the natural merit of man exert perhaps some influence on the Divine election to grace and glory? If we recall the dogma of the absolute gratuity of Christian grace, our answer must be outright negative.
>
> **The Catholic Encyclopedia on Grace.** Beside the necessity of actual grace, its absolute gratuity stands out as the second fundamental question in the Christian doctrine on this subject. The very name of grace excludes the notion of merit. But the gratuity of specifically Christian grace is so great and of such a superior character that even mere natural petition for grace or positive natural dispositions cannot determine God to the bestowal of his supernatural assistance.

Define formally what *strategies* for the players in the game would be and find a property for G's strategy that corresponds to the "absolute gratuity of Grace". If τ is a strategy for G, we say that "**the fate of a_i is predetermined relative to τ** if there is some $\circ \in \{\oplus, \ominus\}$ such that for every run of the game in which G plays according to τ, the fate of a_i is $\circ$.

Prove that if τ satisfies "absolute gratuity of Grace", then the fate of every agent is predetermined relative to τ.

Exercise 16. Let W be a nonempty set of states and $R \subseteq W \times W$ an accessibility relation. We say "state v is conceivable by anyone in state w" for wRv. Let X be a nonempty set of objects, and $E \subseteq W \times X$ a relation. We say "object x exists in state w" for wEx. For each $w \in W$, we have an order $<_w$ of X, and we say "in state w, object x is better than object y" for $y <_w x$.

We call $\langle W, R, X, E, \langle <_w ; w \in W \rangle \rangle$ an **ontological frame** if R is reflexive (*i.e.*, w is conceivable by anyone in state w), and the following principle *"Existence is better than nonexistence"* (EBN) holds:

(EBN) For all x, y and w, if wEx and $\neg wEy$, then $y <_w x$.

The central argument of Anselm's ontologic proof is "if something is such that nothing better can be conceived, then it must exist". Formulate this argument in the language of ontological frames and prove it.

Given an example of an ontological frame where there is no object "such that nothing better can be conceived".

Exercise 17. Many medieval authors think of disjunction as an operator on finite sets of sentences and define $\mathsf{MD}(A_1, ..., A_n)$ to be true if exactly one of the A_i is true.

If f is a binary truth function (i.e., a function from $\{0,1\} \times \{0,1\}$ to $\{0,1\}$), we can use it to recursively define n-ary truth functions by

$$\begin{aligned} f_2(A,B) &:= f(A,B) \\ f_{n+1}(A_0,...,A_n) &:= f(f_n(A_0,...,A_{n-1}),A_n). \end{aligned}$$

We say that an n-ary truth function g is **induced by** f if $g = f_n$.

1. Show that medieval disjunction MD is not induced by any binary truth function.

2. Let h be exclusive disjunction (i.e., $h(0,0) = h(1,1) = 0$ and $h(0,1) = h(1,0) = 1$). Prove that $h_{2n}(A_0,...,A_{2n-1})$ is true if and only if an odd number of the A_i is true and that $h_{2n+1}(A_0,...,A_{2n})$ is true if and only if an even number of the A_i is true.

Exercise 18. Consider a set W of states and a set X of objects. We call the set $\ulcorner X := \{+,-\} \times X := \{\langle +,x\rangle\,;\, x \in X\} \cup \{\langle -,x\rangle\,;\, x \in X\}$ the set of **entities**. We think of $\langle -,x\rangle$ as the imagined object x and $\langle +,x\rangle$ as "$\langle -,x\rangle$ with the added property of existence". We call entities $\langle +,x\rangle$ **existing entities**.

For each $w \in W$, fix a set $X_w \subseteq \ulcorner X$ of **permissible entities** in w. We fix two strict linear ordering $<$ and $\prec$ on $\ulcorner X$, and an accessibility relation R on W. As in Exercise 16, we say that "v is conceivable from w" if wRv. For $xy \in X_w$, we say "in w, x is better (bigger) than y" if $y < x$ ($y \prec x$). A structure $\mathbf{W} := \langle W, \langle X_w\,;\, w \in W\rangle, R, <, \prec\rangle$ is called **Anselmian** if it has the following properties:

- If wRv and $\ulcorner x \in X_w$, then $\ulcorner x \in X_v$.
- For each $w \in W$, if $\langle -,x\rangle \in X_w$, then there is some v such that wRv and $\langle +,x\rangle \in X_v$.
- For each $x \in X$, $\langle -,x\rangle < \langle +,x\rangle$.

If $\mathbf{W}$ is an Anselmian structure and $w \in W$, we say that an entity $\ulcorner x \in \ulcorner X_w$ is **Anselmian** in w if for all v such that wRv and all $\ulcorner y \in X_v$, it is not the case that $\ulcorner x < \ulcorner y$. We say that an entity $\ulcorner x \in \ulcorner X_w$ is **Gaunilan** in w if for all v such that wRv and all $\ulcorner y \in X_v$, it is not the case that $\ulcorner x \prec \ulcorner y$.

The second half of the ontological argument can now be rephrased as follows: In an Anselmian structure, every Anselmian entity is existing. Prove this statement.

Give an example of an Anselmian structure with a state w in which there is a nonexisting Gaunilan entity (*i.e.*, an entity of the form $\langle -,x\rangle$).

There is a simple modification of the notion of an Anselmian structure that we could call a **Gaunilan structure**, for which we can prove that every Gaunilan entity is existing. Give a precise definition of this and prove the statement.

Consider your definition of a Gaunilan structure. It is possible to justify the new axiom as "true" in some natural sense? Could you convince a nonbeliever of the axioms of your Gaunilan structure? Give a brief discussion (at most 10 lines).

Exercise 19. Consider the sentence *tantum omnis philosophus albus est* ("only every philosopher is white", i.e., every philosopher is white but nothing else is).

1. In this sentence, the supposition of *albus* is not *suppositio mobilis*. Why?[1]

2. Give a modern semantics for the *tantum omnis* construction: suppose we have a universe of discourse X and two predicates $\Phi, \Psi \subseteq X$. Give a formal definition such that

   $$\mathbf{tantumomnis}(\Phi, \Psi)$$

   is true if and only if *tantum omnis* Φ *est* Ψ ("only every Φ is Ψ").

Note. The "modern semantics" is not necessarily unique. There might be different semantics that describe the natural language sentences reasonably adequately.

Exercise 20. If X is any set and $\wp(X)$ is its power set (the set of all subsets of X), we call $\mathsf{Q} \subseteq \wp(X)$ a **generalized quantifier**. If $\Phi \subseteq X$ is a predicate on X, we say that $\mathsf{Q}\Phi$ holds (in words: "for Q-many x, $\Phi(x)$ holds") if $\Phi \in \mathsf{Q}$.

1. Let $\forall := \{X\}$ and $\exists := \{A \subseteq X \,;\, A \neq \varnothing\}$. Argue that $\forall\Phi$ and $\exists\Phi$ have the intended meanings "for all x, $\Phi(x)$ holds" and "there is an x such that $\Phi(x)$ holds".

2. (Suppose that X is infinite for this part.) Paraphrase the meanings of $\mathsf{Q}_0\Phi$, $\mathsf{Q}_1\Phi$, and $\mathsf{Q}_2\Phi$ in words: $\mathsf{Q}_0 := \{A \subseteq X \,;\, A \text{ is finite}\}$, $\mathsf{Q}_1 := \{A \subseteq X \,;\, X \setminus A \text{ is finite}\}$, $\mathsf{Q}_2 := \{A \subseteq X \,;\, A \text{ is infinite}\}$.

3. Fix some $x \in X$ and give a definition of a generalized quantifier op_x that corresponds to the *omnis praeter* construction from Exercise 19.

Exercise 21. Consider the sophisma

$$(\star)\ \textit{omnis homo praeter Socratem excipitur}$$

("every man except for Socrates is excepted").

1. Give a background story which describes a situation in which $(\star)$ is true

[1] Note: This part of the problem was removed in the next year.

2. Argue informally that ($\star$) is false.

3. Solve the apparent contradiction by explaining the fallacy as a *secundum quid et simpliciter*.

Exercise 22. Consider the following *obligatio* dialogue. The underlying assumptions are that the two dialogue partners are in Amsterdam, neither of them is actually the Pope, that the Pope is in Rome, and that the opponent is married. Fill in the answers for the respondent according to Burley's system of obligations, once for a respondent who knows the underlying assumptions and can do propositional deductions, and once for a respondent who in addition knows that the Pope is not married and can use that in his reasoning about implications (for example, from "x is married" he can infer "x is not the Pope".) Explain all of the moves according to the rules of Burley's *obligationes*.

One of us two is the Pope.

I admit it.

I am married.

— — —

You are in Amsterdam.

— — —

You are the Pope.

— — —

Cedat tempus.

Exercise 23. In this exercise, we consider the systems of *positio* as described by Walter Burley and Roger Swyneshed. If a *positum* φ^* is given and φ_k (for $0 \leq k \leq n$) are proposed sentences of the **Opponent**, we let Φ_k^{Burley} be the set of **"currently accepted truths"** according to Burley's system on the basis of the sequence $\langle \varphi^*, \varphi_0, \ldots, \varphi_n \rangle$.

Prove the following properties of the two systems:

1. If the *positum* φ^* is consistent, then for all $k \leq n$, the set Φ_k^{Burley} is a consistent set.

2. If the *positum* φ^* is consistent and $k < \ell \leq n$ with $\varphi_k = \varphi_\ell$, then the Res in a Swyneshed-style *positio* will give the same answer in steps k and ℓ of the *obligatio*.

A.3 Exercises for Chapter 4

Exercise 24. For each of the combinations in the Wheel of Reasons (Figure 4.1), come up with an H and an S that fulfil the requirement of the section. For example, for section (1), let H be 'thing' and S be 'cow'. Then being a thing occurs in all cows and in all non-cows.

Exercise 25. Of the sections in the Wheel of Reasons (Figure 4.1), only two give rise to good inferences. Which sections are excluded by the first requirement of the *trairūpya-hetu*? Which sections are excluded by the second? Which sections are excluded by the third?

Exercise 26. Read [Tachikawa, 1971].

- What are the three marks of a correct *hetu*?
- What are the two types of *dṛṣṭānta*, and what are the different roles they play?
- Match the fallacious marks (§3.2) with the relevant quadrants of the Wheel of Reasons (in so far as you can).

A.4 Exercises for Chapter 6

Exercise 27. Let $\mathcal{L}_t$ be a term language containing the basic terms A, I, M, P, W.

1. Translate each of the sentences below into $\mathcal{L}$.
2. Identify the quality and quantity of each sentence (cf. Definitions 6.1.6,6.1.7).
3. Define an interpretation of the basic terms that makes as many of the sentences true as you can (cf. Definition 6.2.3).
4. What is the contradictory of (1)? The subcontrary of (1)? The superaltern of (1)? (cf. Definitions 6.2.11,6.2.13,6.2.15,6.2.17).
5. Identify at least three syllogisms that can be formed from these sentences (cf. Definition 6.1.8). Construct a possible conclusion for each (this conclusion can be, but need not be, one of the sentences below).

Sentences:

1. Some melm doesn't work at the institute.
2. Every painter is a melm.
3. Someone who works at the institute is a painter.
4. Everyone in Andersthal is a weaver.
5. No weavers are melm.
6. Some weaver is a painter.

7. No painter lives in Andersthal.

8. All painters work at the institute.

Exercise 28. Find the valid syllogisms hidden in the following three (completely fictitious) texts and formalize them using the notation for syllogisms used above (e.g., describe the natural language terms corresponding to A, B and C, and then give a mood like "$A\mathrm{a}B, B\mathrm{i}C : A\mathrm{i}C$" for **Darii**).

> **Text A.** "Yesterday, the newspaper reported that some of the Deans inaugurated yesterday published heavily anarchist papers in the 1970s. The whole concept of an anarchist being the Dean at our university is ridiculous. One of the main duties of the Dean is being the head of our administration. Being well-organized and hierarchically minded are necessary conditions for this job, actually even for aspiring to hold that office. Therefore, we can be sure that all new Deans have given up the anarchist convictions from the 1970s if they ever had them."

> **Text B.** "I am afraid that you do not have the right to deny Bartleby access to a desk and a computer. The following is part of the agreement in our special programme that every participating institution has to sign: every participating institution which sends graduate students to the United States has to provide infrastructure for visiting students from US universities in reciprocation. As every single Dutch university sent at least one student to the US between 1991 and 2003, the student Bartleby from Indiana University will have the right to demand a desk and a computer at each university in the Netherlands."

> **Text C.** *A Socratic dialogue overheard on the corridor of an unnamed math department.* "All of my students deserve their grades. My grading is absolutely fair."
>
> "Well, didn't you tell me that you caught a cheating student last week?"
>
> "Yes."
>
> "But could you prove that this student was cheating?"
>
> "Well, not exactly: I lost the cheat sheet after the exam, and when I wanted to fail him, he challenged me to present proof. Since I lost the cheat sheet, I couldn't present proof and had to give him the grade according to his exam book."

"Do I recall correctly that you gave him an A−?"

"Alas, it was even a straight A..."

"And you think that cheating students deserve an A?"

"Of course not."

"So, some of your students didn't deserve their grades."

Exercise 29. Take the set $X \coloneqq \{\mathsf{Socrates}, \mathsf{Aristotle}, \mathsf{Plato}\}$ as a universe of discourse. Define predicates on X to show that the following four moods are invalid.

- $A\mathrm{i}B, C\mathrm{i}B : A\mathrm{i}C$.
- $A\mathrm{o}B, B\mathrm{a}C : A\mathrm{a}C$.
- $A\mathrm{e}B, B\mathrm{i}C : A\mathrm{e}C$.
- $B\mathrm{a}A, C\mathrm{a}B : A\mathrm{a}C$.

Identify the figure of the four invalid moods given above.

Exercise 30.

- Represent each syllogism symbolically and identify the syllogism's figure.
- For those which are valid, identify the syllogism's mood.
- For those which are not, provide a counterexample.

Use the following key:

G Greek

M man

O mortal (thing)

1 Some mortal thing is a man.
Every Greek is mortal.
Some Greek is a man.

2 All Greeks are men.
No man is mortal.
Some mortal thing is Greek.

3 Every man is mortal.
No Greek is mortal.
No Greek is a man.

4 No Greek is mortal.
Some man is Greek.
Some man is not mortal.

5 Some Greek is not mortal.
Some man is not Greek.
Some man is not mortal.

6 No Greek is mortal.
Some man is not Greek.
All men are mortal.

7 Some mortal thing is not Greek.
Some man is mortal.
Some man is not Greek.

8 No mortal is Greek.
Some man is not Greek.
Some man is mortal.

9 Every mortal thing is Greek.
Some mortal thing is not a man.
Some Greek is not a man.

10 No man is mortal.
Some Greek is a man.
Some Greek is not mortal.

11 Some Greek is a man.
No mortal is Greek.
All mortals are men.

12 Some Greek is a man.
All Greeks are mortal.
Some man is not mortal.

Exercise 31. Construct an interpretation that makes as many of the following sentences true as possible:

1. $R\,a\,T$
2. $S\,i\,T$
3. $T\,o\,S$
4. $P\,e\,T$
5. $S\,i\,P$
6. $P\,a\,R$

Exercise 32. Determine whether the following sets of categorical propositions are consistent (cf. Definition 6.2.10). If they are consistent, give an interpretation that makes them all true. If they are not, explain why there is no interpretation.

