Mathematical Background to the Logic of Ibn Sīnā

Wilfrid Hodges

Draft 21 June 2016

Or a person goes for a walk and
stumbles across buried treasure.
Ibn Sīnā, *Burhān* 127.20

This is an incomplete draft; some chapters are missing, and everything needs further editing and checking. But it's virtually complete on the assertoric logic and it gives the main shapes of the two-dimensional part. My vigorous thanks to Stephen Read for sending me some corrections; I will be equally grateful for corrections from other sources.

A lot of the Exercises have solutions attached. It's not intended that there should be many (or any?) solutions given in the final text. These solutions are mainly for safety—I remember Paul Cohn's embarrassment when at a publisher's request he tried to supply solutions to exercises for the second edition of his Algebra text.

A previous draft had a lot more discussion of Ibn Sīnā's text. It became clear that there isn't room for both that and the mathematical development in the same book, so the present draft was got largely by separating out the mathematics. The more textual material is being assembled for a book that Brill invited me to submit. The chapter on Propositional Logic hasn't yet had the textual discussion taken out; but probably that material should go into a separate paper. Two other related papers that I would be happy to have out soon are (1) a paper on the non-circular inconsistent sets and (2) a paper on Ibn Sīnā's view of logic as a science.

Contents

Part I

Preliminary

Chapter 1

Ibn Sīnā and his logic

1.1 Ibn Sīnā

Abū ᶜĀlī al-Ḫusayn bin ᶜAbdallāh bin Sīnā, also known in the West as Avicenna, was born not later than 980 near Bukhara in the south of present-day Uzbekhistan, around the western end of the Silk Road. He died in 1037, after a career spent mostly in the western parts of Persia, within the bounds of modern Iran.

Over the centuries he has had a colossal influence, both as a metaphysician and as a medical writer. (The notion of 'abstraction' in philosophy owes as much to him as to anybody.) But our concern in this book is with his work in formal logic. Not all of his writings in logic have survived, but what we do have runs to many hundreds of pages, mainly in Arabic and some in Persian.

These writings place Ibn Sīnā as a Peripatetic logician, i.e. a logician who treats Aristotle's syllogisms as the starting-point of logic. But they also show him to be an innovative and rebellious Peripatetic, comparable in many ways with Leibniz and Frege, though he lacked the mathematical knowhow of these two later logicians. As far as we know, none of his formal logic was translated into any western language before modern times. The first scholarly edition of his major surviving treatise in formal logic, *Qiyās* ('Syllogism', [55]), was published in Cairo in 1964. So his influence on western logic, if he had any at all, was indirect. (In 1508 a work of his was published in Venice in Latin translation as *Logyca* [51], but it contains no formal logic; mainly it is an introduction to fundamental concepts of philosophy.)

In the Muslim world Ibn Sīnā's logic carried greater weight. True, for over a hundred years after his death, other logical writers quoted him but with little insight, and no significant progress was made in developing his ideas. But in the second half of the 12th century another major figure in Islamic logic, Fak̲r al-Dīn al-Rāzī, diagnosed the main obstacle between Ibn Sīnā and his readers, and coined a name for it: *k̲abṭ*, 'stumbling around' (*Lubāb* [84] 185.10, *Mulak̲k̲as* [82] 150.2—Rāzī makes his complaint about 'the logicians', but certainly he has Ibn Sīnā in his sights). On Rāzī's analysis, Ibn Sīnā had confused himself and his readers by failing to distinguish between temporal modalities like 'permanent' and alethic modalities like 'unavoidable'. So Rāzī proposed to consider sentences that contain both kinds of modality independently, thus inventing what people now describe as 'products of modal logics'. Rāzī's innovation was hugely successful, because it allowed people to incorporate many of Ibn Sīnā's ideas into a framework that they could understand. It led to a flowering of 'Avicennan' logic, whose influence has lasted down to the logical textbooks used in modern Iranian madrasas. But this logic was not Ibn Sīnā's logic.

Ibn Sīnā was a systematic thinker, and his formal logic deserves to be treated in the round. The last millennium has somehow denied him this honour. But now that his main texts are widely available (at least in Arabic), he will surely not have too long to wait. In fact the project of describing and assessing his formal logic as a whole in the light of modern knowledge is already under way, though it has so far been in the hands of a few individuals working separately. One should mention the essays of Nicholas Rescher and his colleague Arnold van der Nat [88], [89] in the 1960s. More recently Zia Movahed [78], [79] and Saloua Chatti [19] have worked on particular texts that define logical notions. There is also a book [75] on Ibn Sīnā's propositional logic by Miklos Maróth. Progress is slow but it is real.

The present book aims to state the main themes of Ibn Sīnā's formal logic in terms that a modern logician can follow, and to work out the logical properties of some of the systems that he introduced.

> *But how can we know that we have correctly identified the logical notions and questions that Ibn Sīnā is discussing, given that he uses a completely different notation from us, some of his aims are clearly different from ours, and in some cases even the Arabic text is insecure?*

Well, we can't always. There are many opportunities for disagreement about the correct translation of a single isolated passage. But in practice the problem is much less severe than it might have been, for three main reasons.

(1) There are very few topics that Ibn Sīnā discusses in just one place. Usually he comes at an issue in several passages in different books, from slightly different angles. This is a tremendous help for clearing up ambiguities. Search engines have a role to play here, though there is no replacement for careful reading of each text in its own context.

(2) Ibn Sīnā is unusually forthcoming about his own intentions and aims. So we know what form he thought logic should take as a theoretical science, and we know quite a bit about how he wanted various parts of his logic to be considered. We also have Ibn Sīnā's own warnings about ways in which some of his writings might mislead the reader, because of what he included or failed to include.

(3) The formal systems that we describe in this book do hang together remarkably well. Not only that, but features that we uncover by our mathematical analyses make sense of moves that Ibn Sīnā himself performs; if we didn't know the logical facts, we would be hard put to explain what he is doing. There are a few places where our formalisations make him say things that are certainly false. Some of these illustrate his own warnings in (2) above; in the remaining cases it seems that we can pin down some points of logic which he frankly misunderstood, and usually we can see how he came to misunderstand them.

It should be clear that all of these three reasons call for a substantial amount of research on Ibn Sīnā's texts and their context in eleventh century philosophy. This research will not be mathematical. But it will need to be informed by the mathematical facts; and surely there will be some mathematical readers who have the resources and the interest to probe the Arabic originals. I have tried to deal with this situation by giving fairly full citations of texts (and with an index of citations), even though not all the texts are available yet in western languages.

There are also references to two other books [47] and [45] now in preparation, both of which will discuss the texts. [45] is intended for non-mathematical readers, and will address the historical task of reconstructing Ibn Sīnā's own understanding of his logic.

Not all recent writers on Ibn Sīnā's logic have accepted the need to use the safeguards mentioned in (1)–(3) above. For example some work on his logic can be described as 'interpretation' of Ibn Sīnā for purposes of raising philosophical questions of independent interest; I have generally kept out of this area. But I have commented on some cases where views or aims have been attributed to Ibn Sīnā that were demonstrably not his.

1.2 Literary sources

This book is meant to be readable without reference to Ibn Sīnā's texts. But I owe the reader a brief statement of the texts used and where one can look for them.

Most of our sources for Ibn Sīnā's logic are the logic sections of various encyclopedias that he wrote. Ibn Sīnā regarded logic as a prerequisite for almost any theoretical study, and so he regularly included logic as the first topic in an encyclopedia. We will use the following texts. (The dates are based on Reisman [87] p. 304. Michot [52] has argued for an earlier dating, but it makes little difference to the relative ordering of the works. See Gutas [34] (second edition) for further information on Ibn Sīnā's writings, and on *Muḵtaṣar* and *Najāt* see Kalbarczyk [65].)

***Muḵtaṣar*:** *Al-muḵtaṣar al-awsat fī al-manṭiq* (Middle Summary on Logic). 1014??. Unpublished; we use the text of the manuscript Nuruosmaniye 2763 (528H), 4894_{54} ff. 253b–303a at Istanbul. (I thank Alexander Kalbarczyk for his help with this.) The text shows Ibn Sīnā's logic in detail but at a slightly less developed state than in *Šifā'* below, and this makes it particularly valuable for studying the development of Ibn Sīnā's thinking.

***Šifā'*:** *Al-šifā'* (The Cure). 1020–1027. This is a massive encyclopedia. The logic section is a loose commentary on Aristotle's *Organon*, which Ibn Sīnā takes to include the *Rhetoric* and the *Poetics*, prefaced with a loose commentary on Porphyry's *Eisagoge*. The heart of the formal logic is in the volume *Qiyās* ('Syllogism'), which runs to over 550 pages [55]. There is an English translation of most of the material on propositional logic in *Qiyās* by Nabil Shehaby [93]. Also some passages of *Qiyās* on modal logic are translated in Tony Street [97], and more translations from *Qiyās* will appear in [47]. At several places in *Šifā'* Ibn Sīnā refers to material that he will put into the *Lawāḥiq* ('Appendices'); some of this material involves interesting mathematical questions. But the *Lawāḥiq* haven't survived, if indeed he ever got around to writing them.

***Najāt*:** *Al-najāt* (The Deliverance). 1027. The logic section is a revised version of the *Al-muḵtaṣar al-aṣḡar fī l-manṭiq* (Shorter Summary on Logic) from about 1014, a lost work close to but distinct from *Muḵtaṣar* above. It is a well-integrated and self-contained account, but it is historically a little problematic because it was published some ten years after it was first written; there was clearly some rewriting before publication but we are not sure how much. The logic section has been translated into English by Asad Q. Ahmed [2].

***Mašriqiyyūn*:** *Al-mašriqiyyūn* (The Easterners). Late 1020s. In a prologue added later to *Šifā'*, Ibn Sīnā explained the relationship between *Šifā'* and *Mašriqiyyūn* as follows:

> I also wrote a book [*Mašriqiyyūn*], in which I presented philosophy as it is naturally [perceived] and as required by an unbiased view which neither takes into account in [this book] the views of colleagues in the discipline, nor takes precautions here against creating schisms .... But as for [*Šifā'*], it is more elaborate and more accommodating to my
> (1.2.1) Peripatetic colleagues. Whoever wants the truth [stated] without indirection, he should seek [*Mašriqiyyūn*]; whoever wants the truth [stated] in a way which is somewhat conciliatory to colleagues, elaborates a lot, and alludes to things which, had they been perceived, there would have been no need for [*Mašriqiyyūn*], then he should read [*Šifā'*]. (Translation slightly adapted from Gutas [34] p. 44f.)

Unfortunately the book was stolen and then burned during Ibn Sīnā's lifetime, but too late for him to think of rewriting it. We do have a few dozen pages of the logic section, and they entirely confirm Ibn Sīnā's account quoted above, though they stop short of material on formal proofs. The logic is that of *Šifā'*, but presented more directly. In spite of its brevity, this is one of our most valuable sources. Regrettably there has never been a critical edition, still less a translation. A text is available [59].

***Dānešnāmeh*:** *Dāneshnāmeh* *ᶜAlā'ī* (Book of Wisdom for ᶜAlā). Late 1020s. This work, in Persian, covers only the most elementary parts of logic. F. Zabeeh [102] has translated the logic section into English. The French translation [1] by Mohammad Achena and Henri Massé has the merit of including the entire work, which gives the logic a context in Ibn Sīnā's philosophical thought.

***Išārāt*:** *Al-išārāt wal tanbīhāt* (Pointers and Reminders). Begun 1030. The work is written in a telegraphic and catalogued style that is a little reminiscent of Wittgenstein's *Tractatus*. Hence much detail is missing, and some developments in *Šifā'* are left out altogether. But there is no doubt about its being a late work; the logic section contains some technical advances on *Šifā'*. The logic section of *Išārāt* is available in a translation by Shams Inati [50].

Translators of Ibn Sīnā tend not to be logicians, so in most cases the engagement with the logic is minimal. But the translations do give a reliable indication of the raw text, and generally they are accompanied by notes containing sound historical scholarship. Translations of other works of Ibn Sīnā that are indirectly relevant to his formal logic include Ahmad Hasnawi [35] and Allan Bäck [9] on *ᶜIbāra* [54], and a forthcoming translation of *Burhān* by Riccardo Strobino.

Street's readable chapter [98] in the Gabbay-Woods Handbook of the History of Logic introduces the main players in medieval Arabic Logic and sets them in their historical context. Its treatment of Ibn Sīnā's formal logic is in the spirit of Rescher's papers. More recent but briefer is Hasnawi and Hodges [36].

1.3 Road map

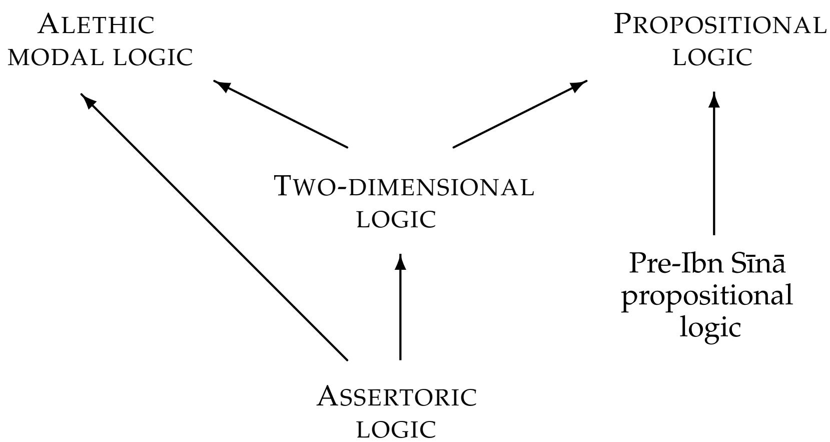
We sketch here the ground to be covered in this book. Some of this sketch will mean little without the details to be added in later chapters, but at least this gives me the chance to highlight Ibn Sīnā's main innovations in logic.

As a Peripatetic logician Ibn Sīnā concerns himself with 'syllogisms', which are roughly speaking the smallest units of inference. But obtusely we will not define 'syllogism'. Ibn Sīnā, and not only he, uses the term so chaotically that it is hardly worth rescuing. For example sometimes he requires a syllogism to have two premises, but sometimes he allows the number of premises to be anything from two upwards. Sometimes he calls a piece of reasoning a syllogism to indicate that it is sound reasoning, but sometimes he distinguishes between sound and unsound syllogisms. Sometimes he takes a syllogism to be a specific mental act, but sometimes he takes it as an inference rule. If we have the notions needed to clear up this mess, we may as well rest our account on those notions instead.

The word 'syllogism' remains available as a rather vague word meaning 'inference of a type considered by Peripatetic logicians'. In translating Ibn Sīnā we have to use the word, because he uses its Arabic equivalent (*qiyas*) all the time.

Ibn Sīnā's contributions to formal logic revolve around the four systems in capital letters in the diagram below. We call them 'logics', though generally they are less well-defined than a modern logician would want. In fact the propositional logic exists in several versions, and the two-dimensional logic is a litle vague in what sentences it includes.

(1.3.1)



At the bottom of the diagram (1.3.1), assertoric logic is Ibn Sīnā's take on the non-modal part of Aristotle's logic, the part sometimes referred to as 'categorical syllogisms'. It underlies all of Ibn Sīnā's logic in very much the same way as boolean algebra underlies modern classical logic. He takes it as a done deal and doesn't claim any originality for his account, which we survey in Part II below. Nevertheless he manages to give assertoric logic several new twists, in his treatment of how complex reasoning in this logic can be built up from simple steps. There are two main aspects to this.

The first has to do with the justification of the less self-evident moods (those that are not 'perfect' in the sense of Section 4.2 below). We discuss Ibn Sīnā's justifications in Chapter 8 below. They follow Aristotle's *Prior Analytics* i.5,6 almost to the letter. But Ibn Sīnā has a requirement that Aristotle didn't have: namely Ibn Sīnā wants to be able to lift the justifications, so far as possible, to justifications of imperfect moods in two-dimensional logic. Our main work in Part III will be to develop two-dimensional logic up to the point where we can say exactly what Ibn Sīnā needed from assertoric logic. Thus he found he needed to add to Aristotle an ecthetic proof of the mood *Baroco*. This solved an immediate problem in an efficient way. But it raises a new problem, namely how to justify this ecthetic proof. A sound answer can be given within Ibn Sīnā's own framework and using tools that he himself introduced. But first, it involves us in giving procedural rather than inferential justifications of some steps. And second, Ibn Sīnā himself never gives this justification, though he sets up the apparatus needed for it. This is one of several places where we can't be sure how far he himself understood the solution of technical problems that he raised.

The second aspect of complex reasoning is to give an account of proofs in assertoric logic that have arbitrary complexity. This is the context in which Ibn Sīnā makes one of his most original contributions. He describes a recursive proof search algorithm, a feat not rivalled in the West until some nine hundred years later. In [41] I put the resulting algorithm into the form of an abstract state machine. The book [47] will review the algorithm more fully; the present text contains the main groundwork that Ibn Sīnā relies on for the algorithm.

Above assertoric in (1.3.1) there are three other logics. Two of these, alethic modal logic and two-dimensional logic, are got from assertoric logic by adding new features.

In the case of alethic modal logic these new features are modal operators 'necessarily', 'possibly' etc.; Aristotle had already introduced these operators in *Prior Analytics* i.8–22. This is probably the least interesting part of Ibn Sīnā's formal logic. None of his modal arguments contain any convincing improvements on those of Aristotle. Worse still, his purely modal methods are all piecemeal, mood by mood; they betray no sign of any overall vision of what a modal logic should do for us.

The position is completely altered when we approach alethic modal logic as Ibn Sīnā himself does, at least from *Qiyās* onwards, namely by way of the arrow from two-dimensional logic in (1.3.1). Two-dimensional logic is a logic that Ibn Sīnā invented. Its outer edges are a little hazy, and in places Ibn Sīnā's account of it is obscured by the *ḳabṭ* that Rāzī complained of. But its central parts form a well-defined and robust logic that is intermediate between assertoric logic and modern first-order logic. Books I to IV of *Qiyās* contain a careful and thorough textbook of this logic, with many interesting details and some striking innovations.

Two-dimensional logic is got from assertoric logic by adding quantification over times. So in this logic sentences can contain multiple quantification, including mixes of universal and existential. There are two sorts, which we call $object$ and $time$; quantification over $object$ corresponds to the single quantifiers in assertoric logic. The sentence forms of two-dimensional logic are defined without any use of intensional notions like 'necessary'. By this and other means, Ibn Sīnā ensures that there is no ambiguity about the truth conditions for sentences of these forms. So two-dimensional logic can be built up as a mathematical system.

In Part III we present this mathematical system, but using modern tools.

Ibn Sīnā gives many of the basic facts, but it's clear that he discovered them by hands-on testing of many examples. As mathematical logicians we can supply the proper metatheoretic proofs where he couldn't. Two-dimensional logic is an order of magnitude more complex than assertoric logic. This complication will feed back into our treatment of assertoric logic, because we will introduce for assertoric logic some methods that we know will generalise to two-dimensional logic; these methods are not always the most efficient ones if our aim was just to develop assertoric logic.

Let me mention two of Ibn Sīnā's discoveries about two-dimensional logic. The first is that Aristotle's justification for some Second Figure syllogisms, such as *Cesare*, doesn't adapt to the corresponding two-dimensional syllogisms. Ibn Sīnā was unable to find any way of using Aristotelian methods to plug this hole (and as of today I haven't found any either). So he advises us instead to fill the gap by using a reduction by paraphrase to assertoric logic. A reduction of this kind was used by Boole to justify propositional logic by reduction to the logic of sets; but for Boole this was a metatheorem, not an ingredient of proofs within propositional logic. By contrast for Ibn Sīnā the reduction was a formal proof step that took one from one logic into another. In this sense it was the first proposal for dealing with 'local formalising', that blemish of the old logics that Leibniz identified and Frege denounced (see Section 4.3 below). Modern classical logic deals with it in a different way, by developing a single logic that contains rules of different types. But some devices of computer science logic fit several logics into a single framework with formal methods for moving between logics, and this can be seen as a return to Ibn Sīnā's proposal.

The second discovery is that minimal inconsistent sets in two-dimensional logic don't always have the circular form that is a characteristic of assertoric logic and determines the shapes of proofs in assertoric logic. As far as I know, this was the first advance after Aristotle in the understanding of the shapes of proofs. Ibn Sīnā seems to have discovered it first with the $(i)I$ types of the dt fragment (cf. Section 10.2 below), but by the time of *Išārāt* he knew other examples outside this fragment. Ibn Sīnā was well aware that his discovery was a challenge to the authority of Aristotle. In fact he presents the discovery as a refutation of a method used by Aristotle in his treatment of modal syllogisms. On the usual reading of Aristotle's text there certainly is an irreparable gap in Aristotle's argument, exactly where Ibn Sīnā pointed it out, but it is a subtle one and in the West it seems to have been first noticed by Paul Thom in 1996.

To return briefly to Ibn Sīnā's alethic modal logic: my own belief is that Ibn Sīnā intended to deduce the (controversial) logical laws of the abstract notion of necessity from the (robust) logical laws of two-dimensional logic, by translating the alethic modal sentences into two-dimensional sentences. In the present book we study in detail only the formal aspects of this deduction. But one of these aspects has a broader interest. Without knowing Ibn Sīnā's logic or any research on it, Spencer Johnston in a recent PhD thesis [64] developed a Kripke semantics for the divided modal logic of the 14th century Scholastic logician Jean Buridan. His semantics turns out to be a notational variant of a major part of two-dimensional logic; I demonstrate this in Chapter 12 below. With hindsight, Johnston's use of Kripke semantics to explicate Buridan's divided modal logic should throw light on Ibn Sīnā's use of two-dimensional logic to support alethic modal logic. This is discussed more fully in [45] and in a joint paper with Johnston [48].

There is a complication—there always is with Ibn Sīnā. He doesn't give a translation of the whole of modal logic into two-dimensional logic. Instead he gives separate and incompatible translations for *pairs* of modalities, and one pair in particular gets shabby treatment. I discuss the reason for this in [45].

There remains Ibn Sīnā's propositional logic. This is more fragmented than his other logics. In fact there are clear signs of three incompatible levels. From the internal evidence these look like three stages in the development of his research, though since the two upper levels are given in detail only in *Qiyās*, we can't show the development taking place. The bottom layer, PL1, is barely distinguishable from what has reached us of propositional reasoning in the logic writings of Ibn Sīnā's predecessor Al-Fārābī. It has very little formal content. The few formal rules that it contains can all be reconstrued as definitions of different notions of 'if' and 'or'; sentences with 'if' are called *muttaṣil* and sentences with 'or' are called *munfaṣil*.

However, this small formal content does include rules such as modus ponens and modus tollens. Ibn Sīnā points out a formal difference between these rules and the two-premise syllogisms of most of his other logics (assertoric, two-dimensional, modal, PL2 below). Namely, in these other two-premise syllogisms there is a term common to the two premises; we form the conclusion by removing this term and recombining the remaining pieces from both premises. So he describes these syllogisms as 'recombinant' (*iqtirānī*). But for example if we reason

(1.3.2) p. If p then q. Therefore q.

then our conclusion takes nothing from the first premise. Instead the first premise is a pure duplicate of part of the second premise, and so Ibn Sīnā describes this rule as 'duplicative' (*istiṯnā'ī*). This division into recombinant and duplicative is a first attempt at cataloguing logic beyond the bounds of what the Peripatetic tradition handed down to Ibn Sīnā, and he justifies it with a description of how our minds work. What he gives is not so much a psychological description as an analysis of what an inference engine would need to do in terms of syntactic manipulation. His inference engine (Arabic *bāl*) is somewhere between a device for association of ideas and a Turing machine. It is considered in more detail in [45], and [47] spells out some of its implications for the design of a proof calculus.