1. $\{G\,a\,F, F\,i\,H, F\,o\,H, T\,e\,G, H\,a\,T, G\,o\,H\}$
2. $\{S\,i\,Q, M\,a\,S, F\,i\,Q, T\,o\,S, J\,a\,M, Q\,e\,J\}$
3. $\{B\,e\,A, C\,a\,B, A\,a\,D, C\,o\,D, D\,i\,B, D\,i\,C\}$

Exercise 33. Give an argument (in English) that $R\,a\,T$ and $R\,e\,T$ cannot both be true at the same time. Give an interpretation that shows that both can be false at the same time.

Exercise 34. Give an argument (in English) that $R\,i\,T$ and $R\,o\,T$ cannot both be false at the same time. Give an interpretation that shows that both can be true at the same time.

Exercise 35. Give an argument (in English) that if $R\,i\,T$ is true then $R\,e\,T$ is false, and vice versa.

Exercise 36. Give an argument (in English) that if $R\,i\,T$ is true then $T\,i\,R$ is true.

Exercise 37. Constuct an interpretation that makes $P\,a\,R$ true and $R\,a\,P$ false.

Exercise 38. Show that the following are not valid syllogisms:

1. $A\,i\,B, A\,i\,C$, therefore $B\,a\,C$.
2. $A\,e\,B, C\,o\,B$, therefore $A\,o\,C$.
3. $A\,a\,B, B\,i\,C$, therefore $A\,o\,C$.
4. $A\,o\,B, A\,a\,C$, therefore $B\,e\,C$.

Exercise 39.

1. Describe a formal system that is able to express sentences of the type "All B are A", "Some B are A", "No B are A" and "Some B are not A" and their negations such that the square of oppositions gives the rules of derivation.
2. Extend the language by expressions for "there are A" and give the appropriate rules. Assume "It is not the case that there is a chimera which is a human being" and derive "There are chimeras" in your system.

Exercise 40. The following three pseudo-syllogisms are sometimes called "indirect moods of the first figure":

AeB, BaC : CeA **Celantes**,
AaB, BiC : CiA **Dabitis**,
AaB, BeC : CoA **Fapesmo**.

1. Why aren't these real syllogisms?
2. Each of these "indirect moods" corresponds to one of the valid moods of the fourth figure. Find the right mood and explain the correspondence.
3. Explain all of the letters in the names **Celantes**, **Dabitis** and **Fapesmo** in terms of the medieval mnemonics. For this, give a formal proof of the indirect moods from the perfect syllogisms.

Exercise 41.

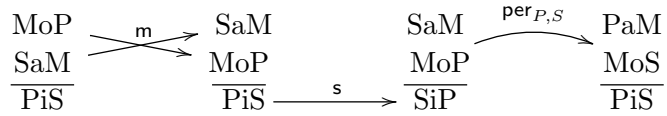
1. Give formal proofs of **Baroco** (BaA, BoC: AoC) and **Camestres** (BaA, BeC: AeC), explaining all the letters in the names.

Exercise 42. A categorical proposition is called **partial** if it has 'i' or 'o' as a copula. Let M be a mood such that both premises of M are partial. Prove that BCDF $\not\vdash M$.

Exercise 43. Let $\mathfrak{B}_{\mathrm{GH}} \coloneqq \{\mathbf{Giliri}, \mathbf{Halodri}\}$

$$\text{where } \mathbf{Giliri} \text{ is } \frac{\mathrm{PiM}\quad \mathrm{MiS}}{\mathrm{PiS}} \qquad \text{and } \mathbf{Halodri} \text{ is } \frac{\mathrm{PaM}\quad \mathrm{MoS}}{\mathrm{PiS}}.$$

For example, the following is a $\mathfrak{B}_{\mathrm{GH}}$-proof:



Following the proof, the mood MoP, SaM:PiS could be called **Homalis**.

Give $\mathfrak{B}_{\mathrm{GH}}$-proofs in the graphic representation and find names consistent with the medieval mnemonics for the following three moods:

$$\frac{\mathrm{SaP}\quad \mathrm{PaM}}{\mathrm{MoS}} \qquad \frac{\mathrm{MiP}\quad \mathrm{SeM}}{\mathrm{PeS}} \qquad \frac{\mathrm{PeM}\quad \mathrm{MiS}}{\mathrm{PeS}}.$$

Exercise 44. Read [Richman, 2004] and give an explanation of what Richman calls a *nonstandard figure*.

Exercise 45. The following is a syntax for a term logic in the Aristotelean style. We have term variables $\mathtt{t}_i$ and symbols **All**, **Some**, **No** and **Somenot**, where, *e.g.*, $\mathbf{Some}(\mathtt{t}_0, \mathtt{t}_1)$ is interpreted as "Some $\mathtt{t}_0$ is $\mathtt{t}_1$," *etc.* If **S** is one of the four symbols and $\mathtt{t}_0$ and $\mathtt{t}_1$ are term variables, then $\mathbf{S}(\mathtt{t}_0, \mathtt{t}_1)$ is called a *clause*. We now add two operators + and − to the language. If C is a clause, then both $+C$ and $-C$ are *statements*, interpreted as "C is true" and "C is false", respectively.

A *rule* for this syntax is a diagram

$$\frac{S}{S'}$$

where S and S' are statements. For instance,

$$\frac{+\mathbf{All}(\mathtt{t}_0, \mathtt{t}_1)}{-\mathbf{No}(\mathtt{t}_0, \mathtt{t}_1)}$$

is a rule, interpreted as "if all $\mathtt{t}_0$ are $\mathtt{t}_1$, then it cannot be true that no $\mathtt{t}_0$ is $\mathtt{t}_1$".

Give all of the rules corresponding to the relationships represented in the square of opposition (for example, the above rule corresponds to one instance of "contraries cannot both be true at the same time").

Exercise 46. Two of four relationships in the Square of Oppositions directly correspond to conversion rules in Aristotelian syllogistics. Which ones and why?

Exercise 47. Which of the following English-language arguments are syllogistic in form?

1. Every number is either even or odd. Every even number is divisible by itself. Every odd number is divisible by itself. Therefore every number is divisible by itself.

2. Socrates is a human. All humans are mortal. Therefore Socrates is mortal.

3. "Every art and every inquiry, and similarly every good action and pursuit, is thought to aim at some good; and for this reason the good has rightly been declared to be that at which all things aim" (Aristotle, *Nicomachean Ethics*.)

4. All water is wet. A river is water. Therefore, a river is wet.

5. Human is a species. That human is human. Therefore, that human is a species.

6. Risible is a property. A human is risible. Therefore, a human is a property.

7. Everything that has feet runs. A bridge has feet. Therefore a bridge runs.

8. Not every person runs. Not all people are Greek. Some Greek person doesn't run.

9. Some person is not Greek. All politicians are people. Some politician is non-Greek.

10. Whenever people put off marriage until they are older, the divorce rate decreases. Today people are putting off marriage until they are older. Therefore, the divorce rate is decreasing today.

11. Venus completes its orbit in less time than the Earth, because Venus is closer to the sun.

12. No canaries are birds. Some parrots are birds. All snakes are canaries. Therefore some parrots are not canaries.

13. All men are stones. Some statues are stones. Therefore some men are statues.

Exercise 48. For the arguments in the previous exercise ones that are not syllogisms, why are they not?

Exercise 49. For the syllogisms in Exercise 47, define a term language and translate the syllogism into that language, and answer the following questions:

a) What figure is each syllogism?

b) Which syllogisms are valid?

c) For the ones that are valid, name the mood.

d) For the ones that are invalid, give a counterexample.

A.5 Exercises for Chapter 7

Exercise 50. In the following, (a) identify the lekta and (b) produce examples of each of the five basic undemonstrated forms of arguments (see §7.1) by replacing "the first", "the second", with lekta of your choice.

1. If there are gods, then the universe is conducted according to divine foresight.
2. If the earth is flying, then the earth exists.
3. If the earth is flying, then it has wings.
4. If he is moving, then he is walking.
5. If it is day, it is light.
6. If it is night, it is dark.
7. If it is day, then I am conversing.
8. If it is night, then I am conversing.
9. If it is night, then it is day.
10. Either you will marry a beautiful woman or you will marry an ugly one.
    If she is beautiful, you will share her with others.
    If ugly, she will be a punishment.
    But neither of these things is desirable.
    Therefore, do not marry.

11. Pleasure is either good or bad or neither good nor bad.

12. Scipio was the son of Paulus and was twice consul and triumphed and was censor and was colleague in the censorship of L. Mummius.

13. Scipio was the son of Paulus and was twice consul and triumphed and was censor and was colleague in the censorship of L. Mummius and he overcame Hannibal in Africa.

Exercise 51. Read *Adv. Math.* VIII, 108ff and 112ff, *Hyp. Pyrrh.* II, 110ff, and *Vitae* VII, 81 [Mates, 1961].

1. What is the antecedent of a conditional? What is the consequent?

2. What are Philo's truth conditions for conditionals? What is his argument for them?

3. What are Diodorus's truth conditions for conditionals? What is his argument for them? Why does he reject Philo's account?

4. Whose view is being articulated in the *Vitae* text?

Exercise 52. Read *Adv. Math.* VIII, 215ff [Mates, 1961].

1. Give an example of the argument form using some of the lekta identified above in place of "the first", "the second", etc.

2. Is the argument form given a good one? Explain your answer.

Exercise 53. Read *Adv. Math.* VIII, 223 [Mates, 1961].

1. What are the two ways in which 'undemonstrated' can be used?

2. What is a simple undemonstrated argument?

3. What is a non simple one?

Exercise 54. For each of the following, identify whether the sentence is (a) simple, (b) a negation, (c) a conjunction, (d) a disjunction, or (e) a conditional. If it is a conjunction, what are the conjuncts? If it is a disjunction, what are the disjuncts? If it is a conditional, what is the antecedent and what is the consequent? Are any of the conjuncts, disjuncts, antecedents, or consequents themselves complex? If so, which of types (a)–(e) are they?

1. Boeing assembles planes or Lockheed constructs rockets.

2. Butane is a hydrogen compound.

3. Either you do your homework or you will not pass the module.
4. Ford builds trucks and Chrysler builds minivans.
5. Honda and Suzuki build motorcycles.
6. If your hook isn't baited, you won't catch a fish.
7. It is false that the sun shines in Durham.
8. James Joyce wrote *Ulysses*.
9. James will come to the party only if Alice does and Andrew doesn't.
10. London is larger than Durham, but Durham is a nicer place to live.
11. Rolex does not make cornflakes.
12. Sara does not enjoy logic.
13. Socrates runs; however, Plato reads.
14. You will not pass the module unless you attend tutorials.
15. You won't catch a fish unless your hook is baited.

Exercise 55. Define a propositional language and translate the sentences in Exercise 54 into that language.

Exercise 56. For each of the following, (1) identify the main connective and (2) construct a parse tree.

1. $((p \to q) \to p) \to p$
2. $(p \to q) \to (p \to p)$
3. $p \to ((q \to p) \to p)$
4. $(p \to (q \to p)) \to p$
5. $(\neg p \vee q) \wedge ((r \to q) \vee (p \vee \neg q))$
6. $\neg((p \vee q) \wedge (r \to q)) \vee (p \vee \neg q)$
7. $\neg((p \vee q) \wedge ((r \to q) \vee (p \vee \neg q)))$

8. $(\neg p \vee (q \wedge r)) \rightarrow (q \vee (p \vee \neg q))$
9. $((p \vee \neg q) \wedge (\neg p \vee q)) \rightarrow (p \vee q)$
10. $(p \vee (\neg q \wedge (\neg p \vee q))) \rightarrow (p \vee q)$
11. $(p \vee (\neg q \wedge \neg p)) \vee (q \rightarrow (p \vee q))$
12. $(p \vee (\neg q \wedge (\neg p \vee q))) \rightarrow (p \vee q)$
13. $p \vee (\neg q \wedge (\neg p \vee ((q \rightarrow p) \vee q)))$

Exercise 57. For each of the wffs in Exercise 56, produce an interpretation (an assignment of truth values to the atomic letters) which makes the wff true, if possible.

Exercise 58. For each of the wffs in Exercise 56, let $V(p) = T$, $V(q) = F$, and $V(r) = T$. Which of the wffs are true on these assignments?

Exercise 59. For each of the wffs in Exercise 56, determine, using truthtables, whether the sentence is a tautology or a contradiction or neither.

Exercise 60. Prove Lemma 7.4.30.

Exercise 61. Find a formula χ such that the only connectives in χ are $\wedge$ and $\neg$ and $\chi \leftrightarrow (\varphi \vee \psi)$ is a tautology.

Exercise 62. Determine whether the following sentences are tautologies:

1. $((p \wedge q) \rightarrow r) \rightarrow ((p \rightarrow r) \vee (q \rightarrow r))$
2. $((p \rightarrow q) \rightarrow p) \rightarrow p$
3. $(p \rightarrow (q \rightarrow p)) \rightarrow p$
4. $(p \rightarrow (q \rightarrow r)) \rightarrow (q \rightarrow (p \rightarrow r))$
5. $(p \rightarrow (p \wedge q)) \rightarrow (p \vee \neg q)$
6. $(\neg(p \wedge q) \wedge (r \vee q)) \rightarrow (r \vee \neg p)$
7. $(((\neg p \vee q) \vee p) \wedge p) \rightarrow \neg q$
8. $((\neg p \vee q) \rightarrow (\neg r \vee s)) \rightarrow ((\neg p \vee r) \rightarrow (\neg s \wedge q))$
9. $((p \rightarrow q) \rightarrow q) \rightarrow p$

10. $((p \to q) \to p) \to q$

11. $\neg p \to (p \to q)$

12. $((p \to q) \wedge (q \vee p)) \to q$

13. $(p \wedge (r \vee s)) \to ((\neg r \vee \neg p) \wedge (\neg s \vee \neg p))$

14. $(((p \wedge q) \to r) \to ((p \to r) \to s)) \to (q \to s)$

Exercise 63. Show that the formulas in Table 7.10 are all tautologies.

Exercise 64. Determine whether the following arguments are valid or invalid. If they are valid, give a proof of them.

1. $a \to (e \to \neg f), h \vee (\neg f \to m), a, \neg h$, therefore $e \to m$.

2. $(a \vee b) \to (c \wedge d), (x \vee \neg y) \to (\neg c \wedge \neg w), (x \vee z) \to (a \wedge e)$, therefore $\neg x$.

3. $n \vee \neg b, p \vee b, p \to q, (n \vee q) \to (r \wedge s), s \to (r \to d), b \to (d \to u)$, therefore u.

4. $(b \to \neg m) \to (d \to \neg s), b \to k, k \to \neg m, \neg s \to n$, therefore $d \to n$.

5. $\neg(j \vee k), b \to k, s \to b$, therefore $\neg s \wedge \neg j$.

6. $(b \to r) \to s, (p \to r) \to \neg s$, therefore $\neg r$.

7. $f \to g, \neg h \vee a, (g \vee a) \to j, \neg j$, therefore $\neg(f \vee h)$.

8. a, therefore $((p \to q) \to p) \to p$.

9. $p \wedge \neg p$, therefore a.

Exercise 65. (1) Translate these arguments into symbolic form; (2) identify the premises and the conclusion; (3) determine if the arguments are valid or not; (4) if they are not, (4a) give a valuation that makes the premises true and the conclusion false, and (4b) formulate (in English) a conclusion that *would* follow validly from the premises.

1. If Izo paints a new painting, then the melm will take it to the anarch and Trasken city will continue to be kept under illusion. Trasken city will only be released from the illusion if Izo agrees to help Thadia or the anarch is overthrown. But the anarch can only be overthrown if Izo agrees to help Thadia and paints a new painting. Therefore the melm must not be allowed to take Izo's painting to the anarch.