Returning to propositional logic, the level PL2 is assertoric logic but with time instead of object quantifiers. So for example in place of 'Every A is a B' we have 'Every time when A is a time when B'. Ibn Sīnā shows that PL2 can also be seen as development of the logic of *muttaṣil* sentences. But formally the logic is isomorphic to assertoric logic. This is probably the first time that any logician gave two logically distinct interpretations of the same formalism.

So PL1 is formally not very interesting and PL2 is formally not new. The level PL3 is got from assertoric logic by developing *munfaṣil* sentences in the same way as PL2 developed *muttaṣil* sentences. This is a more dramatic innovation than PL2, because the facts of *munfaṣil* sentences require Ibn Sīnā to allow negations to occur more or less anywhere in the sentences of this logic. Formally it's equivalent to allowing both affirmative and negative terms throughout assertoric logic; we call the result *metathetic logic*. The system has been studied under other names in some recent computer science literature.

Ibn Sīnā follows Aristotle's procedure for showing when a premise-pair yields no syllogistic conclusion. For PL3 certain changes to the procedure are needed, and Ibn Sīnā omits to make them. As a result several of his claims are wrong. This is one place where he shows every sign of being out of his depth.

Ibn Sīnā springs one more surprise on us in his propositional logic, though it is not clear whether it belongs with PL1 or PL3. (It doesn't fit with PL2.) In a discussion of reductio ad absurdum, he addresses the question how proofs by reductio can be regarded as convincing. His answer is that every such proof contains sentences with hypotheses that are silently intended and not spoken. He shows that in this way every such proof can be regarded as direct, in the sense that each sentence either follows from the ones immediately preceding it, or is an axiom. His account is remarkably close to the one given by Frege in [30]. Moreover it works not only for reductio ad absurdum, but for all cases where an assumption is made and later discharged by $\rightarrow$-introduction. But there is a problem: Ibn Sīnā implicitly uses a principle saying that if an inference is valid then it remains valid if we add hypotheses in the way that Ibn Sīnā describes. In Chapter N we discuss what Ibn Sīnā himself does to demonstrate this principle, and what the strength of the resulting system would be if he adopted the principle as a derived proof rule.

In fact a strong version of this principle, going way beyond anything that we find in Ibn Sīnā's logical practice, would give him the whole of first order logic and hence an undecidable formal system. As it happens, all the formal systems that Ibn Sīnā explicitly describes are decidable, for example because they have the finite model property. One can ask how close he comes to an undecidable system. Mohammed Maarefi discussed this question in a recent thesis in Tehran, and the discussion in Chapter N should be regarded as joint work with him; we also benefited from a conversation with Erich Graedel.

Chapter 2

Mathematical preliminaries

Sections 2.1 and 2.2 consist mainly of definitions and are for reference.

In modern logic both theories and formal proofs are constructed from formulas, but in very different ways. A theory is just a set of formulas, while a formal proof puts the formulas into a highly structured framework. Much of Ibn Sīnā's formal logic takes place in an area somewhere between these two extremes; generally a theory is linearly ordered, and a proof from the theory joins up the two ends of the linear ordering with the conclusion. We will use graph theory, as in Section 2.3, to provide the required structures.

2.1 First-order logic

Nearly all the notions in this section are standard first-order logic, as for example in Shapiro [92] or the earlier parts of the textbook of Shoenfield [95]. With a few exceptions that we explain as we come to them, the logic is formal in the sense that we give no fixed interpretations to the relation symbols etc. in signatures (see Definition 2.1.1 below). In general our languages will not include $=$.

Definition 2.1.1 A *signature* is a set of symbols, each assigned to one of the following classes:

- propositional symbols p, q, r, ...
- individual constant symbols a, b, c, ... (These are often referred to simply as *constants*.)

- relation symbols each of a fixed arity $\geqslant 1$, P, Q, R, ...
- function symbols each of a fixed arity $\geqslant 1$, f, g, h, ...

A relation symbol of arity 1 is said to be *monadic*; a relation symbol of arity 2 is said to be *binary*.

Definition 2.1.2 A signature Σ is *relational* if all the symbols in it are relation symbols. It is *monadic relational* if all the symbols in it are relation symbols of arity 1.

Definition 2.1.3 Each signature Σ has a corresponding first-order language L(Σ). The grammatical expressions of $L(\Sigma)$ are all either terms or formulas; the class of terms and the class of formulas are inductively defined as in Definitions 2.1.4 to 2.1.6 below. The symbols of $L(\Sigma)$ are those in Σ, the logical symbols introduced in the inductive definitions of terms and formulas, and an infinite set of *variables* which we write as x, y, z, x_1, x_2 etc.

Definition 2.1.4 The *terms* of L(Σ) form an inductively defined set: each individual constant symbol in Σ and each variable is a term of L(Σ), and if f is a function symbol of arity n in Σ and $t_1, \ldots, t_n$ are terms of L(Σ) then $f(t_1, \ldots, f_n)$ is a term of L(Σ).

Definition 2.1.5 The *atomic formulas* of L(Σ) are the propositional symbols in Σ, and the expressions $R(t_1, \ldots, t_n)$ where R is a relation symbol of arity n in Σ and $t_1, \ldots, t_n$ are terms of L(Σ).

Definition 2.1.6 The *formulas* of L(Σ) are defined inductively. The symbol $\perp$ is a formula of L(Σ), every atomic formula of L(Σ) is a formula of L(Σ); if ϕ and ψ are formulas of L(Σ) then so are $\neg\phi$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$ and $(\phi \to \psi)$; and if ϕ is a formula of L(Σ) and x an individual variable, then $\forall x\phi$ and $\exists x\phi$ are formulas of L(Σ).

Definition 2.1.7 (a) The formula $\neg\phi$ is called the *negation* of ϕ. A *literal* is either an atomic formula or the negation of an atomic formula.

(b) Expressions of the form $(\phi_1 \wedge \ldots \wedge \phi_n)$ with $n > 1$ are called *conjunctions*; their *conjuncts* are the expressions ϕ_i ($1 \leqslant i \leqslant n$).

(c) Expressions of the form $\forall x$ and $\exists x$ are called respectively *universal quantifiers* and *existential quantifiers*.

Definition 2.1.8 A *subformula* of a formula ϕ is a formula ψ which occurs as a segment of ϕ (possibly the whole of ϕ). An *occurrence* of ψ in ϕ is such a segment of ϕ. (These are intuitive definitions of notions that all logicians use. See the standard texts for a formal definition of 'subformula' which bypasses the notion of an occurrence.)

Definition 2.1.9 If ϕ is a formula, then the *scope* of an occurrence of a quantifier Qx in ϕ is the occurrence of a subformula of ϕ which begins with Qx. An occurrence of a variable y in ϕ is *bound* if it lies within the scope of an occurrence of a quantifier Qy with the same variable; otherwise it is *free*. The *free variables* of ϕ are the variables that have free occurrences in ϕ.

Definition 2.1.10 A term is *closed* if it contains no variables. Likewise a *closed literal* is a literal that contains no variables. A formula is a *sentence* if it has no free variables. A *theory* is a set of sentences.

Definition 2.1.11 A formula is *quantifier-free* if it contains no quantifiers. A formula ϕ is *prenex*, or in *prenex form*, if it has the form

$$Q_0x_0 \ldots Q_{n-1}x_{n-1}\psi \tag{2.1.1}$$

where $n \geqslant 0$ and each Q_i is either $\forall$ or $\exists$, and ψ is quantifier-free. An *instance* of a prenex formula ϕ is a formula got by removing zero or more quantifiers from the beginning of ϕ, and replacing the resulting free variables by closed terms (so that all occurrences of the same variable are replaced by the same closed term).

Definition 2.1.12 We write $\bar{c}$ for a tuple (i.e. finite sequence) $(c_1, \ldots, c_n)$. We can introduce the term t as $t(\bar{x})$ when the variables of t are all included in the tuple $\bar{x}$ of distinct variables; then if $\bar{s}$ is a sequence of terms, of the same length as $\bar{x}$, we write $t(\bar{s})$ for the term got by writing s_i in place of each occurrence of the variable x_i, for each i. Likewise we can introduce the formula ϕ as $\phi(\bar{x})$ when the free variables of ϕ are all included in the tuple $\bar{x}$; then we write $\phi(\bar{s})$ for the formula got by writing s_i in place of each free occurrence of x_i, for each i.

The definitions above describe the first-order language $L(\Sigma)$. The standard semantics of first-order languages interprets the expressions of $L(\Sigma)$ in a class of structures known as Σ-structures, as follows.

Definition 2.1.13 Each signature Σ has corresponding to it a class of structures, known as Σ-*structures*. If A is a Σ-structure, then for each propositional symbol p in Σ, p^A is a truth-value $\in \{True, False\}$. If Σ contains individual constant, relation or function symbols, then A carries a nonempty set dom(A) (the *domain* of A), and the elements of dom(A) are called the *elements* of A. The *cardinality* of A is defined to be the cardinality of its domain. For each individual constant a in Σ, $a^A \in$ dom(A). For each relation symbol R of arity n in Σ, $R^A \subseteq \text{dom}(A)^n$. For each function symbol f of arity n in Σ, $f^A : \text{dom}(A)^n \to \text{dom}(A)$.

Definition 2.1.14 Let A be a Σ-structure and $\bar{a}$ a tuple of (not necessarily distinct) elements of A, of the same length as $\bar{x}$. Then $t^A[\bar{a}]$ (the interpretation of t in A at $\bar{a}$) is an element of A defined in the obvious way. Similarly we define in the obvious way when $\bar{a}$ satisfies ϕ in A, in symbols $A \models \phi[\bar{a}]$. If ϕ is a sentence and the empty tuple satisfies ϕ in A, we say that ϕ is *true in* in A or that A is a *model of* ϕ, in symbols $A \models \phi$. If T is a theory in L(Σ), we say that A is a *model* of T if A is a model of every sentence in T.

Lemma 2.1.15 (Canonical Model Lemma) *Let Σ be a signature and T a set of closed literals of $L(\Sigma)$, none of them containing $=$, such that there is no atomic sentence θ for which T contains both θ and $\neg\theta$. Then there is a Σ-structure M such that for every atomic sentence θ of $L(\Sigma)$ not containing $=$,*

$$M \models \theta \iff \theta \in T. \tag{2.1.2}$$

A model with this property is called a canonical model *of T,*

Proof. See for example the proof of Lemma 1.5.1 on page 18 of [38], ignoring equations. □

Lemma 2.1.16 *Let Σ be a signature, T a set of sentences of $L(\Sigma)$ and M a model of T. Suppose also that no sentence of T contains the symbol $=$, no function symbol is in Σ, and the domain of M has cardinality κ. Then for every cardinal $\lambda \geqslant \kappa$, T has a model N whose domain has cardinality λ.*

For the *proof* we choose an element a of the domain of M, and we take λ new elements a_i ($i < \lambda$); the domain of the Σ-structure N will consist of the domain of M with the new elements added. For every finite sequence $\bar{b}$ of elements of the domain of N, we write $\bar{b}^\star$ for the sequence that results from $\bar{b}$ by replacing each new element a_i with $i < \lambda$ by a. The relations R^N are defined by taking $\bar{b} \in R^N$ if and only if $\bar{b}^\star \in R^M$. The proof that N is a model of T is then by induction on the complexity of formulas. □

Definition 2.1.17 (a) Two formulas $\phi(\bar{x})$, $\psi(\bar{x})$ of $L(\Sigma)$ are said to be *logically equivalent*, in symbols $\phi(\bar{x}) \equiv \psi(\bar{x})$, if for every Σ-structure A and every tuple $\bar{a}$ of appropriate length, $A \models \phi[\bar{a}]$ if and only if $A \models \psi[\bar{a}]$.