2. Sylene is a painkiller and safe to take if and only if it does not cause liver damage. But sylene does cause liver damage, therefore it is neither a painkiller nor safe to take.

3. If we build the tower tall enough, then we will be able to speak to the gods. But to build the tower tall enough, we need both brick-layers and sculptors. The brick-layers and the sculptors will work only if there are priestesses enough to bless them. Therefore, if we wish to speak to the gods, then we must have enough priestesses for the brick-layers and sculptors to build the tower tall enough.

4. Bial will sail across the sea only if Blár will too. Either Roia won't sail across the sea or the navigator will. Therefore, if both Bial and Roia go, then so too will the navigator and Blár.

Exercise 66. (1) Translate these arguments into symbolic form; (2) identify the premises and the conclusion; (3) determine if the arguments are valid or not; (4) if they are, prove the conclusion from the assumption of the premises; (5) if they are not, give a truth value assignment to the atomic letters that makes the premises true and the conclusion false.

1. If a superconducting particle collider is built, then the data yielded will benefit scientists of all nations and it deserves international funding. Either a superconducting particle collider will be built, or the ultimate nature of matter will remain hidden and the data yielded will benefit scientists of all nations. Therefore, the data yielded by a superconducting particle collider will benefit scientists of all nations.

2. Vitamin E is an antioxidant and a useless food supplement if and only if it does not reduce heart disease. But vitamin E neither reduces heart disease nor is it an antioxidant. Therefore, vitamin E is not a useless food supplement.

3. If astronauts attempt interplanetary space travel, then heavy shielding will be required to protect them from solar radiation. If massive amounts of either fuel or water are carried, then the spacecraft must be very large. Therefore, if heavy shielding is required to protect the astronauts from solar raditional only if massive amounts of fuel are carried, then if astronauts attempt interplanetary space travel, then the spacecraft must be very large.

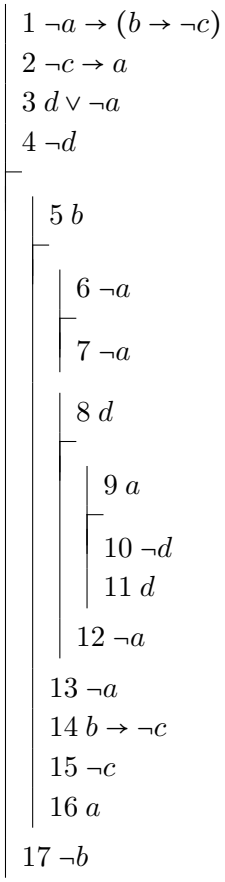
4. John will go to the party only if Harry will go too. Either Emily won't go to the party or Jamie will. Therefore, if both John and Emily go, then so too will Jamie and Harry.

Exercise 67. Add the necessary annotations to the following to show that $(b \to g) \wedge (f \to n), \neg(g \wedge n) \vdash \neg(b \wedge f)$.

$1\ (b \to g) \wedge (f \to n)$
$2\ \neg(g \wedge n)$
—
| $3\ b \wedge f$
| —
| $4\ b \to g$
| $5\ b$
| $6\ g$
| $7\ f \to n$
| $8\ f$
| $9\ n$
| $10\ g \wedge n$
| $11\ \neg(g \wedge n)$
$12\ \neg(b \wedge f)$

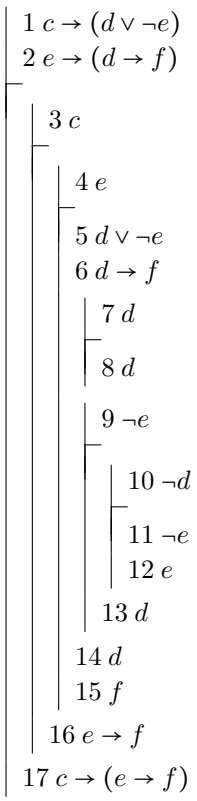
Exercise 68. Add the necessary annotations to the following to show that $\neg a \to (b \to$

$\neg c), \neg c \to a, d \vee \neg a, \neg d \vdash \neg b.$



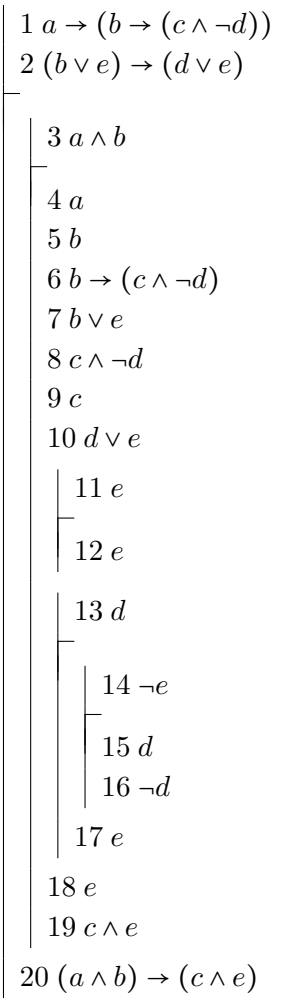
Exercise 69. Add the necessary annotations to the following to show that $c \to (d \vee \neg e), e \to (d \to$

$f) \vdash c \to (e \to f)$.



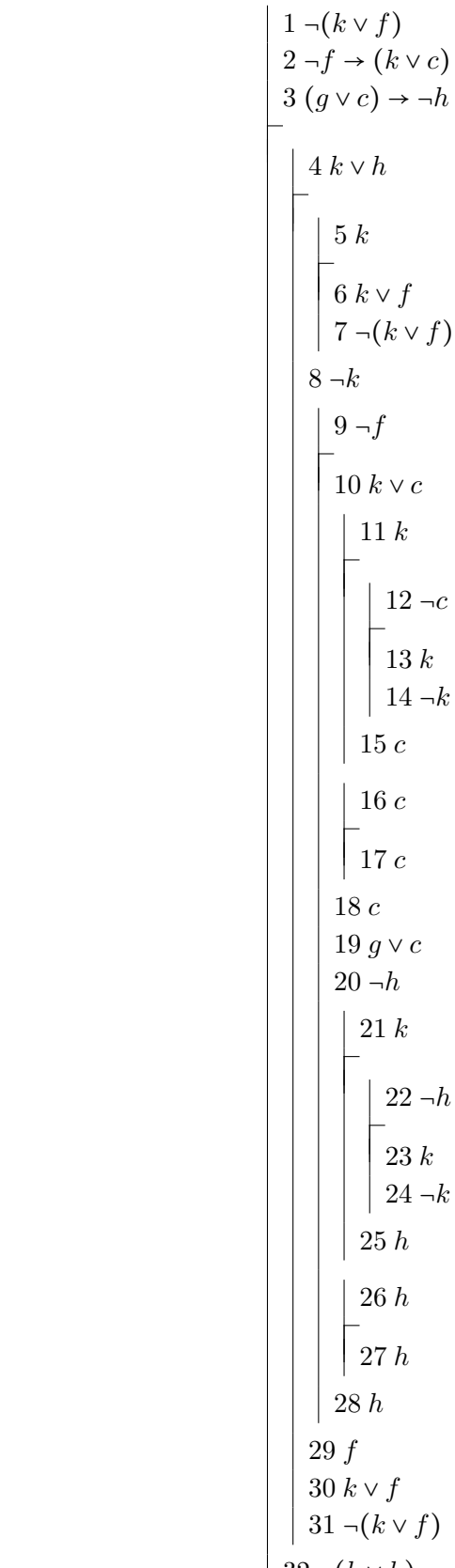
Exercise 70. Add the necessary annotations to the following to show that $a \to (b \to$

$(c \wedge \neg d)), (b \vee e) \to (d \vee e) \vdash (a \wedge b) \to (c \wedge e)$.



Exercise 71. Add the necessary annotations to the following to show that $\neg(k \vee f), \neg f \to$

$(k \vee c), (g \vee c) \to \neg h \vdash \neg(k \vee h)$.



Exercise 72. Give formal proofs of the following using the basic rules only:

1. $\varphi \to \psi, \neg\psi \vdash \neg\varphi$ (modus tollens)
2. $\varphi \vee \psi, \neg\psi \vdash \varphi$ (disjunctive syllogism)
3. $\varphi \wedge \psi \vdash \neg(\neg\varphi \vee \neg\psi)$ (DeMorgan's)
4. $\neg\neg\varphi \vdash \varphi$ (double negation)
5. $(\varphi \wedge \psi) \to \chi \vdash \varphi \to (\psi \to \chi)$ (exportation)

Exercise 73. Prove the following, using both basic and derived rules:

1. $m \to n, m \to o \vdash m \to (n \wedge o)$
2. $\neg s \to k, s \to (r \vee m) \vdash \neg r \to (\neg m \to k)$
3. $c \to (\neg l \to q), l \to \neg c, \neg q \vdash \neg c$
4. $(j \wedge r) \to h, (r \to h) \to m, \neg(p \vee \neg j) \vdash m \wedge \neg p$
5. $\neg(u \wedge w) \to x, u \to \neg u \vdash \neg(u \vee \neg x)$
6. $\neg b \to h, \neg d \to h, \neg(b \wedge d) \vdash h$
7. $(f \wedge h) \to n, f \vee s, h \vdash n \vee s$
8. $(l \vee p) \to u, (m \to u) \to k, p \vdash k$
9. $a \leftrightarrow w, \neg a \vee \neg w, r \to a \vdash \neg(w \vee r)$
10. $h \to (\neg e \to (c \to \neg d)), \neg d \to e, e \vee h, \neg e \vdash \neg c.$
11. $\neg s \to d, \neg s \vee (\neg d \to k), \neg d \vdash k$
12. $(r \to f) \to ((r \to \neg g) \to (s \to q)), (q \to f) \to (r \to q), \neg g \to f, q \to \neg g \vdash s \to f$
13. $k \to l, (m \to n) \wedge s, n \to t, k \vee m \vdash l \vee t$
14. $(n \to b) \wedge (o \to c), q \to (n \vee o), q \vdash b \vee c$
15. $(p \to r) \to (m \to p), (p \vee m) \to (p \to r), m \vee p, \vdash r \vee p$
16. $(u \wedge \neg\neg p) \to q, \neg o \to u, \neg p \to o, \neg o \wedge t \vdash q$

17. $(f \wedge m) \to (s \vee t), (\neg s \vee a) \to f, (\neg s \vee b) \to m, \neg s \wedge g \vdash t$

18. $\neg h \to (\neg t \to r), h \vee (e \to f), \neg t \vee e, \neg h \wedge d \vdash r \vee f$

19. $(m \vee n) \to (f \to g), d \to \neg c, \neg c \to b, m \wedge h, d \vee f \vdash b \vee g$

20. $(\neg m \to p) \wedge (\neg n \to q), \neg(m \wedge n) \vdash p \vee q$

21. $\neg(j \vee k), b \to k, s \to b \vdash \neg s \wedge \neg j$

22. $(j \vee f) \vee m, (j \vee m) \to \neg p, \neg f \vdash \neg(f \vee p)$

23. $(k \wedge p) \vee (k \wedge q), p \to \neg k \vdash q \vee t$

24. $(g \wedge h) \vee (m \wedge g), g \to (t \wedge a) \vdash a.$

25. $(\neg r \vee d) \to \neg(f \wedge g), (f \wedge r) \to s, f \wedge \neg s \vdash \neg(s \vee g)$

26. $(s \vee t) \to (s \to \neg t), (s \to \neg t) \to (t \to k), s \vee t \vdash \neg s \to k$

27. $l \to (\neg m \to (n \wedge o)), \neg n \wedge p \vdash l \to (m \wedge p)$

28. $t \to (h \wedge j), (h \vee n) \to t \vdash t \leftrightarrow h$

29. $(o \to c) \wedge (\neg s \to \neg d), (e \to d) \wedge (\neg e \to \neg c) \vdash o \to s$

30. $j \to (g \to l) \vdash g \to (j \to l)$

31. $(e \to a) \wedge (f \to a), e \vee g, f \vee \neg g \vdash a$

32. $(o \to r) \to s, (p \to r) \to \neg s \vdash \neg r$

33. $(s \vee t) \to (s \to \neg t), (s \to \neg t) \to (t \to k), s \vee t \vdash s \vee k$

34. $p \to a, q \to b \vdash (p \vee q) \to (a \vee b)$

Exercise 74. Prove the soundness of (1) $\wedge\mathsf{E}$, (2) $\vee\mathsf{I}$, (3) $\to\mathsf{E}$, and (4) $\neg\mathsf{I}$.

A.6 Exercises for Chapter 8

Exercise 75. In the following sentences, (1) Identify the constants; (2) Identify the 1-place predicates; and (2) Identify the binary relations.

1. Of the client who ordered the yellowtail roll and the customer who ordered the boston roll, one paid $13.50 and the other is Pam.

2. The person who ordered the yellowtail roll paid 2 dollars less than the person who ordered the volcano roll.

3. The customer who paid $13.50 isn't Orlando.

4. William paid 2 dollars more than the customer who ordered the tiger roll.

5. The client who ordered the futomaki roll paid $9.50.

6. The person who ordered the volcano roll paid 2 dollars less than the customer who ordered the boston roll.

7. The customer who paid $15.50 is either the client who ordered the dragon roll or the client who ordered the yellowtail roll.

Exercise 76. In the following sentences, (1) Identify the constants; (2) Identify the 1-place predicates; and (2) Identify the binary relations.

1. Leaves that are green turn to brown.

2. Cecilia, you're breaking my heart, you're shaking my confidence daily.

3. Not all that glitters is gold, only shoot stars break the mold.

4. Not all heroes wear capes.

5. Everybody loves my baby, but my baby don't love nobody but me.

6. Every breath you take, every move you take, every bond you break, every step you take, I'll be watching you.

7. Hold on to me, I'll hold on to you, the winter long I will always be with you. Hold on to me, I'll hold on to you, I will be the one who will always see you through.

8. She is not the only song I sing, though I sing them all for her.

9. All of my songs can only be composed of the greatest of pains. Every single verse can only be born of the greatest of wishes.

10. It's at night when they come, when I'm alone, and not with anyone.

11. We are the Baldrick's son, and Blackadders.

Exercise 77. Pick your own favorite song lyrics and identify the same three things as in Exercise 76.

Exercise 78. For all the sentences in Exercises 76 and 77, what are the logical connectives? What are the quantifiers?

(Bonus: Start trying to formalise the sentences in a first-order language.)

Exercise 79. Formalize these sentences in predicate logic; use the natural letter choice to represent the predicates and relations (e.g., 'musician' = M, 'trombonist' = T, etc.)

1. Only talented musicians perform in the symphony.

2. The only musicians available are trombonists.

3. Whenever James meets with Albert, Albert is late.

4. Taylor is guilty only if all the witnesses perjured themselves.

5. Balcony seats are never chosen unless all the orchestra seats are taken.

6. The physicists and astronomers at the symposium are listed in the program if they either chair a meeting or read a paper.