(b) When T is a set of sentences of $L(\Sigma)$, two formulas $\phi(\bar{x})$, $\psi(\bar{x})$ of $L(\Sigma)$ are said to be *equivalent modulo* T if the same condition holds as (a), but with A limited to models of T.

Lemma 2.1.18 *Every first-order sentence is logically equivalent to a sentence in prenex form.*

Proof. Cf. Shoenfield [95] pp. 36–39. □

Definition 2.1.19 A *sequent* is an expression of the form

$$T \vdash \psi \tag{2.1.3}$$

where T is a theory in some language $L(\Sigma)$. If T is a finite set $\{\phi_1, \ldots, \phi_n\}$, we can write (2.1.3) as

$$\phi_1, \ldots, \phi_n \vdash \psi. \tag{2.1.4}$$

We say that the sequent (2.1.3) is *valid* if every Σ-structure M that is a model of T is a model of ψ. (It can be shown that this condition is independent of the choice of signature Σ, which can be any signature containing all the relation etc. symbols of the sentences.) The sentences in T are the *premises* of (2.1.3), and ψ is its *conclusion*. We often write '$T \vdash \psi$' to express that the sequent (2.1.3) is valid.

Lemma 2.1.20 *(a) For any sentence ϕ the sequent $\phi \vdash \phi$ is valid.*

(b) If $T \vdash \chi$ is valid and U is any set of sentences, then $T \cup U \vdash \chi$ is valid.

(c) If $T \vdash \phi$ is valid for all $\phi \in U$, and $U \cup V \vdash \psi$ is valid, then $T \cup V \vdash \psi$ is valid.

□

Definition 2.1.21 Let ϕ and ψ be sentences of $L(\Sigma)$. We say that ϕ is *stronger than* ψ, and that ψ is *weaker than* ϕ, if $\phi \vdash \psi$ but not $\psi \vdash \phi$.

Definition 2.1.22 A theory T is *consistent* if it has a model, and *inconsistent* if it has no model.

Theorem 2.1.23 (Compactness Theorem) *Every inconsistent first-order theory has an inconsistent finite subset.*

Proof. See Shoenfield shoe:1 p. 69 or Shapiro [92] §5. □

Definition 2.1.24 For every sequent (2.1.3), its *antilogism* is the theory

$$T \cup \{\neg\psi\}. \tag{2.1.5}$$

Lemma 2.1.25 *Let T be a first-order theory and χ a first-order sentence. Then $T \vdash \psi$ if and only if the antilogism $T \cup \{\neg\chi\}$ is inconsistent.* □

Conversely if T is a theory, then we can take any sentence $\chi \in T$ and form the sequent

$$T \setminus \{\chi\} \vdash \neg\chi.$$

Again we have that the theory is inconsistent if and only if the sequent is valid.

We will go outside the range of ordinary first-order languages in two places. One is in the theory of two-dimensional logics; for this we will need the notion of a *many-sorted language* $L(\Sigma)$. In fact we will only be interested in the case where there are two sort s, *object* and *time*, and we need only explain this case. Each term of $L(\Sigma)$ has one of the two sorts. We have *object variables* x, y, z etc., and *time variables* ρ, σ, τ etc.; the object variables are terms of sort *object* and the time variables are terms of sort *time*. Likewise the individual constants of Σ are each assigned to one of the two sorts. Each of the argument slots of a relation symbol has a sort, and only elements of this sort can fill the slot. Likewise for a function symbol f, except that we also need to say what sort a term has if it begins with f. For this two-sorted signature Σ, a Σ-structure A has two disjoint domains, $\text{dom}_{object}(A)$ consisting of elements of sort *object* and $\text{dom}_{time}(A)$ of elements of sort *time*. A relation $R^A x\tau$ is a subset of $\text{dom}_{object}(A)\times \text{dom}_{time}(A)$. A function $f^A x\tau$ of sort *time* is a function from $\text{dom}_{object}(A)\times \text{dom}_{time}(A)$ to $\text{dom}_{time}(A)$.

We will also consider modal languages, with operators □ 'necessarily' and ◊ 'possibly' that can be put in front of formulas. We will treat these languages as we come to them.

2.2 Theories and expansions

Throughout the previous section we assumed that the relation symbols have no fixed meaning; they can be given a meaning by supplying a Σ-

structure where the symbols are in Σ. But there are a few situations where we want to give some of the symbols either a fixed meaning, or at least a fixed relationship to the other symbols. If this relationship can be written down as a theory T, then we can reason with it by including T among the premises of any inference.

Definition 2.2.1 A set of *meaning postulates* in a first-order language $L(\Sigma)$ is a theory in $L(\Sigma)$ that is intended to express some feature of the meaning of one or more symbols in Σ.

For example some people (though not Ibn Sīnā) count it as part of the meaning of a monadic relation symbol A that A is not empty, i.e. $\exists x Ax$. What these people are doing is to treat the set

(2.2.1) $$\{\exists x Ax : A \text{ a monadic relation symbol}\}$$

as a set of meaning postulates. Another kind of example occurs with definitions, as follows.

Definition 2.2.2 Let Σ be a signature, R a relation symbol not in Σ, and Σ^+ the signature got by adding R to Σ. When we speak of *adding R by definition* we mean that there is some formula $\phi(\bar{x})$ of $L(\Sigma)$ such that we reason in $L(\Sigma^+)$ using the sentence

(2.2.2) $$\forall \bar{x}(R\bar{x} \leftrightarrow \phi(\bar{x}))$$

as a meaning postulate. The sentence (2.2.2) is called an *explicit definition* of R in $L(\Sigma)$.

Many examples occur in set theory, where new set-theoretic symbols are introduced by definitions (usually more complicated than (2.2.2)), and then the definitions are used as tacit axioms.

History 2.2.3 Meaning postulates were described and named by Rudolf Carnap [17]. Explicit definitions are one of his examples.

We will be describing proof theories for various logical systems. These proof theories have one or other of two different purposes. The first purpose is *cognitive*, to express how we can come to know that something is true by deducing it from other things known to be true. This was a matter of intense interest to Ibn Sīnā, and our cognitive proof theories will be constructed so far as possible to reconstruct his views. The second purpose is *metamathematical*, to discover the facts about validity and entailment in a given logic. For this second purpose there is no need to restrict ourselves to methods that Ibn Sīnā himself could have used.

The chief methods that we will use for this second purpose involve Skolem functions and the Herbrand universe. We briefly describe these here. The procedure is standard; essentially it is in Hilbert and Ackermann [37] sections III.8 and III.12. There is a good account in Gallier [31] sections 7.5 and 7.6.

Definition 2.2.4 Let Σ_1 and Σ_2. We write $\Sigma_1 \subseteq \Sigma_2$ to mean that every symbol in Σ_1 is also a symbol in Σ_2, and has the same type (e.g. constant, or relation symbol of arity 2) in both signatures. If M is a Σ_2-structure and $\Sigma_1 \subseteq \Sigma_2$, then we can construct a Σ_1-structure N from M by leaving unchanged the domain and putting $R^M = R^N$ etc. for the symbols R etc. in Σ_1, but leaving out the symbols not in Σ_1. We say that M is an *expansion* of N to Σ_2, and N is the *restriction* of M to Σ_1.

Theorem 2.2.5 *Let Σ be a signature and ϕ a sentence of $L(\Sigma)$. Then there are a signature Σ^{sk} and a sentence ϕ' of $L(\Sigma^{sk})$ such that*

(a) $\Sigma \subseteq \Sigma^{sk}$, and the symbols in $\Sigma^{sk} \setminus \Sigma$ are at most individual constants and function symbols;

(b) ϕ' is prenex with at most universal quantifiers in its initial quantifier string;

(c) for every Σ-structure M that is a model of ϕ there is a Σ^{sk}-structure M^+ that is an expansion of M and a model of ϕ';

(d) every Σ^{sk}-structure that is a model of ϕ' is also a model of ϕ.

Proof sketch. A skolemisation of a sentence ψ is also a skolemisation of any sentence logically equivalent to ψ. So by Lemma 2.1.18 we can assume that ϕ is in prenex form (Definition 2.1.11). Let ϕ_0 be ϕ with all existential quantifiers in the quantifier prefix left out. If for example $\exists y$ was one of these quantifiers, then any occurrences of y in ϕ_0 will be free. For each such variable y we introduce a term $f(\bar{x})$, where f is a new function symbol and $\bar{x}$ are the variables in the universal quantifiers before $\exists y$ in the quantifier prefix. Then we replace each occurrence of y by the term $f(\bar{x})$. If $\bar{x}$ is empty, i.e. there are no universal quantifiers before $\exists y$, then instead of $f(barx)$ we introduce a new constant c. The sentence ϕ' is the result of making all these replacements in ϕ_0, and Σ^{sk} is formed by adding the new function symbols f and constants c to Σ. See for example Shoenfield [95] p. 56. □

Definition 2.2.6 (a) Any sentence ϕ' as in Theorem 2.2.5 is called a *skolemisation* of ϕ. The symbols added to the signature are called the *Skolem constants* and *Skolem functions*.

(b) Let Σ be a signature and Φ a basic set of formulas of $L(\Sigma)$. Let T be a theory in $L(\Sigma)$. The *skolemisation* Sk(T) of T is the theory consisting of a skolemisation of each sentence in T, *using different added symbols for different sentences* (and the proof of Theorem 2.2.7 below will show the reason for this crucial requirement). Generalising the notation of Theorem 2.2.5, we write Σ^{sk} for the signature that consists of exactly the symbols of Σ together with the Skolem constants and Skolem functions used to form Sk(T).

For example a possible skolemisation of the theory

$$\forall x \exists y Bxy,\ \exists x \forall y \exists z (Bxy \wedge (Bxz \to Byx)) \tag{2.2.3}$$

is

$$\forall x Bxf(x),\ \forall y (Bcy \wedge (Bcg(y) \to Byc)). \tag{2.2.4}$$

Theorem 2.2.7 *Let Σ be a signature. Let T be a theory in $L(\Sigma)$ and Sk(T) its skolemisation in $L(\Sigma^{sk})$. Then T is consistent if and only if Sk(T) is consistent.*

Proof. If M is a model of T, then for each sentence $\phi \in T$ we can expand M to a model of the skolemisation ϕ' by Theorem 2.2.5(c). Since the new symbols are distinct for each sentence, we can make all these expansions simultaneously, and the result is a model of Sk(T) with signature Σ^{sk}. Conversely if N is a model of Sk(T), then N is a model of the skolemisation ϕ' of each sentence $\phi \in T$, and so N is a model of ϕ by Theorem 2.2.5(d). □

Definition 2.2.8 Let Σ be a signature and T a theory in $L(\Sigma)$. Let Sk(T) be the skolemisation of T in the language $L(\Sigma^{sk})$. Then the *Herbrand universe* of T is the set of closed terms (cf. Definition 2.1.10) of $L(\Sigma^{sk})$. The *Herbrand theory* Hr(T) of T is the set of sentences of $L(\Sigma^{sk})$ that are got by removing the initial universal quantifiers on any sentence of Sk(T) and replacing each free variable (at all occurrences) by a closed term in the Herbrand universe, in all possible ways.

Lemma 2.2.9 *Let T be a theory in a language $L(\Sigma)$ with at least one individual constant, and Hr(T) its Herbrand theory in the corresponding language $L(\Sigma^{sk})$.*

Suppose Σ^{sk} contains at least one individual constant. Then T is consistent if and only if Hr(T) is consistent.

Proof. Suppose first that T has a model M. Taking a sample sentence $\forall_1 x_1 \ldots \forall_n x_n \phi(x_1, \ldots, x_n) \in T$,

$$M \models \forall_1 x_1 \ldots \forall_n x_n \phi(x_1, \ldots, x_n). \tag{2.2.5}$$

Let $t_1, \ldots, t_n$ be terms in the Herbrand universe, and for each i let a_i be the element t_i^A named by t_i in M. Then removing the quantifiers,

$$M \models \phi[a_1, \ldots, a_n]. \tag{2.2.6}$$

But this is the condition for M to be a model of $\phi(t_1, \ldots, t_n)$.

Second, suppose that the Σ^{sk}-structure N is a model of Hr(T). Let N_0 be the substructure of N consisting of the elements named by terms in the Herbrand universe. (This is a well-defined substructure since the presence of an individual constant guarantees that the Herbrand universe is not empty.) Then since Hr(T) is quantifier-free, N_0 is also a model of Hr(T). Considering any sentence $\forall_1 x_1 \ldots \forall_n x_n \phi(x_1, \ldots, x_n)$ in T, and any elements $a_1, \ldots, a_n$ of N_0, there are terms $t_1, \ldots, t_n$ in the Herbrand universe that name $a_1, \ldots, a_n$ in N_0. By assumption

$$N_0 \models \phi(t_1, \ldots, t_n) \tag{2.2.7}$$

and hence

$$N_0 \models \phi[a_1, \ldots, a_n]. \tag{2.2.8}$$

It follows that

$$N_0 \models \forall_1 x_1 \ldots \forall_n x_n \phi(x_1, \ldots, x_n). \tag{2.2.9}$$

Hence N_0 is a model of T. □

2.3 Labelled digraphs

Definition 2.3.1 By a *directed graph* or *digraph*, or an *abstract digraph* to distinguish it from the labelled digraphs below, we mean a sequence $\Gamma = (N, A, \sigma, \tau)$ where

(a) N is a nonempty set and A is a set disjoint from N; the elements of N are called *nodes* and the elements of A are called *arrows*;

(b) σ and τ are 1-ary functions from A to N; for each arrow a in A, the node $\sigma(a)$ is called the *source* of a and the node $\tau(a)$ is called the *target* of a.

(c) We say that a is *from* its source and *to* its target. We also say (not distinguishing source from target) that an arrow is *between* its source and its target. Two nodes μ and ν of Γ are said to be *neighbours* if they are distinct and there is an arrow between them.

The digraph is called *loopless* if it has no arrow a with $\sigma(a) = \tau(a)$. It is called *finite* if it has finitely many nodes and finitely many arrows. Its *node size* is the number of its nodes; its *arrow size* is the number of its arrows.

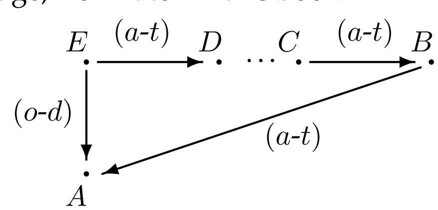
Definition 2.3.2 By a *labelled digraph* we mean a sequence $\Gamma = (N, A, \sigma, \tau, \lambda_N, \lambda_A)$ where (N, A, σ, τ) is a digraph and moreover

(c) λ_N and λ_A are 1-ary functions with domains N and A respectively; for each node ν in N, $\lambda_N(\nu)$ is called the *label on* ν, and for each arrow a, $\lambda_A(a)$ is called the *label on* a.

The digraph (N, A, σ, τ) is called the *underlying abstract digraph* of Γ.

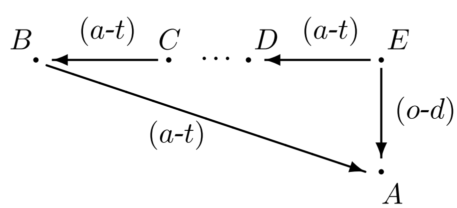
The digraphs and labelled digraphs that we deal with will always be finite, i.e. the sets of nodes and arrows will be finite. They will also be planar, i.e. they can be drawn on a page with dots for the nodes, and an arrow drawn from node μ to node ν to represent each arrow with source μ and target ν, in such a way that the arrows never cross each other or pass through nodes. But this planarity is a happy accident and not something we need to build into the definitions. Here is an example of a labelled digraph drawn on the page, from later in this book:

(2.3.1)



(2.3.1) is the same digraph as

(2.3.2)



But in practice we will adopt some conventions for placing the digraph on the page; see Definition 5.3.9 below. In this instance our conventions would rule out the diagram (2.3.2).

In the definitions that follow, sometimes we define a notion only for digraphs, but the notion applies equally well to labelled digraphs via their underlying abstract digraphs.

Definition 2.3.3 Let Γ and Δ be digraphs. By an *isomorphism of digraphs* from Γ to Δ we mean a pair (i, j) such that

(a) i is a bijection from the set of nodes of Γ to the set of nodes of Δ;

(b) j is a bijection from the set of arrows of Γ to the set of arrows of Δ;

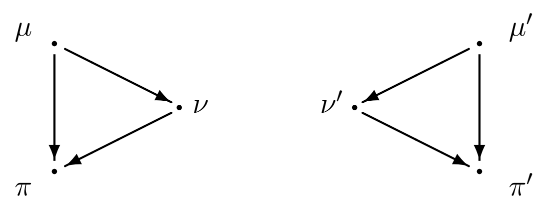
(c) if μ and ν are any nodes of Γ, then j takes the arrows with source μ and target ν to the arrows with source $i(\mu)$ and target $i(\nu)$.

We say that the digraph Γ is *isomorphic* to Δ if there is an isomorphism of digraphs from Γ to Δ.

We can verify that if (i, j) is an isomorphism of digraphs from Γ to Δ then (i^{-1}, j^{-1}) is an isomorphism of digraphs from Δ to Γ; so the relation of isomorphism between digraphs is symmetric. Similar arguments show that the relation is reflexive and transitive, so it is an equivalence relation on digraphs.

For example the two digraphs drawn below are isomorphic; take i so that $i(\mu) = \mu'$, $i(\nu) = \nu'$ and $i(\pi) = \pi'$, and then define j in the obvious way. (In these diagrams the symbols μ, μ' etc. are for purposes of identifying the nodes; the digraphs are not labelled digraphs.)

(2.3.3)



We are said to be counting digraphs *up to isomorphism* if we count two digraphs as the same digraph whenever they are isomorphic. Up to isomorphism, only one digraph appears in (2.3.3), though it is shown twice from different angles. We will call this digraph the *trinity digraph*, for a reason given in Exercise 1.1.1 on page 5 of [38].

Definition 2.3.4 Let Γ and Δ be labelled digraphs. By an *isomorphism of labelled digraphs* from Γ to Δ we mean a pair (i, j) such that (a), (b) and (c) of Definition 2.3.3 hold, together with two more clauses:

(d) if μ is any node of Γ, then the label on μ is the same as the label on $i(\mu)$;

(e) if a is any arrow of Γ, then the label on a is the same as the label on $j(a)$.

We say that Γ is *isomorphic* to Δ if there is an isomorphism of labelled digraphs from Γ to Δ. This is an equivalence relation on labelled digraphs.

Definition 2.3.5 Suppose Γ is a digraph. We can get another digraph by performing either of the following operations on Γ (provided at least one node is left at the end):

(a) Remove one node and all the arrows that have that node as either source or target.

(b) Remove one arrow, but keep all the nodes including the source and target of the arrow.

By a *subgraph* of Γ we mean a digraph that is either Γ itself or can be got from Γ by applying (a) or (b) any number of times.

Definition 2.3.6 A *path* in a digraph is a sequence

(2.3.4) $$\nu_1, a_1, \nu_2, \ldots, \nu_{m-1}, a_{m_1}, \nu_m$$

where the ν_i are distinct nodes except that ν_1 and ν_m can be the same node, and for each i $(1 \leqslant i < m)$, a_i is an arrow between ν_i and ν_{i+1}. In a picture:

where each horizontal line represents either a $\rightarrow$ or a $\leftarrow$. The *length* of the path is the number of arrows in it, i.e. $m - 1$. The path is *closed* if $\nu_1 = \nu_m$. If $\nu_1 \neq \nu_m$ the path is said to be *open*, and its *end nodes* are ν_1 and ν_m.

Definition 2.3.7 (a) If all of the arrows in the path (2.3.4) point in the same direction, so that the target of one arrow is the source of the next, we say that the path is *directed*, and we call it a *track*.

(b) Suppose (2.3.4) is a track. The *initial node* of the track is the node ν_1 that is not a target, and the *terminal node* is the node ν_n that is not a source; the track is *from* its initial node *to* its terminal node. If the track has positive length then its *initial arrow* is the arrow whose source is the initial node, and its *terminal arrow* is the arrow whose target is the terminal node. A node is a *successor* if it is not initial. If $i < n$, the *immediate successor* of ν_i is ν_{i+1}; we write ν^+ for the immediate successor of ν. Likewise ν_i is the *immediate predecessor* of ν_{i+1}, and we write ν^- for the immediate predecessor of ν.

(c) Suppose ν is one of the nodes of the track (2.3.4). By $\overleftarrow{\nu}$ we mean the initial segment of (2.3.4) that starts at ν_1 and finishes at ν, and by $\overrightarrow{\nu}$ we mean the final segment of (2.3.4) that starts at ν and finishes at ν_n.

Note that a path that is not a track doesn't have an intrinsic direction. We can give it a direction, but this will be imposed from outside (cf. Definition 2.3.15 below). When we do put a direction on a path, we will alter the vocabulary, and speak of 'first' and 'final' rather than 'initial' and 'terminal'. The expressions 'initial' and 'terminal' are special cases of more general notions defined from the digraph itself; cf. Definition 2.3.11(b) below.

Definition 2.3.8 We say that a digraph Γ is *connected* if for every pair μ, ν of distinct nodes of Γ there is an open path from μ to ν.

Lemma 2.3.9 *For every (abstract or labelled) digraph Γ there is a unique family of connected subgraphs $\{\Gamma_x : x \in X\}$ of Γ such that*

(a) every node of Γ is a node of Γ_x for exactly one x, and

(b) every arrow of Γ is an arrow of Γ_x for exactly one x.

Definition 2.3.10 Let Γ be a digraph (abstract or labelled). Then the subgraphs Γ_x of Lemma 2.3.9 are called the *connected components* of Γ.

Proof sketch. By Zorn's lemma every connected subgraph of Γ is contained in a unique maximal connected subgraph. Take $\{\Gamma_x : x \in X\}$ to be the family of all maximal connected subgraphs of Γ. □

Definition 2.3.11 (a) Each node of a digraph Γ has an *out-valency*, which is the number of distinct arrows that have it as source, and an *in-valency*, which is the number of distinct arrows that have it as target. The *valency* of a node is its out-valency plus its in-valency.

(b) Generalising Definition 2.3.7(b) above, we say that a node of a digraph is *initial* if its in-valency is 0, and *terminal* if its out-valency is 0.

For example in (2.3.3) above, the node μ has out-valency 2 and in-valency 0, while the node π has out-valency 0 and in-valency 2. The node ν has out-valency and in-valency both 2. So all the nodes have valency 2.

Lemma 2.3.12 *Let Γ be a finite digraph. Then the sum of the valencies of the nodes of Γ is an even number.*

Proof. Consider the set T of all triples (a, μ, i) such that a is an arrow of Γ, $i \in \{0, 1\}$, and if $i = 0$ then μ is the source of a, and if $i = 1$ then μ is the target of a. Then for each arrow a there are exactly two triples (a, μ, i) in T. (We included i to ensure this even when an arrow has the same node as its source and its target.) So T has an even number of elements. But also for each node μ the number of triples (a, μ, i) in T is equal to the valency of μ, so the number of elements of T is the sum of the valencies. □

Lemma 2.3.13 *Let Γ be a digraph with a finite number n of nodes, and let μ_1 be a node of Γ. Then the following are equivalent:*

(a) *Γ is connected.*

(b) *We can list the nodes of Γ as*

(2.3.5) $$\mu_1, \ldots, \mu_n$$

so that each μ_h ($1 < h \leqslant n$) is a neighbour of some μ_j with $j < h$.