7. Either you do your homework or you will not pass the module.

8. Honda and Suzuki build motorcycles.

9. If your hook isn't baited, you won't catch a fish.

10. James will come to the party only if Alice does and Andrew doesn't.

11. London is larger than Durham, but Durham is a nicer place to live.

12. You will not pass the module unless you attend tutorials.

13. You won't catch a fish unless your hook is baited.

Exercise 80. Let $\mathcal{L}_q$ be a language consisting of the one-place predicates A 'is fat', B 'is blue', C 'is a car', F 'is a fish', H^1 'is a hat', L 'is little', P 'is a place', R 'is red', S 'is star', T 'is thin', U 'is funny', Y 'is yellow', and the binary relations H^2 'has', I 'in'. Translate the following into English:

1. $\exists x \exists y(Fx \wedge Fy \wedge x \neq y \wedge Bx \wedge Rx)$.
2. $\exists x \exists y(H^2xy \wedge Sy \wedge Ly)$.
3. $\exists x \exists y(H^2xy \wedge Sy \wedge Cy)$.
4. $\exists x(Fx \wedge Tx) \wedge \exists y(Fy \wedge Ay)$.
5. $\exists x(Fx \wedge Ax \wedge \exists y(Yy \wedge H^1y \wedge H^2xy)$.
6. $\forall x(Px \rightarrow \exists y(Fy \wedge Iyx))$.

Exercise 81. Let $\mathcal{L}_q$ be a language consisting of the following:

Unary predicates

B	'is a bear'
D	'is a day'
E	'is beautiful'
G	'is grass'
I	'is big'
L	'is long'
S	'is scared'
W	'is wavy'

Binary relations

C	'catches'
H	'hunts'
O	'goes over'
T	'goes through'
U	'goes under'

Translate the following into English:

1. $\exists x \exists y \exists z(Bz \wedge (Hxz \wedge Hyz))$
2. $\exists x \exists y \exists z(Bz \wedge Iz \wedge (Cxz \wedge Cyz))$
3. $\forall x(\exists y(By \wedge Hxy) \rightarrow \neg Sx)$
4. $\exists x((Gx \wedge Lx) \wedge Wx)$

5. $\exists x \exists y \forall z((Gz \wedge (Lz \wedge Wz)) \rightarrow \neg(Oxz \wedge Oyz))$

6. $\exists x \exists y \forall z((Gz \wedge (Lz \wedge Wz)) \rightarrow \neg(Uxz \wedge Uyz))$

7. $\exists x \exists y \forall z((Gz \wedge (Lz \wedge Wz)) \rightarrow (Txz \wedge Tyz))$

Exercise 82. Read the following two articles:

- Bertrand Russell, "On Denoting," *Mind*, n.s. 14, no. 56 (1905): 479–493, `https://www.jstor.org/stable/2248381`
- Delia Graff Fara, "Socratizing," *American Philosophical Quarterly* 48, no. 3 (2011): 229–238, `http://apq.press.illinois.edu/48/3/fara.html`

And answer these questions:

1. What is a denoting phrase, according to Russell?
2. What three cases of denoting phrases can we identify?
3. How does Russell say we should analyse the sentence "All men are mortal"?
4. How does Russell analyse the phrase "the father of Charles II"?
5. What is the general method that Russell gives for reducing any proposition with a denoting phrase into one which contains none?
6. How does Russell analyse the sentence "The present king of France is bald"?
7. Why does he argue that this sentence is false, rather than meaningless?
8. Briefly explain Russell's three puzzles on p. 485.
9. How does Russell's analysis of denoting phrases solve these three puzzles?
10. What is a primary occurrence of a denoting phrase? What is a secondary occurrence?
11. According to philosophical folklore, how does Quine propose to remove proper names from language?
12. How can we understand proper names to be disguised definite descriptions?
13. How does Quine solve the problem of negative existential claims?

14. What is the problem Graff identifies with defining 'Pegasus' as "winged horse captured by Bellerophon" on p. 231?

15. What alternative form of descriptivism does Quine adopt in response to this problem?

16. What does Quine gain by shifting from "is Pegasus" to "Pegasizes"?

17. How does the proper name 'John' differ from the proper name 'Pegasus'?

18. What, according to Graff, is Quine's account of proper names in *Word and Object*?

Exercise 83. For the following wffs:

1. List the terms occurring in the wff.

2. List the atomic wffs occurring in the wff.

3. For each wff, identify the bound and free variables.

1	$\forall x \exists y Rxyz \wedge \exists y Rxyy$
2	$\forall x(\exists y Rxyz \wedge \exists y Rxyy)$
3	$\exists y(x = y) \rightarrow \exists z(x = z)$
4	$\forall x \exists x Rxx$
5	$\exists y Qxy \wedge \forall z \exists y Qxy$
6	$\exists y Qxy \wedge \forall z \exists y Qzy$
7	$\forall x(F(x) = \mathsf{c}) \rightarrow \exists y(F(x) = y)$
8	$\forall x \forall y \exists z(Rxyz \rightarrow \exists y Rxyy \wedge (F(x) = z))$
9	$\forall x(Px \rightarrow Qxy) \rightarrow (\exists y Py \rightarrow \exists z Qyz)$
10	$\exists x Qxy \rightarrow (Px \rightarrow \neg \exists z Qxz)$

For each of 1–5, give a model and a variable assignment which makes the formula true (if possible). (The models need not be the same for each formula.)

For each of 6–10, give a model and a variable assignment which makes the formula false (if possible). (The models need not be the same for each formula.)

Exercise 84. Render the following sentences in (more or less idiomatic) English:

1. $\exists x(Tx \wedge (Ox \vee Rx))$.

2. $\forall x((Ax \wedge Ox) \rightarrow Gx)$.

3. $\exists x(f(f(\mathsf{h})) = x \wedge Px)$.

4. $\exists x \forall y (y \neq f(x))$.

5. $\forall x ((Lx \vee Ax) \rightarrow \exists y (f(y) = x))$.

6. $\exists x Zx\mathsf{g}$.

Exercise 85. Identify the terms in the sentences in Exercise 84. Which ones are atomic?

Exercise 86. Give parse trees of the formulas in Exercise 84 (cf. Definition 8.2.10).

Exercise 87. For each of the following:

- List the terms occurring in the formula.
- List the atomic formulas occuring in the formula.
- Identify which variables are bound and which are free.
- Identify whether the formula is a sentence or not.

1. $\forall x \forall y ((Gx \wedge Fzy) \rightarrow E(f(\mathsf{g}), x, z))$
2. $\forall x \exists y E(xyz) \wedge \exists y Z(x_1, y)$
3. $\forall x \exists y (Cx \wedge Sy) \rightarrow (S\mathsf{h} \vee Gf(z))$
4. $\exists y (x = y) \rightarrow \exists z (x = z)$
5. $\forall x \exists x (x < x)$
6. $\exists y (Px \wedge \forall z \exists y F(xy))$
7. $\exists y Ly \wedge \forall z \exists y E(zyx)$
8. $\forall x (f(x) = \mathsf{g}) \rightarrow \exists y (f(z) = y))$
9. $\forall x \forall y \exists z (Z(xyz) \rightarrow \exists y Az) \wedge (r(x) = z))$
10. $\forall x (P(x) \rightarrow Q(xy)) \rightarrow (\exists y Py \rightarrow \exists z (y < z))$
11. $\exists x Bx \rightarrow (P(x) \rightarrow \neg \exists z (Bx \vee O\mathsf{l}))$

Exercise 88. For each wff φ in Exercise 87, calculate $\varphi(\mathsf{a}/x)$.

Figure A.1: Domain

Exercise 89. Let $\mathcal{L}$ be a language containing the constants c and d; the unary predicate symbols B, C, O, R, and S; the binary relation symbol Z; the ternary relation symbol E; and the unary function symbols f and g.

Let $\mathfrak{M} = \langle D, I\rangle$ be a model where D is given in Figure A.1, and where I is as follows:

- $I(B) = \{\star, \star, \boxtimes\}$
- $I(\mathsf{c}) = \circledast$
- $I(C) = \{\boxtimes, \boxtimes, \circledast\}$
- $I(\mathsf{d}) = \diamond$
- $I(E) = \{\langle \star, \star, \boxtimes\rangle, \langle \diamondsuit, \diamondsuit, \square\rangle\}$
- $I(f)$ is the function that returns the object immediately above (if there is one), and is identity otherwise.
- $I(O) = \{\diamondsuit, \diamond, \square, \star, \star, \diamondsuit\}$
- $I(r)$ is the function that returns the object immediately below (if there is one), and is identity otherwise.
- $I(R) = \{\circledast, \diamondsuit, \square, \diamondsuit\}$
- $I(S) = \{\star, \star, \star, \circledast\}$
- $I(Z) = \{\langle \diamondsuit, \star\rangle, \langle \star, \star\rangle, \langle \square, \star\rangle, \langle \star, \star\rangle, \langle \boxtimes, \star\rangle, \langle \star, \star\rangle, \langle \circledast, \star\rangle, \langle \boxtimes, \star\rangle, \langle \diamond, \star\rangle, \langle \diamondsuit, \star\rangle\}$

Let v be a variable assignment such that $v(x) = \star$, $v(y) = \circledast$, $v(z) = \boxtimes$, and for any other variable w, $v(w) = \star$.

For each of the following, determine if the claim is true or false.

1. $\mathfrak{M}, v \vDash \exists x Bx$
2. $\mathfrak{M}, v \vDash \neg\exists x(r(\mathsf{d}) = x)$
3. $\mathfrak{M}, v \vDash \exists x(Ox \wedge \exists y \exists z(Exyz))$
4. $\mathfrak{M}, v \vDash \exists x \forall y(xZy)$
5. $\mathfrak{M}, v \vDash \exists x \forall y(yZx)$

6. $\mathfrak{M}, v \vDash \exists x(r(\mathsf{c}) = x \wedge (r(x) = y \wedge Cy))$
7. $\mathfrak{M}, v \vDash \forall x(Bx \to Ox)$
8. $\mathfrak{M}, v \vDash \exists x \exists y \exists z(Exyz \wedge \neg Ezyx)$
9. $\mathfrak{M}, v \vDash \exists x(Rx \wedge (Sx \vee Bx))$
10. $\mathfrak{M}, v \vDash \forall x((Bx \wedge Rx) \to Cx)$
11. $\mathfrak{M}, v \vDash \exists x(f(f(\mathsf{d})) = x \wedge Ox)$
12. $\mathfrak{M}, v \vDash \forall x(f(x) = \mathsf{c} \to \exists y(f(x) = y))$
13. $\mathfrak{M}, v \vDash \exists x Bx \to (C(x) \to \neg \exists z(Zzx \vee S\mathsf{d}))$

If the sentence is true, explain why (following the definitions of truth). If the sentence is false, indicate which parts of the model that makes it false.

Exercise 90. Define a formula φ and a substitution instance of that formula which is *not* allowed.

Exercise 91. Prove the following (note that this proof only requires propositional proof rules): $G\mathsf{a} \to Gs(\mathsf{m}), (G\mathsf{n} \vee G\mathsf{o}) \to \neg G\mathsf{m}, Gs(\mathsf{m}) \to G\mathsf{m} \vdash \neg G\mathsf{a} \vee (\neg G\mathsf{n} \wedge \neg G\mathsf{o})$.

Exercise 92. Add the missing annotations to the following proof.

1	$\forall x(Ax \to Bx)$
2	$A\mathsf{m} \vee A\mathsf{n}$
	⊢
	3 $A\mathsf{m}$
	⊢
	4 $A\mathsf{m} \to B\mathsf{m}$
	5 $B\mathsf{m}$
	6 $\exists x Bx$
	7 $A\mathsf{n}$
	⊢
	8 $A\mathsf{n} \to B\mathsf{n}$
	9 $B\mathsf{n}$
	10 $\exists x Bx$
11	$\exists x Bx$

Exercise 93. Prove the following.

1. $\forall x(Ax \to Bx), \forall x(Bx \to Cx) \vdash \forall x(Ax \to Cx)$.
2. $\forall x(Ax \to Bx), \neg B\mathsf{m} \vdash \exists x \neg Ax$.
3. $\forall x(Ax \to (Bx \vee Cx)), A\mathsf{g} \wedge \neg B\mathsf{g} \vdash C\mathsf{g}$.
4. $\forall x((Ax \vee Bx) \to Cx), \exists y(Ay \wedge Dy) \vdash EyCy$.
5. $\forall x(Ax \to Bx), A\mathsf{m} \vee A\mathsf{n} \vdash B\mathsf{m} \vee B\mathsf{n}$.
6. $\exists x Ax \to \forall x(Bx \to Cx), \exists x Dx \to \exists x \neg Cx, \exists x(Ax \wedge Dx) \vdash \exists x \neg Bx$.
7. $\forall x(Ax \to Bx), \forall x(Ax \to Cx) \vdash \forall x(Ax \to (Bx \wedge Cx))$.
8. $\exists x Ax \to \exists x(Bx \wedge Cx), \exists x(Cx \vee Dx) \to \forall x Ex \vdash \forall x(Ax \to Ex)$.
9. $\exists x Ax \to \exists x(Bx \wedge Cx), \neg \exists x Cx \vdash \forall x \neg Ax$.
10. $\forall x(Ax \to Cx), \exists x Cx \to \exists x(Bx \wedge Dx) \vdash \exists x Ax \to \exists x Bx$.
11. $\exists x Ax \to \forall x Bx, A\mathsf{n} \to \neg B\mathsf{n}, \vdash \neg A\mathsf{n}$.
12. $\forall x((Ax \vee Bx) \to Cx), \forall x((Cx \vee Dx) \to \neg Ax) \vdash \forall x((Ax \wedge Dx) \to \neg Bx)$.
13. $\exists(Ax \vee Bx), \exists x Ax \to \forall x(Cx \to Bx), \exists x Cx \vdash \exists x Bx$.

Exercise 94. Prove the following:

1. $\forall x(Px \to Qx) \to \exists x(Rx \wedge Sx), \forall x(Px \to Sx) \wedge \forall x(Sx \to Qx) \vdash \exists x Sx$
2. $\forall x(Ax \to Bx), A\mathsf{m} \wedge A\mathsf{n} \vdash B\mathsf{m} \wedge B\mathsf{n}$.
3. $\forall x(Bx \to Cx), \exists x(Ax \wedge Bx) \vdash \exists x(Ax \wedge Cx)$.
4. $\forall x(Jx \to (Kx \wedge Lx)), \exists y \neg Ky \vdash \neg \forall z Jz$.
5. $\forall x(Ax \to (Bx \vee Cx)), \exists x(Ax \wedge \neg Cx) \vdash \exists x Bx$.
6. $\forall x(Bx \vee Ax), \forall x(Bx \to Ax) \vdash \forall x Ax$.
7. $\neg \exists x(Ax \wedge \neg Bx), \neg \exists x(Ax \wedge \neg Cx) \vdash \forall x(Ax \to Cx)$.
8. $\forall x((Ax \wedge Bx) \to Cx), \neg \forall x(Ax \to Cx) \vdash \exists x \neg Bx$.

9. $\exists x \neg Ax \to \neg \exists x Bx, \neg \forall x Ax \to \exists x Bx, \forall x(\neg Ax \vee Cx) \vdash \forall x Cx$.

10. $\neg \exists x(Ax \vee Bx), \exists x Cx \to \exists x Ax, \exists x Dx \to \exists x Bx \vdash \neg \exists x(Cx \vee Dx)$.

Exercise 95. Prove the following, using both basic and derived rules:

1. $\exists x \forall y(Axy \to Bxy), \forall x \exists y \neg Bxy \vdash \neg \forall x \forall y Axy$

2. $\forall x \exists y Axy \to \forall x \exists y Bxy, \exists x \forall y \neg Bxy \vdash \exists x \forall y \neg Axy$.

3. $\forall x \exists y Axy \vee \forall x \forall y Bxy, \forall x \exists y(Cx \to \neg Bxy) \vdash \forall x \exists y(Cx \to Axy)$.

4. $\forall x \exists y(Ax \wedge By) \vdash \exists y \forall x(Ax \wedge By)$.

5. $\forall x \exists y(Ax \vee By) \vdash \exists y \forall x(Ax \vee By)$.

6. $\forall x(Ax \to \forall y(By \to Cxy)), A\mathsf{m} \wedge B\mathsf{n} \vdash C\mathsf{mn}$.

A.7 Exercises for Chapter 10

Exercise 96. Prove that any model for the language $\{\mathtt{0}, \mathtt{S}\}$ satisfying PA1 and PA2 must be infinite.

Exercise 97. Prove that Add as defined in Example 10.3.12 is primitive recursive.

Exercise 98. Calculate $h(0,1)$, $h(1,1)$, $h(2,1)$, etc., given the definitions of h, f, and g in Example 10.3.11. What function is this?

Exercise 99. Prove the following:

$$\vdash_{\mathrm{PA}} \underline{2} \times \underline{3} = \underline{6}$$

Exercise 100. For all natural numbers n, m, we have:

$$\vdash_{\mathrm{PA}} \mathtt{Add}(\underline{m}, \underline{n}) = \underline{m+n}$$

Exercise 101. For all natural numbers n, m, we have:

$$\vdash_{\mathrm{PA}} \mathtt{Mult}(\underline{m}, \underline{n}) = \underline{m \times n}$$

Exercise 102. [Transitivity of $\leq$]

$$\vdash_{\mathrm{PA}} \mathtt{x} \leq \mathtt{y} \wedge \mathtt{y} \leq \mathtt{z} \to \mathtt{x} \leq \mathtt{z}$$

Exercise 103.

$$\vdash_{\mathrm{PA}} \neg(x < 0)$$

Exercise 104. Prove the commutativity of addition, that is,

$$\vdash_{\mathrm{PA}} \forall xy(x + y = y + x).$$

Do this by first proving two lemmas:

Lemma A.7.1. $\vdash_{\mathrm{PA}} \forall x(0 + x = x)$.

Lemma A.7.2. $\vdash_{\mathrm{PA}} \forall x \forall y(Sx + y = S(x + y))$.

Exercise 105. Prove the following:

$$\vdash_{\mathrm{PA}} x < Sy \leftrightarrow x < y \vee x = y$$

Exercise 106. Show that

$$\vdash_{\mathrm{PA}} (p|x \wedge r \times s = p) \rightarrow r|x$$

Exercise 107. Show that

$$\vdash_{\mathrm{PA}} (r \times s = p \wedge r \geq p > \underline{1}) \rightarrow s = \underline{1}$$

Exercise 108. Define a formula $\varphi(x)$ with one free variable x such that

$$\mathbb{N} \vDash \varphi(n) \text{ iff } n \text{ is even}$$

Exercise 109. $\forall x \leq 5 \exists y \leq 3(x + y < 9)$ is a closed bounded formula. By Lemma 4.9.6, it is equivalent to a quantifier free formula. Produce the formula it is equivalent to, and prove the equivalence in PA.

Exercise 110. By Lemma 4.9.4, $\forall x \leq (x_1 + Sx_2) \exists w \exists y \leq w(x + y = w)$ is equivalent to a Σ_1^0 formula. Produce the formula it is equivalent to.

A.8 Exercises for Chapter 11

Exercise 111. Prove the remaining half of Lemma 11.3.7.

Exercise 112. Prove Corollary 11.3.8.

Exercise 113. Prove Corollary 11.3.9.

Exercise 114. Determine whether the following are valid:

1. $\Box(\varphi \wedge \psi) \rightarrow (\Box\varphi \wedge \Box\psi)$.
2. $\Diamond \Box \varphi \rightarrow \Box \Diamond \varphi$.
3. $\Box\varphi \rightarrow \Box \Box \varphi$.
4. $\Box \Diamond \varphi \rightarrow \varphi$.
5. $\Box\varphi \rightarrow \varphi$.

Exercise 115. For the formulas in Exercise 114, determine whether they are satisfiable.

Exercise 116. Prove the following in K:

- $\Box(p \rightarrow q) \rightarrow (\Box(q \rightarrow r) \rightarrow \Box(p \rightarrow r))$
- $\Box(p \rightarrow q) \rightarrow (\Diamond p \rightarrow \Diamond q)$
- $\Box(p \rightarrow q) \rightarrow (\Diamond(p \wedge r) \rightarrow \Diamond(q \wedge r))$
- $\Diamond(p \rightarrow p) \rightarrow (\Box q \rightarrow \Diamond q)$
- $(\Box p \wedge \Diamond q) \rightarrow \Diamond(p \wedge q)$

Exercise 117. Let α be a constant wff. Then if $\tau(\alpha)$ is PL-valid, $\vdash_{\mathrm{D}} \alpha$; otherwise, $\vdash_{\mathrm{D}} \neg\alpha$.

Exercise 118. Let D be the system obtained by adding to K the axiom:

$$(\mathsf{D}): \qquad \Box p \rightarrow \Diamond p$$

Prove that D is sound and complete with respect to the class of frames satisfying the condition

$$\forall x \exists y (xRy)$$

Exercise 119. Prove the following in S4:

- $\Box(p \rightarrow q) \rightarrow \Box(\Box p \rightarrow \Box q)$.
- $(\Box p \vee \Box q) \rightarrow \Box(\Box p \vee \Box q)$.
- $\Box(\Box(p \rightarrow q) \rightarrow r) \rightarrow \Box(\Box(p \rightarrow q) \rightarrow \Box r)$.

Exercise 120. Prove the following in S5:

- $\Box(\Box p \vee q) \rightarrow (\Box p \vee \Box q)$
- $(\Diamond p \wedge \Diamond q) \rightarrow \Diamond(\Diamond p \wedge q)$
- $\Box(\Box p \leftrightarrow \Box\Box p)$
- $\Box(\Box p \leftrightarrow \Diamond\Box p)$
- $\Box(\Diamond p \leftrightarrow \Diamond\Diamond p)$
- $\Box(\Diamond p \leftrightarrow \Box\Diamond p)$

Exercise 121. Prove that the following frame conditions identify the same class of frames:

- Reflexive and Euclidean.
- Reflexive, symmetric, and transitive.
- Serial, symmetric, and transitive.
- Serial, symmetric, and Euclidean.

Exercise 122. Given the result in Exercise 121, give four distinct axiomatisations of S5.

Exercise 123. Reduce the following to OMCNF, and determine whether the resulting formula "passes the test". If it does, give a proof of it in S5. If it does not, given an S5-countermodel.

1. $\Box(\Box(\Box p \rightarrow \Box q) \rightarrow \Box(p \rightarrow q))$.
2. $\Diamond\Box(p \rightarrow \Box\Diamond p)$.
3. $\Box(\Box(p \rightarrow \Box p) \rightarrow \Box p) \rightarrow (\Diamond\Box p \rightarrow \Box p)$.
4. $\Diamond(p \wedge q) \rightarrow \Box(\Diamond(\Box p \rightarrow \Diamond q) \rightarrow \Box q)$.
5. $\Box(p \rightarrow q) \rightarrow \Box(\Diamond(p \wedge \neg\Box p) \rightarrow \Diamond(q \wedge \Box(p \rightarrow \Box p)))$.

Exercise 124. Let L1 be the system obtained by adding to K the axiom:

$$(\mathsf{L1}): \qquad (p \wedge \Box p) \rightarrow \Box\Box p$$

Prove that the frame of $\mathfrak{M}^{\mathrm{L1}}$ satisfies the condition

$$\forall xyz((xRy \wedge yRz \wedge x \neq z) \rightarrow xRz)$$

Exercise 125. Let K4Lem$_0$ be the system obtained by adding to K4 the axiom

$$(\mathsf{Lem}_0): \qquad \Box((p \wedge \Box p) \to q) \vee \Box((q \wedge \Box q) \to p)$$

Prove that the frame of $\mathfrak{M}^{\mathrm{K4Lem}_0}$ satisfies the condition

$$\forall wxy((wRx \wedge wRy \wedge x \neq y) \to (xRy \vee yRx))$$

You may assume that the frame is transitive.

A.9 Exercises for Chapter 12

Exercise 126. Let Believe, Fear, and Doubt be operators corresponding to the natural language expressions "I believe", "I fear", and "I doubt", *i.e.*, the meaning of $\mathsf{Fear}(p)$ is "I fear that p", *etc.*.

Give examples (in terms of a little story that provides the necessary background information required for evaluating the natural language expressions) for the **in**validity of the following rules (2 points each):

- If $\mathsf{Believe}(p \vee q)$, then $\mathsf{Believe}(p) \vee \mathsf{Believe}(q)$.
- If $\mathsf{Fear}(p \wedge q)$, then $\mathsf{Fear}(p) \wedge \mathsf{Fear}(q)$.
- If $\mathsf{Doubt}(p \wedge q)$, then $\mathsf{Doubt}(p) \wedge \mathsf{Doubt}(q)$.
- If Fear$(\neg p)$, then not Fear(p).

Exercise 127. Read Alan Code, "Aristotle's response to Quine's objections to modal logic", **Journal of Philosophical Logic** 5 (1976), p. 159–186, and answer the following questions.

1. Code pseudo-deduces the false statement (3) "Ford resigned last August" from the true statements (1) and (2). If Ford didn't resign, who did and when did he resign exactly?
2. Paraphrase Smullyan's solution to the problem of "The president resigned last August" in one sentence.
3. Does Code believe that Aristotle had something like Smullyan's solution in mind? (Give a brief argument.)
4. Explain briefly (at most 100 words) what Code means when he says "Ford is not a spatio-temporal worm but rather ... a hydra".

A.10 Exercises for Chapter 13

Exercise 128. Prove the correspondence between the following frame conditions and temporal formulas. That is, prove (1) that if the frame meets the condition, the temporal formula is valid (true at every world in every model on that frame), and (2) that if the frame fails to meet the condition, it is possible to define a model on that frame with a world in the model where the temporal formula is false.

Backwards linearity	$\forall xyz((x < z \wedge y < z) \rightarrow (x < y \vee y < x \vee x = y))$	$FPp \rightarrow (Pp \vee p \vee Fp)$
Forwards density	$\forall xy(x < y \rightarrow \exists z(x < z \wedge z < y))$	$Fp \rightarrow FFp$
Forwards discreteness	$\forall xy(x < y \rightarrow \exists z(x < z \wedge \neg\exists t(x < t < z))$	$G(G(p \rightarrow Gp) \rightarrow p) \rightarrow (FGp \rightarrow p)$

Exercise 129. A **naumachic model** is a quadruple $\langle M, U, \leq, S\rangle$ where M and U are finite sets, $\leq$ is a binary relation between M and U (*i.e.*, $\leq \subseteq M \times U$) and S is a function from U to $\{\mathsf{seabattle}, \mathsf{no\text{-}seabattle}\}$.

We call the elements of M **tomorrows**, the elements of U **DATs** (for "**D**ay **A**fter **T**omorrow"), if $m \leq u$, we say that "u is a possible future of m", and if $S(u) = \mathsf{seabattle}$ we say that "there is a sea battle at u" (similarly, if $S(u) = \mathsf{no\text{-}seabattle}$ we say that "there is no sea battle at u".

Given a naumachic model $\mathbf{N} = \langle M, U, \leq, S\rangle$, we say

- $\mathbf{N} \vDash$ "There will be a sea battle the day after tomorrow" if for all $m \in M$ and all u such that $m \leq u$, $S(u) = \mathsf{seabattle}$.
- $\mathbf{N} \vDash$ "There will be a sea battle the day after tomorrow" if for all $m \in M$ and all u such that $m \leq u$, $S(u) = \mathsf{no\text{-}seabattle}$.
- $\mathbf{N} \vDash$ "Tomorrow it will be determined whether there is a sea battle the day after tomorrow" if for all $m \in M$ the following holds: all u such that $m \leq u$ have the same value of $S(u)$.

We consider the following four naumachic models (t represents "today", not represented in the formal model, the m_i are the tomorrows, the u_i are the day-after-tomorrows, the arrows indicate the $\leq$ relation, and u_i:$\mathsf{seabattle}$ means $S(u_i) = \mathsf{seabattle}$).

Are the following statements true or false?

1. In $\mathbf{N}_0$, there will be a sea battle the day after tomorrow.
2. In $\mathbf{N}_1$, there will be a sea battle the day after tomorrow.
3. In $\mathbf{N}_2$, there will be a sea battle the day after tomorrow.
4. In $\mathbf{N}_0$, it will be determined tomorrow whether there is a sea battle the day after tomorrow.
5. In $\mathbf{N}_1$, it will be determined tomorrow whether there is a sea battle the day after tomorrow.
6. In $\mathbf{N}_2$, it will be determined tomorrow whether there is a sea battle the day after tomorrow.

Appendix B

Answers to Selected Exercises

Proof of Exercise 96. Let $\mathfrak{M}$ be a model for the language $\{\mathtt{S}, O\}$, and suppose that $\mathfrak{M}$ is finite, containing elements $\{m_0, \ldots, m_n\}$. Fix an element m_j of $\mathfrak{M}$ as the interpretation of O. The interpretation of $\mathtt{S}$ is a function mapping $\{m_0, \ldots, m_n\}$ to $\{m_0, \ldots, m_n\}$. By PA1 we know that every m_i, $0 \leq i \leq n$, $m_j \neq \mathtt{S}m_i$. This means that we have n possible values of $\mathtt{S}m_i$, since it cannot be m_j. Then for some m_k, $i \neq k$, $\mathtt{S}m_i = \mathtt{S}m_k$. But by PA2, it follows that $m_i = m_k$, which contradicts our assumption. □

Proof of Exercise 97. Straightforward from the proof of Exercise 98. □

Proof of Exercise 100. We rewrite the proof in [Goldstern and Judah, 1998] to make it clear when the object-language functions are used and when the meta-language ones. We fix m and use induction on n.

Basis case:

$$\vdash_{\mathrm{PA}} \mathtt{Add}(\underline{m}, \underline{0}) = \underline{m + 0}$$

Note that $\underline{0} = 0$, and this follows from Axiom PA3. Note also that $\underline{k+1} = \mathtt{S}(\underline{k})$. Then:

1. $\vdash_{\mathrm{PA}} \mathtt{Add}(\underline{m}, \underline{k}) = \underline{m + k}$ (inductive hypothesis)
2. $\vdash_{\mathrm{PA}} \mathtt{Add}(\underline{m}, \underline{k+1}) = \mathtt{S}(\mathtt{Add}(\underline{m} + \underline{k}))$, by Axiom PA4.
3. $\vdash_{\mathrm{PA}} \mathtt{S}(\mathtt{Add}(\underline{m}, \underline{k})) = \mathtt{S}(\underline{m + k})$ (by 1, and equality axioms).
4. $\vdash_{\mathrm{PA}} \mathtt{Add}(\underline{m}, \underline{k+1}) = \mathtt{S}(\underline{m + k})$ (by equality).

□

Proof of Exercise 101. We fix m and use induction on n.