Proof. (a) ⇒ (b): Assume (a) and list the nodes of Γ as $\nu_1, \ldots, \nu_n$ with $\nu_1 = \mu_1$. The nodes $\mu_2, \ldots, \mu_n$ are chosen inductively. When μ_h has been chosen, we take μ_{h+1} to be the first ν_i which is not among the nodes $\mu_1, \ldots, \mu_h$ but is a neighbour of one of them. The fact that Γ is connected shows that such a ν_i exists when $h < n$.

(b) ⇒ (a): Assume (b). By induction on h we can show that each μ_h is in the same connected component of Γ as μ_1. □

Lemma 2.3.14 *Let Γ be a loopless finite digraph with $n \geqslant 2$ nodes. Then the following are equivalent:*

(a) Γ is connected, all the nodes of Γ have valency $\leqslant 2$ and at least one node has valency 1.

(b) Γ has $n-1$ arrows, and we can list the nodes of Γ as

$$\mu_1, \ldots, \mu_n \tag{2.3.6}$$

so that whenever $1 \leqslant j < h \leqslant n$, μ_j and μ_h are neighbours if and only if $h = j+1$.

When (a) and (b) hold, there are exactly two nodes of Γ with valency 1.

Proof. (a) $\Rightarrow$ (b): Assume (a), and take the listing (2.3.6) as in (b) of Lemma 2.3.13 with μ_1 a node of valency 1. We show first that if $1 < j+1 < h$ then μ_j and μ_h are not neighbours. For otherwise take a counterexample with h as small as possible. Then μ_j and μ_h are neighbours. But since $j+1 < h$, the choices of h and (2.3.6) show that μ_j and μ_{j+1} are neighbours too. Since μ_j has valency $\leqslant 2$, μ_j is not also a neighbour of μ_i for any $i < j$. But then j must be 1 and so μ_j has valency 1, contradiction.

It follows that if $1 \leqslant j < n$ then μ_j and μ_{j+1} are neighbours. For by choice of (2.3.6), μ_h is a neighbour of some μ_i with $i < h$, and we have just shown that i is not $< j$. Moreover there is only one arrow between μ_j and μ_{j+1}; for otherwise, since μ_j has valency $\leqslant 2$, we must have $j = 1$, which implies that μ_j has valency 1. Hence there are exactly $n-1$ arrows.

(b) $\Rightarrow$ (a): Assume (b). Then for each h $(1 \leqslant h < n)$ there is an arrow between μ_h and μ_{h+1}. There are no other arrows, since Γ has $n-1$ arrows. We can read off that μ_1 and μ_n, which are distinct since $n \geqslant 2$, each have valency 1, and all other nodes have valency 2. □

Definition 2.3.15 A finite digraph Γ meeting either of the conditions (a), (b) of Lemma 2.3.13 is said to be *linear*. The two nodes with valency 1 form the end nodes of the open path, so we call them the *end nodes* of the digraph. Unlike a path, a linear digraph doesn't have a direction; in Definition 3.3.1 we will give a linear digraph a directiion by specifying which of the end nodes counts as first and which as last. A linear digraph with a direction is essentially the same thing as a path, cf. Definition 2.3.6 above.

Lemma 2.3.16 *Let Γ be a loopless digraph with a finite number $n \geqslant 2$ of nodes. Then the following are equivalent:*

(a) Γ *is connected and every node of* Γ *has valency 2.*

(b) Γ *has* n *arrows, and we can list the nodes of* Γ *as*

(2.3.7) $$\mu_1, \ldots, \mu_n$$

so that whenever $1 \leqslant j < h \leqslant n$, μ_j *and* μ_h *are neighbours if and only if either* $h = j + 1$, *or* $j = 1$ *and* $h = n$.

Proof. (a) ⇒ (b): Assume (a). Since Γ has at least 2 nodes and is connected, it has at least one arrow, say a with source μ and target ν. Let Γ' be Γ with a removed. We claim that Γ' is still connected. For otherwise Γ' has two connected components Δ_1 and Δ_2, one containing μ and ν respectively. The valencies of the nodes in Δ_1 are the same as they were in Γ, except that the valency of μ drops by 1. But since all the other nodes in Δ_1 have valency 2, this contradicts Lemma 2.3.12, proving the claim.

Now Γ' satisfies (a) of Lemma 2.3.14, and so Γ' has $n-1$ arrows and the nodes of Γ' can be listed as in (b) of that lemma. The nodes μ_1 and μ_n must be μ and ν in some order. So restoring the arrow a gives us (b) as required.

(b) ⇒ (a) is like the corresponding argument for Lemma 2.3.14 but simpler. □

Definition 2.3.17 A finite digraph Γ which has at least two nodes and meets the equivalent conditions (a), (b) of Lemma 2.3.16 is said to be *circular*. Observe that a circular digraph is essentially the same thing as a closed path of length at least 2. A digraph Γ is *acyclic* if it is loopless and has no circular subgraph. (Warning: some of the graph-theoretical literature defines a 'directed acyclic graph' to be a digraph with no closed track of length at least 1. This is a weaker notion than ours, because it allows the digraph to contain closed paths that are not tracks.)

Lemma 2.3.18 *Let* Γ *be a finite loopless connected graph with* n *nodes. Then the following are equivalent:*

(a) Γ *is acyclic.*

(b) Γ *has* $n-1$ *arrows, and we can list the nodes of* Γ *as*

(2.3.8) $$\mu_1, \ldots, \mu_n$$

so that for each h $(1 < h \leqslant n)$ *there is a unique arrow that is between* μ_h *and some node in* $\{\mu_1, \ldots, \mu_{h-1}\}$.

Proof. If $n = 1$ then both (a) and (b) are true. Thus Γ is acyclic because it is loopless and every circular graph has at least two nodes; also it has no arrows because it is loopless, and the listing $(2.3.8)$ consists of the unique node. So assume henceforth that $n \geqslant 2$.

(a) $\Rightarrow$ (b): Assume (a), and let the listing of nodes in (2.3.8) be as in (b) of Lemma 2.3.13. We claim that if $1 \leqslant j < h \leqslant n$ then Γ contains a path that has end nodes μ_j and μ_h and contains only nodes μ_i with $i \leqslant h$. This is left to the reader.

Suppose for contradiction that there is some h $(1 < h \leqslant n)$ such that μ_h has an arrow between it and a node μ_i, and a distinct arrow between it and a node $\mu_{i'}$, where both i and i' are $< h$. If $i = i'$ then the two arrows form a circular digraph with nodes μ_h and μ_i, contradicting (a). If $i \neq i'$ then there is a linear subgraph as in the claim, with end nodes μ_i and $\mu_{i'}$; the two arrows between these and μ_h join up with the linear subgraph to make a circular digraph, again contradicting (a). So the uniqueness in (b) is proved. Since Γ is loopless, each arrow of Γ is between a node and an earlier node in (2.3.8). We have just shown that there is a unique such arrow associated with the later node, so the total number of arrows is $n - 1$.

(b) $\Rightarrow$ (a): Assuming (b), suppose (a) is false and choose the earliest h such that Γ has a circular subgraph whose nodes all lie in $\mu_1, \ldots, \mu_h$. This subgraph contains two arrows between μ_h and earlier nodes in the list (2.3.8), contradicting (b). □

Introductory accounts of graph theory (and that's all we will need) are Wilson [101] and Chartrand and Zhang [18].

2.4 Exercises

All digraphs in these exercises are assumed to be finite.

2.1. Let T be a theory in a signature consisting of a single binary relation symbol A. Explain why it would be permissible to introduce a new relation symbol B defined as in (a), but neither (b) nor (c) will work as a definition of B:

(a) $\forall x \forall y (Bxy \leftrightarrow \exists z Axz)$.

(b) $\forall x (Bxx \leftrightarrow \exists z (Axz \wedge Azx))$.

(c) $\forall x \forall y \forall z (Bxy \leftrightarrow (Axz \wedge Azy))$.

Solution. (b) The definition doesn't tell us when Bxy holds with x and y standing for different elements.
(c) Suppose M is a Σ-structure with at least two elements 0 and 1, and Aab holds for elements a, b of M if and only if $a = b$. Assume the definition (c). Then in M, reading (c) from right to left and taking 0 for each of x, y, z, we have

$$(A00 \wedge A00) \to B00.$$

Reading (c) from left to right, and taking 0 for x, y and 1 for z we have

$$B00 \to (A01 \wedge A10).$$

But in M, $A00$ is true and $A01$ is false. So the definition has led to a contradiction.

2.2.

(a) Show that if T is a consistent finite theory whose skolemisation contains a constant but no function symbols, then T has a finite model. [For simplicity you can assume there is no $=$ in the language.]

(b) Write down a consistent finite theory with no function symbols but no finite model.

Solution. (a) Since T is finite, we can choose a finite signature Σ such that T is in $L(\Sigma)$. Then the construction of Σ^{sk} makes it finite too, and we can suppose that Σ^{sk} contains a constant but no function symbols. So the Herbrand universe of T is nonempty and finite, since it consists of just the finitely many constants in Σ^{sk}. The Herbrand theory Hr(T) will be a finite set of quantifier-free sentences. Since Hr(T) is consistent by REF, there is a model M of Hr(T). Now the truth or falsehood in M of any quantifier-free atomic sentence θ is unaffected if we remove from M all elements not named by constants in θ. So we can remove from M all elements not named by any of the finitely many constants in the Herbrand universe, and the result is a finite model of Hr(T). By THE PROOF OF ABOVE, this model is also a model of T.
(b) $\forall x \forall y \forall z (Bxy \wedge Byz \to Bxz)$, $\forall x \forall y (Bxy \to \neg Byx)$, $\forall x \exists y Bxy$.

2.3. Skolemise the following sentences.

(a) $\forall x \exists y (Bx \to Ay)$.

(b) $\forall x \exists y Bxy$.

(c) $\forall x(\forall y Bxy \to \exists z \forall w(Bzx \to Bzw))$.

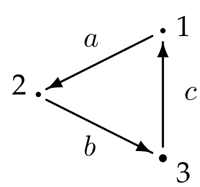
Which of them have skolemisations with only Skolem constants (as opposed to Skolem functions)? Solution. (a) $\forall x(Bx \to Af(x))$.
(b) $\forall x Bxg(x)$.
(c) $\forall x(Bxf_1(x) \to \forall w(Bf_2(x)x \to Bf_2(z)w))$.
In (a) we can first move the quantifiers inside: $(\exists x Bx \to \exists y Ay)$, and then move them out again but in the opposite order: $\exists y \forall x(Bx \to Ay)$. This allows the skolemisation $\forall x(Bx \to Ac)$. But this is impossible with (b). For suppose that Bxy means $x < y$; then in the natural numbers $\forall x \exists y Bxy$ is true, but there is no number c such that $\forall x Bxc$. A similar argument shows that we need Skolem functions for (c).

2.4. Let T consist of the sentence $\exists x \forall y(Cxy \leftrightarrow \neg Cyy)$. Show that T is inconsistent by showing that Hr(T) is inconsistent.

Solution. The simplest skolemisation is $\forall y(Ccy \leftrightarrow \neg Cyy)$ with the Skolem constant c. The Herbrand universe consists of just c, so Hr(T) has just the one sentence $(Ccc \leftrightarrow \neg Ccc)$. This sentence is clearly inconsistent.