Basis case:

$$\vdash_{\text{PA}} \mathtt{Mult}(\underline{m}, \underline{0}) = \underline{m \times 0}$$

This is obvious from Axiom PA5 once you note that $\underline{0} = 0$. Note that $\underline{k+1} = \mathtt{S}(\underline{k})$. Then:

1. $\vdash_{\text{PA}} \mathtt{Mult}(\underline{m}, \underline{k}) = \underline{m \times k}$ (inductive assumption)
2. $\vdash_{\text{PA}} \mathtt{Mult}(\underline{m}, \underline{k+1}) = \mathtt{Add}(\mathtt{Mult}(\underline{m}, \underline{k}), \underline{m})$ (by PA6).
3. $\vdash_{\text{PA}} \mathtt{Mult}(\underline{m}, \underline{k+1}) = \mathtt{Add}(\underline{m \times k}, \underline{m})$ (by application of inductive hypothesis).
4. $\vdash_{\text{PA}} \mathtt{Mult}(\underline{m}, \underline{k+1}) = \underline{(m \times k) + m}$ (by Example 4.2.2 of the book).
5. $\vdash_{\text{PA}} \mathtt{Mult}(\underline{m}, \mathtt{S}(\underline{k})) = \underline{(m \times k) + m}$.
6. $\vdash_{\text{PA}} \underline{(m \times k) + m} = \underline{m \times \mathtt{S}(k)}$ (by the reflexivity of identity).
7. $\vdash_{\text{PA}} \mathtt{Mult}(\underline{m}, \mathtt{S}(\underline{k})) = \underline{m \times \mathtt{S}(k)}$.

□

Proof of Exercise 102. Fix y and z and prove for x. By Example 4.3.1, we know that every number is either 0 or a successor.

- Case 1: $x = 0$. Need to show:

$$\vdash_{\text{PA}} 0 \leq y \wedge y \leq z \rightarrow 0 \leq z$$

Proof.

1	$0 \leq y, y \leq z$	$\vdash_{\text{PA}}$	$\exists r(0 + r = y) \wedge \exists s(y + s = z)$	(def. of $\leq$)
2	$0 \leq y, y \leq z$	$\vdash_{\text{PA}}$	$\exists r, s(0 + r + s = z)$	(1)
3	$0 \leq y, y \leq z$	$\vdash_{\text{PA}}$	$\exists t(0 + t = z)$	(2)
4	$0 \leq y, y \leq z$	$\vdash_{\text{PA}}$	$0 \leq z$	(def. of $\leq$)
5		$\vdash_{\text{PA}}$	$0 \leq y \wedge y \leq z \rightarrow 0 \leq z$	(Ded.Thm.)

- Case 2: $x = \mathtt{S}t$. Need to show:

$$\mathtt{Sx} \leq \mathtt{y} \wedge \mathtt{y} \leq \mathtt{z} \rightarrow \mathtt{Sx} \leq \mathtt{z}$$

Proof.

1	$\mathtt{S}x \leq y, y \leq z$	$\vdash_{\mathrm{PA}}$	$\exists t(\mathtt{S}x + t = y) \wedge \exists r(y + r = z)$	(def. of $\leq$)
2	$\mathtt{S}x \leq y, y \leq z$	$\vdash_{\mathrm{PA}}$	$\exists t, s(\mathtt{S}x + t + s = z)$	(1)
3	$\mathtt{S}x \leq y, y \leq z$	$\vdash_{\mathrm{PA}}$	$\exists r(\mathtt{S}x + r = z)$	(2)
4	$\mathtt{S}x \leq y, y \leq z$	$\vdash_{\mathrm{PA}}$	$\mathtt{S}x \leq z$	(def. of $\leq$)
5		$\vdash_{\mathrm{PA}}$	$\mathtt{S}x \leq y \wedge y \leq z \rightarrow \mathtt{S}x \leq z$	(Ded.Thm.)

□

Proof of Exercise 103. Note that $\vdash_{\mathrm{PA}} \neg(x < 0)$ is short-hand for $\vdash_{\mathrm{PA}} \neg(x \leq 0 \wedge x \neq 0)$, which is in turn short-hand for $\vdash_{\mathrm{PA}} \neg(\exists r(x + r = 0) \wedge x \neq 0)$.

1	$x = 0$	$\vdash_{\mathrm{PA}}$	$\neg(x \neq 0)$	(double negation)
2	$x = 0$	$\vdash_{\mathrm{PA}}$	$\neg(\varphi \wedge x \neq 0)$	(for any φ)
3	$x = 0$	$\vdash_{\mathrm{PA}}$	$\neg(\exists r(x + r = 0) \wedge x \neq 0)$	
4	$x = \mathtt{S}y$	$\vdash_{\mathrm{PA}}$	$x \neq 0$	(PA2)
5	$x = \mathtt{S}y$	$\vdash_{\mathrm{PA}}$	$\forall r((\mathtt{S}y + r) = \mathtt{S}(y + r))$	(PA4)
6	$x = \mathtt{S}y$	$\vdash_{\mathrm{PA}}$	$\forall r(\mathtt{S}(y + r) \neq 0)$	(PA1)
7	$x = \mathtt{S}y$	$\vdash_{\mathrm{PA}}$	$\forall r(\mathtt{S}y + r \neq 0)$	(identity)
8	$x = \mathtt{S}y$	$\vdash_{\mathrm{PA}}$	$\neg\exists r(\mathtt{S}y + r = 0)$	(quantifier interchange)
9	$x = \mathtt{S}y$	$\vdash_{\mathrm{PA}}$	$\neg\exists r(\mathtt{S}y + r = 0 \wedge x \neq 0)$	
10	$x = 0 \vee x = \mathtt{S}y$	$\vdash_{\mathrm{PA}}$	$\neg\exists r(\mathtt{S}y + r = 0 \wedge x \neq 0)$	(3, 9)
11		$\vdash_{\mathrm{PA}}$	$x = 0 \vee x = \mathtt{S}y \rightarrow$ $\neg\exists r(\mathtt{S}y + r = 0 \wedge x \neq 0)$	(Ded.Thm)
12		$\vdash_{\mathrm{PA}}$	$\neg\exists r(\mathtt{S}y + r = 0 \wedge x \neq 0)$	(4.3.1, (11), MP)
13		$\vdash_{\mathrm{PA}}$	$\neg(x \leq 0 \wedge x \neq 0)$	
14		$\vdash_{\mathrm{PA}}$	$\neg(x < 0)$	

□

Proof of Exercise 98. $h(0, 1) = f(1) = P^1_1(x) = 1$.

$$\begin{aligned} h(1, 1) &= h(S0, 1) \\ &= \mathtt{S}(h(0, 1) \\ &= \mathtt{S}1 = 2 \end{aligned}$$

(Final case left as exercise). This is the addition function. □

Proof of Exercise 117. First, recall what a constant formula and what the PL-transform τ is. A constant formula is a formula whose truth-value is guaranteed to be 1 or 0 (i.e., a tautology or a contradiction) irrespective of the values assigned to the propositions. Constant formulas are built from $\top$ 'constant true proposition' and $\bot$ 'constant false proposition', with the usual truth-functional and modal connectives [Hughes and Cresswell, 1996, p. 47]. The PL-transform $\tau(\alpha)$ of a formula α is α with all the modal operators removed [Hughes and Cresswell, 1996, p. 65].

To prove this lemma, we must prove (i) that it holds for $\bot$; (ii) that if it holds for α it also holds for $\neg\alpha$; (iii) if it holds for α it also holds for $L\alpha$; and (iv) if it holds for α and β then it holds for $\alpha \vee \beta$. (i) and (iii) are proven in the book [Hughes and Cresswell, 1996, p. 66]; (iv) is provable by purely PL principles. We prove (ii) here:

Assume the lemma holds for some constant α. $\tau(\alpha)$ is either PL-valid or PL-unsatisfiable. If it is valid, then $\tau(\neg\alpha)$ is unsatisfiable; if it is unsatisfiable, then $\tau(\neg\alpha)$ is valid. If $\tau(\alpha)$ is valid, then by assumption, $\vdash_{\mathrm{D}} \alpha$. It follows that $\vdash_{\mathrm{D}} \neg\neg\alpha$. Thus if $\tau(\neg\alpha)$ is unsatisfiable, $\vdash_{\mathrm{D}} \neg\neg\alpha$. If $\tau(\alpha)$ is unsatisfiable, then by assumption $\vdash_{\mathrm{D}} \neg\alpha$. But then, if $\tau(\neg\alpha)$ is valid, $\vdash_{\mathrm{D}} \neg\alpha$, as required. □

Proof of Exercise 118.

- Soundness: To prove that $Lp \to \Diamond p$ is sound with respect to the class of serial frames we simply need to show that D cannot be falsified on any serial model. Suppose it can. Then there is some serial model $\mathfrak{M} = \langle W, R, V\rangle$ and world $w \in W$ such that $w \nvDash Lp \to \Diamond p$. That is, (1) $w \vDash Lp$ and (2) $w \vDash \neg Mp$. By LMI, (2) is equivalent to $w \vDash L\neg p$ (3). Since $\mathfrak{M}$ is serial, there is an x such that wRx. By (1), $x \vDash p$. By (3), $x \vDash \neg p$. This is a contradiction, so we reject our assumption that D can be falsified in a serial model.

- Completeness: To prove that D is complete with respect to the class of serial models, we have to show that the frame of the canonical model for D, $\mathfrak{M}^{\mathrm{D}}$, is a frame for D, that is, it is a member of the class of frames satisfying this condition:

  $$\forall x \exists y (xRy)$$

  From D, it is possible to prove theorem $\mathsf{D1}$: $M(p \to p)$. Since it is a theorem, it is in every world of the canonical model. By the truth-and-membership lemma, $V(M(p \to p), w) = 1$ for every world in the model. So, for any arbitrary w, there must be a w' such that wRw' and $V(p \to p, w') = 1$. As a result, for every world in the canonical model there is a world that can it can see, so the frame of the canonical model is serial, as required.

□

Proof of Exercise 123.

1.

2.

3.

4.

5. Reduce $\Box(p \to q) \to \Box(\Diamond(p \wedge \neg \Box p) \to \Diamond(q \wedge \Box(p \to \Box p)))$ to a first-degree formula:

   Step 1 $\neg \Box (\neg p \vee q) \vee \Box(\neg \Diamond (p \wedge \neg \Box p) \vee \Diamond(q \wedge \Box(\neg p \vee \Box p)))$
   Step 2 $\Diamond(p \wedge \neg q) \vee \Box(\Box(\neg p \vee \Box p) \vee \Diamond(q \wedge \Box(\neg p \vee \Box p)))$
   Step 4 $\Diamond(p \wedge \neg q) \vee (\Box \Box (\neg p \vee \Box p) \vee \Diamond(q \wedge \Box(\neg p \vee \Box p)))$
   Step 3 $\Diamond(p \wedge \neg q) \vee (\Box(\neg p \vee \Box p) \vee \Diamond(q \wedge \Box(\neg p \vee \Box p)))$
   Step 4 $\Diamond(p \wedge \neg q) \vee ((\Box \neg p \vee \Box p) \vee \Diamond(q \wedge \Box(\neg p \vee \Box p)))$
   Step 4 $\Diamond(p \wedge \neg q) \vee ((\Box \neg p \vee \Box p) \vee (\Diamond q \wedge \Box(\neg p \vee \Box p)))$
   Step 4 $\Diamond(p \wedge \neg q) \vee ((\Box \neg p \vee \Box p) \vee (\Diamond q \wedge (\Box \neg p \vee \Box p)))$

   Convert to OMCNF:

   1 $\Diamond(p \wedge \neg q) \vee (\Box \neg p \vee \Box p) \vee (\Diamond q \wedge (\Box \neg p \vee \Box p))$
   2 $\Diamond(p \wedge \neg q) \vee ((\Box \neg p \vee \Box p) \vee \Diamond q) \wedge ((\Box \neg p \vee \Box p) \vee (\Box \neg p \vee \Box p))$
   3 $(\Diamond(p \wedge \neg q) \vee (\Box \neg p \vee \Box p) \vee \Diamond q)) \wedge (\Diamond(p \wedge \neg q) \vee ((\Box \neg p \vee \Box p) \vee (\Box \neg p \vee \Box p))$
   4 $(\Box \neg p \vee \Box p \vee \Diamond(q \vee (p \wedge \neg q))) \wedge (\Box \neg p \vee \Box p \vee \Box \neg p \vee \Box p) \vee \Diamond(p \wedge \neg q))$

   There are two conjuncts to perform the test on: (1) $(\Box \neg p \vee \Box p \vee \Diamond(q \vee (p \wedge \neg q)))$ and (2) $(\Box \neg p \vee \Box p \vee \Box \neg p \vee \Box p) \vee \Diamond(p \wedge \neg q))$. If we perform the test on (1), the two disjuncts we must test are (a) $\neg p \vee (q \vee (p \wedge \neg q))$ and (b) $p \vee (q \vee (p \wedge \neg q))$. If we distribute (1a), we get (1a′) $\neg p \vee (q \wedge p) \vee (q \wedge \neg q)$. The final disjunct of (1a′) is a contradiction, so we can ignore it; the result, $\neg p \vee (p \wedge q)$ is not a tautology. Applying similar techniques to (1b), we get $p \vee (p \wedge q)$, which is also not a tautology. So, conjunct (1) fails the test. Because the first conjunct fails the test, it doesn't matter what the second ones does.

□

Appendix C

Errata in [Goldstern and Judah, 1998]

In this chapter is the errata for [Goldstern and Judah, 1998] that the advanced logic classes at Durham University have collected between 2015 and 2019.

C.1 4.3

- page 197, proof line 10: for *4.3.2* read *4.3.1*
- page 198, example 4.3.6, proof of (1): For *PA1* read *PA3*.
- page 198, example 4.3.6, proof of (2): For *PA2* read *PA4*.
- page 200, definition of 'prime': Add a final closing parenthesis to the definition.
- page 201, (A2): For $\neg z|y$, read $z|y$ (twice).

C.2 4.4

- page 204, example 4.4.2: In all cases, for 56,700, read 567,000.
- page 208, first line: For *1.3.12* read *1.3.14*.

C.3 4.5

- page 209, definition 4.5.1, line beginning with ∘: For *out* read *our*.

- page 210, last line: For $\mathtt{t} = \underline{2}^{\mathtt{o}} \cdot \underline{3}^{\mathtt{v}} \cdot \underline{5}^{\mathtt{f}_1}$ read $\mathtt{f} = \underline{2}^{\mathtt{o}} \cdot \underline{3}^{\mathtt{v}} \cdot \underline{5}^{\mathtt{f}_1}$.
- page 211, two lines before Definition 4.5.11: For *existential quantification* read *universal quantification*.
- page 211, last line of first par. after Definition 4.5.11: For $\varphi_k = \exists \mathtt{x}_n \varphi_j$ read $\varphi_k = \forall \mathtt{x}_n \varphi_j$.
- page 213, Definition 4.5.19: For *GroupII(*$\mathtt{p}$*)* read *GroupIV(*$\mathtt{p}$*)*.
- page 213, Definition 4.5.19: Replace the definition of *GroupIV(*$\mathtt{p}$*)* with the following:

$$[\exists \mathtt{f}_1 \exists \mathtt{f}_2 \exists \mathtt{x} \mathit{implies}(\mathtt{f}_1, \mathtt{f}_2, \mathtt{p}) \wedge \mathit{var}(\mathtt{x}) \wedge \mathtt{f}_2 = \underline{2}^{\ulcorner \forall \urcorner} \cdot \underline{3}^{\mathtt{x}} \cdot \underline{5}^{\mathtt{f}_1} \wedge \mathit{formula}(\mathtt{f}_1) \wedge \mathit{formula}(\mathtt{f}_2) \wedge \mathit{notfree}(\mathtt{x}, \mathtt{f}_1)]$$

C.4 4.6

- page 214, Definition 4.6.1: Add to the end another conjunction:

$$\forall \mathtt{k}(\mathit{entry}(\mathtt{c}, \mathtt{k}, \mathtt{t}) \rightarrow \mathit{entry}(\mathtt{c}', \mathtt{k}, \mathtt{t}'))$$

- page 215, Definition 4.6.3: Correct $\mathit{atomicfm}(\mathtt{c}, \mathtt{k}, \mathtt{e})$ to $\mathit{atomicfm}(\mathtt{c}, \mathtt{k})$
- page 215, Definition 4.6.3: For the case where $\mathtt{x} \neq \mathtt{v}$, replace the consequent with:

$$\mathit{fobtained}(\mathtt{c}', \mathtt{k}, \mathtt{l}, \underline{\ulcorner \forall \urcorner}, \mathtt{x}, \mathtt{m})$$

- page 215, Definition 4.6.3: Add to the end another conjunction:

$$\forall \mathtt{k}(\mathit{entry}(\mathtt{c}, \mathtt{k}, \mathtt{f}) \rightarrow \mathit{entry}(\mathtt{c}', \mathtt{k}, \mathtt{f}')$$

C.5 4.7

- page 216, Definition 4.7.1: For *If φ is a formula* read *If φ is a formula with one free variable* $\mathtt{x}_0$.