2.5. Let Γ be the following digraph, where the nodes and edges have been labelled for convenience of reference.

(2.4.1)

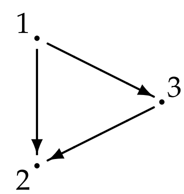


(a) Show that Γ has exactly seventeen distinct subgraphs. (b) Some of these subgraphs are isomorphic to each other; show that counting up to isomorphism, there are just seven distinct subgraphs of Γ.

Solution: (b) Any subset of $\{1, 2, 3, a, b, c\}$ describes a subgraph of Γ provided that it contains at least one node, and if it contains an arrow then it contains the source and target of that arrow. This allows the following possibilities. With three arrows: $\{1, 2, 3, a, b, c\}$. With two arrows: $\{1, 2, 3, a, b\}$, $\{1, 2, 3, b, c\}$, $\{1, 2, 3, a, c\}$. With one arrow: $\{1, 2, 3, a\}$, $\{1, 2, a\}$, $\{1, 2, 3, b\}$, $\{2, 3, b\}$, $\{1, 2, 3, c\}$, $\{1, 3, c\}$. With no arrows: $\{1, 2, 3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, $\{1\}$, $\{2\}$, $\{3\}$. (b) Eliminating subgraphs that are isomorphic to ones earlier in the list leaves the following: $\{1, 2, 3, a, b, c\}$, $\{1, 2, 3, a, b\}$, $\{1, 2, 3, a\}$, $\{1, 2, a\}$, $\{1, 2, 3\}$, $\{1, 2\}$, $\{1\}$.

2.6. A digraph is said to be *rigid* if there is exactly one isomorphism from the digraph to itself. Show that the trinity digraph is rigid. (This was part of Aquinas' argument in support of adding the *filioque* clause to the Creed; see *Summa Theologica* part 1 q. xxxvi art.2.)

Solution. Number the nodes of the trinity digraph Γ:



Let (i, j) be an isomorphism from Γ to Γ. The node 1 is the only node with in-valency 0, so $i(1) = 1$. The node 2 is the only node with out-valency 0, so $i(2) = 2$. This implies that $i(3) = 3$, and it forces j to be the identity as well.

2.7. Let Γ be the following digraph. The arrows are the words in the set W below, and the nodes are the consonants that occur in these words; the source of an arrow is the consonant at the beginning of the word, and the target of the arrow is the consonant at the end of the word.

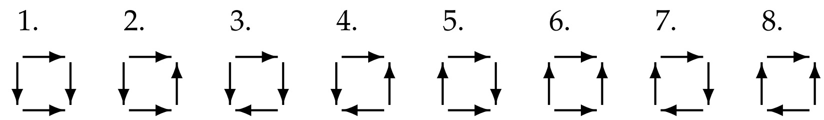
$$W = \{ \text{bib, cow, din, gas, men, peg, rot, rub, sip, tut, wok} \}$$

Show that Γ has four connected components, one of them circular, one of them a track, one of them linear but not a track, and one of them with two loops.

Solution: {peg, gas, sip} gives a circular component, {cow, wok} a track, {men, din} a linear digraph that is not a track, and {bib, rot, rub, tut} gives the fourth connected component with loops provided by bib and tut.

2.8. Show that, counting up to isomorphism, there are four circular digraphs of node size 4.

Solution. We could simply list them, but this wouldn't prove that there are no possibilities that we missed. A better argument: take any circular digraph of node size 4, choose one arrow and draw it at the top, pointing to the right. The remaining three arrows can each go in either of two directions, giving $2^3 = 8$ possibilities:



Now inspection shows that 1 and 6 are isomorphic; 2, 3, 5 and 8 are isomorphic; 4 and 7 are not isomorphic to any of the others; and these are all the isomorphisms.

2.9. Let Γ be a finite acyclic connected digraph, whose nodes are listed as $\mu_1, \ldots, \mu_n$ in such a way that for each h with $1 < h \leqslant n$, μ_h is a neighbour of some node in $\{\mu_1, \ldots, \mu_{h-1}\}$. Show that if $1 \leqslant j < h \leqslant n$ then there is a path that has end nodes μ_j and μ_h, and contains only nodes μ_i with $i \leqslant h$. (Cf. Lemma 2.3.18 above.)

Solution. Consider any h with $1 < h \leqslant n$. By a *ladder to* h we mean a sequence

$$\mu_{i_1}, \ldots, \mu_{i_k}$$

where $1 = i_1 < i_2 < \ldots < i_k = h$, and for each j ($1 \leqslant j < k$), μ_{i_j} and $\mu_{i_{j+1}}$ are neighbours.

We claim that if $1 \leqslant h \leqslant n$ then there is a ladder to h. If $h = 1$ then this is trivial, taking the sequence consising of just μ_1. If $1 < h$ it is proved by starting at μ_h, choosing some $\mu_i \in \{\mu_1, \ldots, \mu_{h-1}\}$ that is a neighbour of μ_h, and likewise some neighbour of μ_i in $\{\mu_1, \ldots, \mu_{i-1}\}$, and so on backwards until we reach μ_1.

Now if $1 \leqslant j < h \leqslant n$ then by the claim there are a ladder to j and a ladder to h. These two ladders have at least one node in common, namely μ_1. Let i be the greatest number such that μ_i is in both ladders. The required path is got by taking the ladder to h, running it backwards from μ_j as far as μ_i, and then running forwards from μ_i to μ_h along the ladder to h. □

Chapter 3

Peripatetic preliminaries

3.1 Subject and predicate

The definitions in this section are not traditional. They set up a modern framework that allows us to discuss several different parts of Ibn Sīnā's logic in parallel.

Definition 3.1.1 (a) By a *subject-predicate tag*, or for short a *tag*, we mean an ordered pair (B, A) of distinct relation symbols. (For this purpose we regard propositional symbols as relation symbols of arity 0, so that propositional logic can be included.) The symbols B and A are called respectively the *subject symbol* and the *predicate symbol* of the tag.

(b) By a *sentence form* we mean a function f whose domain is a class of tags, such that for each tag (B, A) in the domain, $f(B, A)$ is a sentence of a formal language (first-order unless we specify otherwise), satisfying the two conditions

If (B, A) and (B', A') are tags in the domain of f, then $f(B', A')$ is the sentence got from $f(B, A)$ by substituting B' for each occurrence of B and A' for each occurrence of A.

(In some logics the sentences $f(B, A)$ may contain other relation symbols besides B and A. These other relation symbols will always be reserved symbols that can't be used in tags.)

(c) By a *subject-predicate logic* $\mathcal{L}$ we mean a family of sentence forms, all with the same domain, such that if f and f' are sentence forms of $\mathcal{L}$

and (B,A), (B',A') are tags in their domain, then

If $f(B,A) = f'(B',A')$ then $f = f'$, $B = B'$ and $A = A'$.

This condition allows us to speak unambiguously of the *form* of the sentence $f(B,A)$ and the *subject symbol* and *predicate symbol* of this sentence.

(d) The *sentences* of $\mathcal{L}$ are the sentences $f(B,A)$ where f is any sentence form of $\mathcal{L}$ and (B,A) any tag in the domain of f. A *theory* of (or in) $\mathcal{L}$ is a set of sentences of $\mathcal{L}$. More generally a *formal subject-predicate sentence* is a sentence of some subject-predicate logic.

(e) We use Greek letters ϕ, ψ etc. to range over formal subject-predicate sentences, and letters A, B, C etc. for relation symbols. (See BELOW for the reason why they are often taken in inverse alphabetical order.) If a formal subject-predicate sentence is introduced as $\phi(B,A)$, this means that the sentence ϕ has subject symbol B and predicate symbol A. In this context, $\phi(D,C)$ means the same sentence with B replaced by D and A replaced by C.

(In Arabic 'subject' is *mawḍū*^{c}, 'predicate' is *maḥmūl* and 'sentence' is *qawl*.)

Note the requirement in (a) that the two relation symbols in a tag are distinct.

History 3.1.2 Aristotle used a phrase of the form 'A is predicated of all B' to express 'Every B is an A'. The Arabic translators preferred the phrasing with 'Every' at the front, but they preserved Aristotle's ordering of the letters. Hence the habit of taking the relation symbols in inverse alphabetical order.

History 3.1.3 The requirement in (a) in Definition 3.1.1, that B and A must be distinct relation symbols, is purely syntactic; nothing prevents two different relation symbols from being given the same interpretation. But for a Peripatetic writer like Ibn Sīnā, syntax is always an outward manifestation of underlying structured meanings. So does Ibn Sīnā intend that the subject and predicate of a sentence should have or be *different meanings*? At *Qiyās* [55] i.7, 66.4 he gives an example

> If there is movement then there is movement;
> but there is movement;
> so there is movement.

where the propositional predicate of the first sentence repeats the subject; he describes the example as 'ugly'. But in this example the subject and predicate are syntactically the same, so (as often in Ibn Sīnā) the example doesn't resolve what to us are the obvious questions. It seems that Ibn Sīnā has nothing approaching a workable way of individuating meanings; so we miss nothing essential if we regard requirement (a) as purely syntactic. But there might be more to be said about this.

Definition 3.1.4 An *instance* of a formal subject-predicate sentence ϕ is a natural-language sentence got by choosing texts for each of the relation symbols, and then translating the formal sentence into natural language, using the chosen texts to translate the relation symbols. The chosen texts are called the *terms* of the instance; we distinguish them as the *subject term* and the *predicate term* of the instance. ('Term' in Arabic is *ḥadd*, plural *ḥudūd*.)

There are a number of other properties that subject-predicate logics tend to have, though for simplicity we stop short of adding them to the definition. One such property is that each form has a contradictory negation; see (1) in Section 3.2 below. Another is that the sentences of the logic never use the symbol $=$.

It will be helpful to have a working example of a subject-predicate logic, as follows.

Definition 3.1.5 We write $\mathcal{L}_{uas}$ for the subject-predicate logic called *unaugmented assertoric logic*. This logic has four sentence forms:

$$(a, uas),\ (e, uas),\ (i, uas),\ (o, uas). \tag{3.1.1}$$

The domain of each of these forms is the class of pairs (B, A) of monadic relation symbols (cf. Definition 2.1.1 above). These sentence forms take tags (B, A) to sentences as follows:

$$
\begin{array}{rcl}
(a, uas)(B, A) & : & \forall x(Bx \to Ax) \\
(e, uas)(B, A) & : & \forall x(Bx \to \neg Ax) \\
(i, uas)(B, A) & : & \exists x(Bx \wedge Ax) \\
(o, uas)(B, A) & : & \exists x(Bx \wedge \neg Ax)
\end{array}
$$

Figure 3.1: Unaugmented assertoric sentences

(See History 5.1.3 below for the origin of the labels a, e, i, o, and Definition 5.1.2(c) for the name 'unaugmented'.)

History 3.1.6 Unaugmented assertoric logic is not one of the logics that Ibn Sīnā studies. Nor does it support the procedures of Aristotle's assertoric syllogistic (for example we don't have '(a)-conversion' $(a, uas)(B, A) \vdash (i, uas)(A, B)$, cf. Definition 3.3.9(c)). It appears very briefly at p. 24 in Frege's *Begriffsschrift* [29], and Frege has been taken to task for misrepresenting Aristotle by mentioning it. But presumably Frege was indicating correlations between his own formalism and sentences studied by Aristotelian logicians of his own time, not offering a scholarly interpretation of Aristotle.

Definition 3.1.7 Let $\mathcal{L}$ be a subject-predicate logic and T a theory in $\mathcal{L}$. The *subject-predicate digraph* of T, $\Gamma(T)$, is a labelled digraph (cf. Definition 2.3.2) such that

(a) The nodes of $\Gamma(T)$ are in 1–1 correspondence with the set of relation symbols occurring as subject or predicate symbols of sentences of T; each node is labelled with the corresponding relation symbol.