C.6 4.8

- page 219, second paragraph: For *We found in section 3.7* read *We found in section 4.7*

C.7 4.9

- page 221, first line of section: For *section 3.1* read *section 4.1.*
- page 223, in (e), define $\varphi := \forall x \leq \tau(\exists y \psi)$.
- page 225, in the example of B: cases (1), (2), and (3) should all have 2 instead of n.
- page 225, line (3) of proof of Lemma 4.9.8: For φ' read $\ulcorner\varphi\urcorner$, twice.
- page 226, in final line of Definition 4.9.11: For *(i.e., the complement of A* read *(i.e., the complement of A).*
- page 227, proof of (2) in Example 4.9.14: In both displayed formulas, add a final closing parenthesis. In the second, for the terminal period read a comma.

C.8 4.10

- page 230, proof of (5) in Fact 4.10.1: For k^{k^k} read 2^{k^k}, throughout; for m_n read m_k.
- page 230, Lemma 4.10.2(c): For *nextprime*$(\mathtt{p}, \mathtt{q})$ read *nextprime*(x, y).
- page 231, statement of 4.10.2(C): For $f(k, s)$ read $f(s, k)$.
- page 231, proof of 4.10.2.(e): For `s` read `x`, throughout.
- page 231, proof of 4.10.2.(g): For `c` read `s`, throughout.
- page 231, proof of 4.10.2.(B): For $p^s | e \wedge p^{s+1} \nmid e$ read $p^e | s \wedge p^{e+1} \nmid s$.
- page 231, proof of 4.10.3.(d): The bound should be `c` rather than `t`.

Index

Π^0_1 formula, 213
Σ^0_1 formula, 213

Adam of Balsham, 19
Albert the Great, 21
Albertus Magnus, *see* Albert the Great
Aquinas, Thomas, 21
Aristotle, 14–16, 18, 21, 25, 57, 76, 78, 82, 90, 91, 97, 111, 112, 353, 355, 357, 373, 375
Aristotle's theses, 442
axiomatic basis, 272

Bacon, Roger, 2, 22, 23, 25, 34
Boethius, 16, 25
Boethius's theses, 442
Brouwer, L.E.J., 303

Celarent, 89, 91, 94, 99–101, 104, 105
Champeaux, William of, 19
Colebrooke, H.T., 70, 71
Consistency
  Of a predicate logic wff, 178
  Of a set of categorical propositions, 86
  Of a set of first-order axioms, 180
  Of a set of modal axioms, 272
  Of a set of predicate logic wffs, 178

Dharmakīrti, 66
Dignāga, 63, 65, 70
Duns Scotus, Johannes, 23, 57

Ferio, 89, 91, 95, 99, 100, 102, 104, 105

Gōngsūn Lóng, 67, 68, 71
Ghent, Henry of, 27
Grosseteste, Robert, 22

Henry of Ghent, *see* Ghent, Henry of
Henry, King of England, 22

insolubles, *see also* paradox

James of Venice, 19

Kilwardby, Robert, 23

Law of Excluded Middle, 74, 75, 116, 117, 153, 173, 434
logic
  connexive, 441, 442

Master Argument, 20, 372, 374–376, 379–381
model
  canonical, 328, 334
mood, 83, 88–91
  perfect, 91

Normal Form
Conjunctive, 124
Disjunctive, 124
Modal Conjunctive, 313
Ordered Modal Conjunctive, 313

Peter Abelard, 19
PL-transform, 287, 308, 309
Porphyry, 18
pre-order, 266, 436
Principle of Non-Contradiction, 74, 75, 116, 117, 153, 173, 273, 434
proof
modal, 272

Quine, W.V.O., 400, 405–407, 478, 479, 488

seriality, 265, 268, 269, 292, 293, 338, 352, 365, 380
set
definition of, 8
empty, 8
finite, 8
Sherwood, William, 24
Sherwood, William of, 22, 99
Square of Opposition, 16
syllogism
negative, *see also* Celarent, Ferio

Thomas Aquinas, *see* Aquinas, Thomas

Validity
Of a predicat logic wff, 178
Of a predicate logic argument, 178
Of a propositional argument, 122
Of a propositional wff, 121
Of a Stoic argument, 111
Of the four basic moods, 100
Vasubandhu, 63–66, 70

wff, *see* well-formed formula
William of Champeaux, *see* Champeaux, William of
William of Sherwood, *see* Sherwood, William of

Bibliography

[Adams, 1998] Adams, E. W. (1998). *A Primer of Probability Logic*. CSLI Publications.

[Anacker, 1984] Anacker, S. (1984). *Seven Works of Vasubandhu, the Buddhist Psychological Doctor*. Motilal Banarsidass Publications.

[Aristotle,] Aristotle. *Prior Analytics*.

[Asher and McCready, 2007] Asher, N. and McCready, E. (2007). Were, would, might and a compositional account of counterfactuals. *Journal of Semantics*, 24:93–129.

[Bacon, 2009] Bacon, R. (2009). *The Art and Science of Logic*. Pontifical Institute of Medieval Studies. Trans. by Thomas S. Maloney.

[Beonio-Brocchieri Fumagelli, 1969] Beonio-Brocchieri Fumagelli, M. T. (1969). *The Logic of Abelard*. Synthese Historical Library. Dordrecht: Reidel.

[Bezhanishvili and de Jongh, 2005] Bezhanishvili, N. and de Jongh, D. (2005). Intuitionistic logic. Technical report, European Summer School in Logic, Language, and Information.

[Biard, 2011] Biard, J. (2011). Albert of Saxony. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring edition. `http://plato.stanford.edu/archives/spr2011/entries/albert-saxony/`.

[Blackburn and Jørgensen, 2012] Blackburn, P. and Jørgensen, K. F. (2012). Indexical hybrid tense logic. In Bolander, T., Braüner, T., Ghilardi, S., and Moss, L., editors, *Advances in Modal Logic*, pages 144–160. College Publications.

[Boolos and Clark, 1993] Boolos, G. and Clark, P. (1993). Basic Law V. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 67:213–233+235–249.

[Bowen, 1895] Bowen, F. (1895). *A Treatise on Logic, or, The Laws of Pure Thought; Comprising Both the Aristotelic and Hamiltonian Analyses of Logic Forms, and Some Chapters of Applied Logic*. Boston: Allyn and Bacon.

[Braakhuis, 1998] Braakhuis, H. (1998). Obligations in early thirteenth century Paris: the *Obligationes* of Nicholas of Paris (?). *Vivarium*, 36(2):152–233.

[Braakhuis, 1981] Braakhuis, H. A. G. (1981). English tracts on syncategorematic terms from Robert Bacon to Walter Burley. In Braakhuis, H. A. G., Kneepkens, C. H., and de Rijk, L. M., editors, *English Logic and Semantics From the End of the Twelfth Century to the Time of Ockham and Burley*, pages 131–165. Ingenium Publishers.

[Bradwardine,] Bradwardine, T. *Insolubilia*. Stephen Read, ed.

[Brinkley et al., 1995] Brinkley, R., Spade, ed., P. V., and Wilson, ed., G. A. (1995). *Richard Brinkley's Obligationes: A Late Fourteenth Century Treatise on the Logic of Disputation*. Beiträge zur Geschichte der Philosophie und Theologie des Mittelalters. Aschendorff.

[Brogaard and Salerno, 2013] Brogaard, B. and Salerno, J. (2013). Fitch's paradox of knowability. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter edition. `http://plato.stanford.edu/archives/win2013/entries/fitch-paradox/`.

[Bull, 1964] Bull, R. A. (1964). The axiomatization of prior's modal calculus q. *Notre Dame Journal of Formal Logic*, 5(3):211–214.

[Buridan and Klima, trans., 2001] Buridan, J. and Klima, trans., G. (2001). *Summulae de Dialectica*. Yale University Press.

[Burley, 1963] Burley, W. (1963). Tractatus de obligationibus. In Green, R., editor, *An Introduction to the Logical Treatise 'De Obligationibus' with Critical Texts of William of Sherwood (?) and Walter Burley*. Louvain.

[Burley, 1988a] Burley, W. (1988a). Obligations (selections). In Kretzmann, N. and Stump, E., editors, *Cambridge Translations of Medieval Philosophical Texts*, volume 1, pages 369–412. Cambridge University Press.

[Burley, 1988b] Burley, W. (1988b). Obligations (selections). In Kretzmann, N. and Stump, E., editors, *Cambridge Translations of Medieval Philosophical Texts*, volume 1: Logic and the Philosophy of Language, pages 369–412. Cambridge University Press.

[Chagrov and Zakharyaschev, 1997] Chagrov, A. and Zakharyaschev, M. (1997). *Modal logic*. Clarendon Press.

[Chittaro and Montanari, 2000] Chittaro, L. and Montanari, A. (2000). Temporal representation and reasoning in artificial intelligence: Issues and approaches. *Annals of Mathematics and Artificial Intelligence*, 28(1–4):47–106.

[Copeland, 2008] Copeland, B. J. (2008). Arthur Prior. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall edition. `http://plato.stanford.edu/archives/fall2008/entries/prior`.

[Corish, 1976] Corish, D. (1976). Aristotle's attempted derivation of temporal order from that of movement and space. *Phronesis*, 21(3):241–251.

[Cross, 1993] Cross, C. B. (1993). From worlds to probabilities: A probabilistic semantics for modal logic. *Journal of Philosophical Logic*, 22:169–192.

[Davidson, 1967] Davidson, D. (1967). The logical form of action sentences. In Rescher, N., editor, *The Logic of Decision and Action*, pages 81–95. University of Pittsburgh Press.

[de Rijk, 1974] de Rijk, L. M. (1974). Some thirteenth century tracts on the game of obligation. *Vivarium*, 12(2):94–123.

[de Rijk, 1975] de Rijk, L. M. (1975). Some thirteenth century tracts on the game of obligation II. *Vivarium*, 13(1):22–54.

[de Rijk, 1976] de Rijk, L. M. (1976). Some thirteenth century tracts on the game of obligation III. *Vivarium*, 14(1):26–49.

[Demey et al., 2019] Demey, L., Kooi, B., and Sack, J. (2019). Logic and probability. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*. Summer edition. `https://plato.stanford.edu/archives/sum2019/entries/logic-probability/`.

[Dod, 1982] Dod, B. G. (1982). Aristoteles Latinus. In Kretzmann, N., Kenny, A., and Pinborg, J., editors, *Cambridge History of Later Medieval Philosophy*, pages 45–79. Cambridge University Press.

[Ebbinghaus et al., 1994] Ebbinghaus, H.-D., Flum, J., and Thomas, W. (1994). *Mathematical Logic*. Springer, 2nd edition.

[Findlay, 1941] Findlay, J. N. (1941). Time: A treatment of some puzzles. *Australasian Journal of Psychology and Philosophy*, 19:216–235.

[Fine, 1978] Fine, K. (1978). Model theory for modal logic. part i—the *de re/de dicto* distinction. *Journal of Philosophical Logic*, 7:125–156.

[Fisk, 1965] Fisk, M. (1965). The logic of either-or. *Notre Dame Journal of Formal Logic*, VI(1):39–50.

[Fitch, 1963] Fitch, F. B. (1963). A logical analysis of some value concepts. *Journal of Symbolic Logic*, 28(2):135–142.

[Fitting and Mendelsohn, 1998] Fitting, M. and Mendelsohn, R. L. (1998). *First-Order Modal Logic*, volume 277 of *Synthese Historical Library*. Kluwer Academic Publishers.

[Foster, 2008] Foster, C. (2008). The problem of the perfect agent: investigations into determinism. Master's thesis, Universiteit van Amsterdam. ILLC Publications MoL-2008-01, `http://www.illc.uva.nl/Publications/ResearchReports/MoL-2008-01.text.pdf`.

[Freeman and Sellars, 1971] Freeman, E. and Sellars, W., editors (1971). *Basic Issues in the Philosophy of Time*. Open Court.

[Gabbay, 1976] Gabbay, D. M. (1976). *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*. D. Reidel.

[Galton, 1987] Galton, A., editor (1987). *Temporal Logics and Their Applications*. Harcourt, Brace, Jovanovich.

[Ganeri, 2001] Ganeri, J., editor (2001). *Indian Logic: A Reader*. Curzon Press.

[Garson, 2006] Garson, J. W. (2006). *Modal Logic For Philosophers*. Cambridge University Press.

[Gillon, 2016] Gillon, B. (2016). Logic in classical Indian philosophy. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall edition. `https://plato.stanford.edu/archives/fall2016/entries/logic-india/`.

[Gödel, 1933] Gödel, K. (1933). Eine interpretation des intuitionistischen aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums*, 4:39–40.

[Gold, 2018] Gold, J. C. (2018). Vasubandhu. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer edition. `https://plato.stanford.edu/archives/sum2018/entries/vasubandhu/`.

[Goldberg et al., 1974] Goldberg, H., Leblanc, H., and Weaver, G. (1974). A strong completeness theorem for 3-valued logic. *Notre Dame Journal of Formal Logic*, 15(2).

[Goldblatt, 1992] Goldblatt, R. (1992). *Logics of Time and Computation*. Number 7 in CLSI Lecture Notes. Center for the Study of Language and Information, 2nd edition.

[Goldrei, 1996] Goldrei, D. C. (1996). *Classical Set Theory: For Guided Independent Study*. Chapman & Hall/CRC.

[Goldstern and Judah, 1998] Goldstern, M. and Judah, H. (1998). *The Incompleteness Phenomenon*. A K Peters, Ltd.

[Goranko et al., 2004] Goranko, V., Montanari, A., and Sciavicco, G. (2004). A road map of interval temporal logics and duration calculi. *Journal of Applied Non-Classical Logics*, 14(1):11–56.

[Grabmann, 1909] Grabmann, M. (1909). *Die Geschichte der scholastischen Methode: Nach d. gedruckten u. ungedr. Quellen dargest.* Schwabe.

[Hájek, 2009] Hájek, A. (2009). Two interpretations of two Stoic conditionals. *History of Philosophy and Logical Analysis*, 12(1):206–221.

[Hájek, 2010] Hájek, P. (2010). On fuzzy modal logics $\mathrm{S5}(\mathcal{C})^{\star}$. *Fuzzy Sets and Systems*, 161:2389–2396.

[Harel et al., 2000] Harel, D., Kozen, D., and Tiuryn, J. (2000). *Dynamic Logic*. MIT Press.

[Harel et al., 2002] Harel, D., Kozen, D., and Tiuryn, J. (2002). Dynamic logic. In Gabbay, D. and Guenther, F., editors, *Handbook of Philosophical Logic*, volume 4. Kluwer, 2 edition.

[Heytesbury, 1494] Heytesbury, W. (1494). *De Sensu Composito et Diviso*. Venice: Bonetus Locatellus.

[Heytesbury, 1988] Heytesbury, W. (1988). The compounded and divided senses. In Kretzmann, N. and Stump, E., editors, *Cambridge Translations of Medieval Philosophical Texts*, volume 1: Logic and the Philosophy of Language, pages 415–434. Cambridge University Press.

[Hintikka, 1962] Hintikka, J. (1962). *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press.

[Hintikka, 1973] Hintikka, J. (1973). *Time and Necessity: Studies in Aristotle's Theory of Modality*. Clarendon Press.

[Hughes and Cresswell, 1996] Hughes, G. E. and Cresswell, M. J. (1996). *A New Introduction to Modal Logic*. Routledge.

[Indraccolo, 2017] Indraccolo, L. (2017). The "white horse is not horse" debate. *Philosophy Compass*, 12:1–11.

[Japaridze and de Jongh, 1998] Japaridze, G. and de Jongh, D. (1998). The logic of provability. In *Handbook of Proof Theory*, pages 475–550. Elsevier.

[Jarrett, 1978] Jarrett, C. (1978). The logical structure of Spinoza's "Ethics", part I. *Synthese*, 37(1):15–65.

[Kilwardby, 1993] Kilwardby, R. (1993). *On Time and Imagination*. Oxford University Press. trans. by A. Broadie.

[Kleene, 1950] Kleene, S. C. (1950). *Introduction to Metamathematics*. D. Van Nostrand Company, Inc.

[Koslow, 1992] Koslow, A. (1992). *A Structuralist Theory of Logic*. Cambridge University Press.

[Koslow, 1999] Koslow, A. (1999). The implicational nature of logic: A structuralist account. In Varzi, A. C., editor, *European Review of Philosophy, The Nature of Logic, vol. 4*, pages 111–155. CSLI Publications.

[Kretzmann, 1982] Kretzmann, N. (1982). Syncategoremata, sophismata, exponibilia. In Kretzmann, N., Kenny, A., and Pinborg, J., editors, *Cambridge History of Later Medieval Philosophy*, pages 211–245. Cambridge University Press.

[Kripke, 1959] Kripke, S. A. (1959). A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24:1–14.

[Kröger, 1987] Kröger, F. (1987). *Temporal Logic of Programs*. Springer-Verlag.

[Kröger and Merz, 2008] Kröger, F. and Merz, S. (2008). *Temporal Logic and State Systems*. Springer.

[Lemmon, 1957] Lemmon, E. J. (1957). New foundations for Lewis modal systems. *Journal of Symbolic Logic*, 22(2):176–186.

[Lewis and Langford, 1932a] Lewis, C. I. and Langford, C. H. (1932a). *Symbolic Logic*. The Century Company, 2nd edition.

[Lewis and Langford, 1932b] Lewis, C. I. and Langford, C. H. (1932b). *Symbolic Logic*. New York, London: The Century Co.

[Lewis, 1979] Lewis, D. (1979). Possible worlds. In Loux, M., editor, *The Possible and the Actual*, pages 182–189. Cornell University Press.

[Łukasiewicz, 1920] Łukasiewicz, J. (1920). O logice trójwartościowej (on three-valued logic). *Ruch Filozoficzny*, 5:169–171.

[Manna and Pnueli, 1992] Manna, Z. and Pnueli, A. (1992). *The Temporal Logic of Reactive and Concurrent Systems*. Springer-Verlag.

[Mates, 1949] Mates, B. (1949). Diodorean implication. *Philosphical Review*, 58:234–244.

[Mates, 1961] Mates, B. (1961). *Stoic Logic*. University of California Press.

[McTaggart, 1908] McTaggart, J. E. (1908). The unreality of time. *Mind n.s.*, 17(68):457–474.

[Mironov, 2005] Mironov, A. M. (2005). Fuzzy modal logics. *Journal of Mathematical Sciences*, 128(6):3461–3483.

[Moody, 1953] Moody, E. A. (1953). *Truth and Consequence in Mediaeval Logic*. North-Holland Publishing Company.

[Morikawa, 1989] Morikawa, O. (1989). Some modal logics based on a three-valued logic. *Notre Dame Journal of Formal Logic*, 30(1):130–137.

[Murdoch and Thijssen, 2001] Murdoch, J. E. and Thijssen, J. M. M. H. (2001). John Buridan on infinity. In Thijssen, J. M. M. H. and Zupko, J., editors, *The Metaphysics and Natural Philosophy of John Buridan*, pages 127–149. Brill.

[of Holland and Bos, ed., 1985] of Holland, J. and Bos, ed., E. P. (1985). *Four Tracts on Logic*. Ingenium Publishers.

[of Inghen, 1489] of Inghen, M. (1489). *Obligationes*. Paris. Erroneously published under the name Pierre d'Ailly.

[of Pergola, 1961] of Pergola, P. (1961). *Logica and Tractatus de Sensu Composito et Diviso*. Franciscan Institute. Mary Anthony Brown, ed.

[of Saxony, 2010] of Saxony, A. (2010). Questiones circa logicam *(Twenty-Five Disputed Questions on Logic)*, volume 9 of *Dalls Medieval Texts and Translations*. Peeters. Introduction, Translation, and Notes by Michael J. Fitzgerald.

[of Sherwood, 1963] of Sherwood, W. (1963). Tractatus de obligationibus. In Green, R., editor, *An Introduction to the Logical Treatise 'De Obligationibus' with Critical Texts of William of Sherwood (?) and Walter Burley*. Louvain.

[of Sherwood, 1995] of Sherwood, W. (1995). *Introductiones in logicam / Einführung in die Logik*. Hamburg: Felix Meiner Verlag. ed. and trans. by H. Brands & Ch. Kann.

[of Spain, 1992] of Spain, P. (1992). *Syncategoreumata*. Brill. L.M. de Rijk, ed., Joke Spruyt, trans.

[of Venice, 1979] of Venice, P. (1979). *Pauli Veneti: Logica Magna, Prima Pars: Tractatus de Terminis*. Oxford University Press. Ed. and trans. by Norman Kretzmann.

[of Venice and Ashworth (ed. and trans.), 1988] of Venice, P. and Ashworth (ed. and trans.), E. J. (1988). *Logica Magna Part II, Fascicule 8: Tractatus de Obligationibus*. Oxford University Press.

[Øhrstrøm and Hasle, 2006a] Øhrstrøm, P. and Hasle, P. (2006a). A.N. Prior's logic. In Gabbay, D. M. and Woods, J., editors, *Handbook of the History of Logic*, volume 7, pages 400–446. Elsevier.

[Øhrstrøm and Hasle, 2006b] Øhrstrøm, P. and Hasle, P. (2006b). Modern temporal logic: The philosophical background. In Gabbay, D. M. and Woods, J., editors, *Handbook of the History of Logic*, volume 7, pages 447–498. Elsevier.

[Øhrstrøm and Hasle, 1995] Øhrstrøm, P. and Hasle, P. F. V. (1995). *Temporal Logic: From Ancient Ideas to Artificial Intelligence*, volume 57 of *Studies in Linguistics and Philosophy*. Kluwer Academic Publishers.

[Ostermann, 1988] Ostermann, P. (1988). Many-valued modal propositional calculi. *Zeitschrift für math. Logik und Grundlagen der Mathematik*, 34:343–354.

[Priest, 2001] Priest, G. (2001). *An Introduction to Non-Classical Logics*. Cambridge University Press.

[Prior, 1967] Prior, A. (1967). *Past, Present, and Future*. Clarendon Press.

[Prior, 1955] Prior, A. N. (1955). Diodoran modalities. *Philosophical Quarterly*, 5(20):205–213.

[Prior, 1957] Prior, A. N. (1957). *Time and Modality*. Clarendon Press.

[Prior, 1968] Prior, A. N. (1968). *Papers on Time and Tense*. Clarendon Press.

[Prior and Fine, 1977] Prior, A. N. and Fine, K. (1977). *Worlds, Times, and Selves*. Duckworth.

[Reichenbach, 1971] Reichenbach, H. (1971). *The Direction of Time*. University of CA Press.

[Ren and Liu, 2019] Ren, Y. and Liu, Y. (2019). A pro-realist account of Gongsun Long's "white horse dialogue". *Philosophy East and West*, 69(2):464–483.

[Rendsvig and Symons, 2019] Rendsvig, R. and Symons, J. (2019). Epistemic Logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edition.

[Rescher, 1966] Rescher, N. (1966). On the logic of chronological propositions. *Mind*, 75(297):75–96.

[Rescher and Urquhart, 1971] Rescher, N. and Urquhart, A. (1971). *Temporal logic*. Springer-Verlag.

[Richman, 2004] Richman, F. (2004). Equivalence of syllogisms. *Notre Dame Journal of Formal Logic*, 45:215–233.

[Schotch et al., 1978] Schotch, P. K., Jensen, J. B., Larsen, P. F., and MacLellan, E. J. (1978). A note on three-valued modal logic. *Notre Dame Journal of Formal Logic*, 19(1):63–68.

[Sedley, 1977] Sedley, D. (1977). Diodorus Cronus and Hellenstic philosophy. *Proceedings of the Cambridge Philological Society*, 203, n.s. 23:74–120.

[Segerberg, 1967] Segerberg, K. (1967). Some modal logics based on a three-valued logic. *Theoria*, 33:53–71.

[Segerberg, 1971] Segerberg, K. (1971). *An Essay in Classical Logic*. Uppsala: Filosofiska Studier.

[Sioson, 1964] Sioson, F. M. (1964). Further axiomatizations of the Łukasiewicz three-valued calculus. *Notre Dame Journal of Formal logic*, 5(1):62–70.

[Spade, 1978] Spade, P. V. (1978). Richard Lavenham's *Obligationes*: Edition and comments. *Rivista Critica di Storia della Filosofia*, 33:224–241.

[Spade, 1982] Spade, P. V. (1982). Three theories of *Obligationes*: Burley, Kilvington and Swyneshed on counterfactual reasoning. *History and Philosophy of Logic*, 3:1–32.

[Spade, 1996] Spade, P. V. (1996). *Thoughts, Words, and Things: An Introduction to Late Medieval Logic and Semantic Theory*. WWW: PVSpade.com, 1.0 edition. `http://pvspade.com/Logic/docs/thoughts.pdf`.

[Spade and Yrjönsuuri, 2013] Spade, P. V. and Yrjönsuuri, M. (2013). Medieval theories of *obligationes*. In Zalta, E. N., editor, *Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall edition. `http://plato.stanford.edu/archives/fall2013/entries/obligationes/`.

[Stalnaker, 1976] Stalnaker, R. (1976). Possible worlds. *Noûs*, 10(1):65–75.

[Strobino, 2009] Strobino, R. (2009). Concedere, Negare, Dubitare*: Peter of Mantua's Treatise on Obligations*. PhD thesis, Scuola Normale Superiore.

[Stump, 1985] Stump, E. (1985). The logic of disputation in Walter Burley's treatise on obligations. *Synthese*, 63:355–374.

[Tachikawa, 1971] Tachikawa, M. (1971). A sixth century manual of Indian logic. *Journal of Indian Philosophy*, 1:111–145.

[Thom, 2007] Thom, P. (2007). *Logic and Ontology in the Syllogistic of Robert Kilwardby*. Brill.

[Troelstra and Schwichtenberg, 2000] Troelstra, A. S. and Schwichtenberg, H. (2000). *Basic Proof Theory*. Cambridge University Press, 2nd edition.

[Trzęsicki, 1987] Trzęsicki, K. (1987). Is discreteness of time necessary for Diodorean master argument? *Bulletin of the Section of Logic*, 16(3):125–131.

[Tweedale, 1982] Tweedale, M. M. (1982). Abelard and the culmination of the old logic. In Kretzmann, N., Kenny, A., and Pinborg, J., editors, *Cambridge History of Later Medieval Philosophy*, pages 143–158. Cambridge University Press.

[Uckelman, 2009] Uckelman, S. L. (2009). Anselm's logic of agency. *Logical Analysis and History of Philosophy*, 12:248–268.

[Uckelman, 2011a] Uckelman, S. L. (2011a). Deceit and indefeasible knowledge: The case of *Dubitatio*. *Journal of Applied Non-Classical Logics*, 21(3/4):503–519.

[Uckelman, 2011b] Uckelman, S. L. (2011b). A dynamic epistemic logic approach to modeling *Obligationes*. In Grossi, D., Minica, S., Rodenhäuser, B., and Smets, S., editors, *LIRa Yearbook*, pages 147–172. Institute for Logic, Language, and Computation.

[Uckelman, 2013] Uckelman, S. L. (2013). Medieval *Disputationes de obligationibus* as formal dialogue systems. *Argumentation*, 27(2):143–166.

[Uckelman, 2014] Uckelman, S. L. (2014). Reasoning about obligations in *Obligationes*: A formal approach. In Goré, R., Kooi, B., and Kurucz, A., editors, *Advances in Modal Logic*, pages 553–568. College Publications.

[Uckelman et al., 2014] Uckelman, S. L., Alama, J., and Knoks, A. (2014). A curious dialogical logic and its composition problem. *Journal of Philosophical Logic*, 43(6):1065–1100.

[Uckelman et al., 2018] Uckelman, S. L., Maat, J., and Rybalko, K. (2018). The art of doubting in *Obligationes Parisienses*. In Kann, C., Löwe, B., Rode, C., and Uckelman, S. L., editors, *Modern Views of Medieval Logic*, volume 16 of *Recherches de Théologie et Philosophie Médiévales–Bibliotheca*, pages 11–27. Peeters.

[Uckelman and Uckelman, 2007] Uckelman, S. L. and Uckelman, J. (2007). Modal and temporal logics for abstract space-time structures. *Studies in the history and philosophy of modern physics*, 38:673–681.

[van Benthem, 1983] van Benthem, J. (1983). *The Logic of Time: A Model-Theoretic Investigation into the Varieties of Temporal Ontology and Temporal Discourse*. D. Reidel.

[van Benthem, 2004] van Benthem, J. (2004). What one may come to know. *Analysis*, 64(282):95–105.

[Verbrugge, 2017] Verbrugge, R. L. (2017). Provability Logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2017 edition.

[von Wright, 1951] von Wright, G. H. (1951). Deontic logic. *Mind*, 60:1–15.

[Wajsberg, 1931] Wajsberg, M. (1931). Aksjomatyzacja trójwartościowego rachunku zdań (an axiomatization of the three-valued propositional calculus). *Comptes Rendus*, III(24):126–148. Reprinted in *Polish Logic*, Storrs McCall, ed., (Oxford: Clarendon Press: 1967), pp. 264–284.

[Wang, 1957] Wang, H. (1957). The axiomatization of arithmetic. *Journal of Symbolic Logic*, 22(2):145–158.

[Welton, 1891] Welton, J. (1891). *A Manual of Logic*, volume 1. London: W. B. Clive & Co.

[White, 1984] White, M. J. (1984). The necessity of the past and modal-tense logic incompleteness. *Notre Dame Journal of Formal Logic*, 25(1).

[Woodruff, 1974] Woodruff, P. W. (1974). A modal interpretation of three-valued logic. *Journal of Philosophical Logic*, 3:433–439.

[Yrjönsuuri, 1994] Yrjönsuuri, M. (1994). *Obligationes: 14th Century Logic of Disputational Duties*, volume 55 of *Acta Philosophica Fennica*. Societatis Philosophia Fennica.