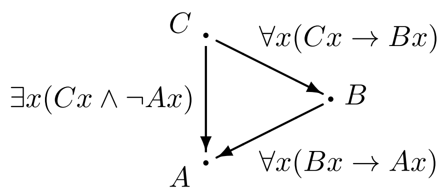
(b) The arrows of $\Gamma(T)$ with source labelled B and target labelled A are in 1–1 correspondence with the set of sentences of T with subject symbol B and predicate symbol A; each arrow is labelled with the corresponding sentence.

Example 3.1.8 Here is a theory in $\mathcal{L}_{uas}$:

(3.1.2) $$\forall x(Cx \to Bx),\ \forall x(Bx \to Ax),\ \exists x(Cx \wedge \neg Ax)$$

The subject-predicate digraph of this theory is

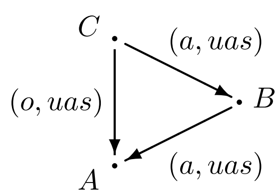
(3.1.3)



Definition 5.3.9 below will describe some conventions for writing subject-predicate digraphs.

There is some redundancy in the diagram (3.1.3): the terms are given twice, at the nodes and in the sentences. There would be no loss of information if we replaced the sentences by their sentence forms:

(3.1.4)



Definition 3.1.9 Let $\mathcal{L}$ be a subject-predicate logic, T a theory in $\mathcal{L}$ and Γ the subject-predicate digraph of T. Then we say that T is *graph-circular* if Γ is circular (cf. Definition 2.3.17); similarly *graph-linear* (cf. Definition 2.3.15), *graph-acyclic* etc.

Thus for example the theory (3.1.2) is graph-circular.

3.2 Negation

Negation in Peripatetic logic is more complicated than in most modern logics. Ibn Sīnā used five distinct forms of negation, and not all of them are straightforward to formalise.

(1) **Contradictory negation**

Definition 3.2.1 Let $\mathcal{L}$ be a subject-predicate logic. We say that $\mathcal{L}$ *has contradictory negations* if for every sentence form f of $\mathcal{L}$ there is a unique sentence form g of $\mathcal{L}$ such that

For each tag (R, S), $g(R, S)$ is logically equivalent to $\neg f(R, S)$.

We call g the *contradictory negation* of f, and $g(R, S)$ the *contradictory negation* of $f(R, S)$; we write $g = \overline{f}$ and $g(R, S) = \overline{f(R, S)}$. (The Arabic for 'contradictory negation' is *naqīḍ*.)

Lemma 3.2.2 $\overline{\overline{\phi}} = \phi$*, and if* $\psi = \overline{\phi}$ *then* $\phi = \overline{\psi}$*.* □

When Ibn Sīnā introduces a logic, he first sets out the logical relations between single sentences. We will follow this practice. After setting out the sentence forms and their meanings, we will determine, for any sentence forms f and g, where there are entailments between

$$\begin{array}{cccc} f(R,S), & \overline{f(R,S)}, & f(S,R), & \overline{f(S,R)}, \\ g(R,S), & \overline{g(R,S)}, & g(S,R), & \overline{g(S,R)}, \end{array} \tag{3.2.1}$$

including the case where $f = g$. For a logic like core 2D logic, which has sixteen sentence forms, the task is not negligible.

Example 3.2.3 In unaugmented assertoric logic $\mathcal{L}_{uas}$ the sentence forms (a, uas) and (o, uas) are the contradictory negations of each other, and the sentence forms (e, uas) and (i, uas) are the contradictory negations of each other.

(2) **Simple negation**

For most subject-predicate logics the sentence forms come in pairs; in each pair one form is counted as an *affirmation* (*ījāb*) or as *affirmative* (*mūjib*), and the other form is counted as a *denial* (*salb*) or as *negative* (*sālib*); the second is called a *simple negative* (*sālib basīṭ*, e.g. ^{c}*Ibāra* 78.10). All of this terminology carries over also to the resulting sentences and propositions. In most such pairs the main difference between the affirmative sentence and its negative twin is that the negative sentence has a negation $\neg$ in front of the predicate symbol, though there are usually other differences too. In

PROPOSITIONAL LOGIC PL3 BELOW the negative sentence has its negation added to the subject symbol of the affirmative, not the predicate symbol.

Example 3.2.4 In unaugmented assertoric logic $\mathcal{L}_{uas}$ the forms (a, uas) and (e, uas) make an affirmative/negative pair; likewise (i, uas) and (o, uas). This example is unusual in that the simple negative is exactly the affirmative with the predicate symbol negated.

(3) **Metathesis**

The Arabic tradition retained a notion that was not always present in western Peripatetic logic, namely *metathesis* (*ᶜudūl*, *ᶜIbāra* [54] 82.4, 92.2f). In metathesis a negation occurs, but it is reckoned to be a part of the term where it occurs, not a part of the form. So for example an affirmative sentence ϕ stays affirmative if a metathetic negation is put on the predicate symbol; the resulting sentence need not be logically equivalent to the simple negation of ϕ.

Metathesis is a challenge to our notion of a subject-predicate logic. One possible way to incorporate it is to allow tags to contain not just relation symbols but also negated relation symbols, as for example $(B, \neg A)$. But this is only a cosmetic patch for a major problem, which is that the wholesale use of metathesis is tantamount to allowing negation to occur wherever you want, and this makes large parts of the Peripatetic view of logic pointless. We will face this problem in Chapter ON METATHETIC LOGIC, with Ibn Sīnā's most advanced form of propositional logic.

(4) **Sentence negation**

Writing in Arabic, Ibn Sīnā could negate a sentence by putting 'It is not the case that' (*laysa*) at the beginning of it. This was a gift from Arabic, and not a feature of Peripatetic logic in general. Though Ibn Sīnā uses this device constantly, he tends to use it in ways that don't really increase the expressive power of his logic. So for example he sometimes expresses a contradictory negation $\overline{f}$ as '*laysa* f'. He also uses this sentence negation as a computational device for calculating the contradictory negations of complicated sentence forms: he begins by putting *laysa* at the front and then moves the negation inwards by de-Morgan-like rules.

History 3.2.5 Even in Arabic this negation has to be handled with care. For example '*laysa* some As are Bs' means 'Some As are not Bs' and not 'It is not the case that some As are Bs'; this results from a freak interaction between $laysa$ and *ba*^{c}*ḍ* 'some', which Ibn Sīnā notes at ^{c}*Ibāra* [54] 54.6f, *Qiyās* [55] 37.10f. But *laysa kull* does mean 'Not every' (with *kull* = 'every'), and in several places Ibn Sīnā accepts *laysa laysa kull* as equivalent to just *kull*. See [44] for examples and discussion.

(5) **Privation**

Ibn Sīnā also recognises a form of lexical negation known as *privation* (Arabic ^{c}*adam*); for example the privation of 'seeing' is 'blind', whereas its metathetic negation is 'not seeing'. Privation doesn't show up in formal logic, so we say no more about it.

3.3 Productivity and conclusions

The notion of 'theory' in Definition 3.1.1(d) above is from modern logic, not from Peripatetic logic. In fact for Ibn Sīnā the premises of an inference are not a set but a structure. The structure consists of a graph-linear set of sentences, together with a direction for the linear ordering. Likewise the conclusion of an inference is not just any sentence that follows from the premises, but a sentence related to the structure of the premises in a specific way. Here are modern versions of Ibn Sīnā's definitions.

Definition 3.3.1 Let $\mathcal{L}$ be a subject-predicate logic.

(a) By a *premise-sequence* in $\mathcal{L}$ we mean a graph-linear theory T in $\mathcal{L}$ together with an ordered pair $[B, A]$ of distinct relation symbols such that B and A are the node labels of the two end nodes of the subject-predicate graph $\Gamma(T)$ of T; we call $[B, A]$ a *direction* of T. We write the premise-sequence as $T[B, A]$. (Ibn Sīnā calls a premise-sequence either *ta'līf* (literally 'composition') or *qarīna* (literally 'linkage'). 'Premise' is *muqaddama*.)

(b) The *length* of the premise-sequence $T[B, A]$ is the number of sentences in T. When the length is 2 we also call $T[B, A]$ a *premise-pair*.

(c) The symbols B and A are called the *extremes* of $T[B, A]$; B is the *minor extreme* and A is the *major extreme*. The relation symbols of T that are not either B or A are called the *middle symbols* of T. (In Arabic 'extreme' is *ṭaraf*; 'middle' is *'awsaṭ*. 'Minor extreme' and 'major extreme' should be *al-ṭaraf al-'aṣḡar* and *al-ṭaraf al-'akbar* respectively, though in practice Ibn Sīnā uses less direct descriptions in place of these.)

(d) The unique sentence of T containing the minor extreme is called the *minor premise*, and the unique sentence of T containing the major extreme is called the *major premise*. (In Arabic 'minor premise' is *al-muqaddamat al-ṣuḡrā* and 'major premise' is *al-muqaddamat al-kubrā*.)

(e) We write the premise-sequence $T[B, A]$ as $<\phi_1, \ldots, \phi_n>$ $[B, A]$ where ϕ_1 is the minor premise, ϕ_n is the major premise, and the sentences $\phi_1, \ldots, \phi_n$ are the sentences of T in the order in which they appear in the subject-predicate graph $\Gamma(T)$. Note that if $n \geq 2$ then the direction $[B, A]$ can be recovered from $<\phi_1, \ldots, \phi_n>$ and so can be omitted in the notation.

If ϕ is a sentence whose relation symbols are B and A, note the difference between the sentence $\phi(B, A)$ and the theory $\{\phi\}[B, A]$. The notation $\phi(B, A)$ means that ϕ has subject symbol B and predicate symbol A. The notation $\{\phi\}[B, A]$ means that the one-sentence theory $\{\phi\}$ is taken to have B as its first extreme and A as its last extreme, regardless of which way round they are in ϕ. The next definition will show these two notions interacting.

With the aid of Definition 3.3.1 and the notions of 'stronger' and 'weaker' from Definition 2.1.21, we can define the two notions that Ibn Sīnā picks out as the main contributions of logic to science in general, as follows.

Definition 3.3.2 Let $\mathcal{L}$ be a subject-predicate logic and $T[B, A]$ a premise-sequence in $\mathcal{L}$.

(a) We say that $T[B, A]$ is *productive* (Arabic *muntij*) if there is a sentence $\chi(B, A)$ of $\mathcal{L}$ such that $T[B, A] \vdash \chi(B, A)$. If $T[B, A]$ is not productive, then it is said to be *sterile* (Arabic *ᶜaqīm*).

(b) Suppose $\chi(B, A)$ is a sentence of $\mathcal{L}$. Then we say that $T[B, A]$ *yields* (Arabic *yuntiju*) $\chi(B, A)$, and that $\chi(B, A)$ is a *conclusion* (Arabic *natīja*) of $T[B, A]$, in symbols

$$T[B, A] \rhd \chi(B, A),$$

if:

(i) $T[B, A] \vdash \chi(B, A)$ (so that $T[B, A]$ is productive), and

(ii) if $\theta(B,A)$ is any sentence of $\mathcal{L}$ which is stronger than $\chi(B,A)$, then $T[B,A] \nvdash \theta(B,A)$.

In fact there is usually a single strongest χ as in (i), (ii) above, so that we can speak of 'the conclusion' of the premise-sequence. See Exercise 9.3 for a premise-pair in two-dimensional logic with two incomparable strongest conclusions.

Example 3.3.3 In unaugmented assertoric logic $\mathcal{L}_{uas}$, the theory of Example 3.1.8 above is inconsistent, so any two of its sentences entail the contradictory negation of the third. In particular

(3.3.1) $$\exists x(Cx \wedge \neg Ax),\ \forall x(Bx \to Ax) \ \vdash\ \exists x(Cx \wedge \neg Bx).$$

Now

(3.3.2) $$\exists x(Cx \wedge \neg Ax),\ \forall x(Bx \to Ax)$$

forms a premise-pair with minor extreme C and major extreme B, call it $U[C,B]$. By (3.3.1) this premise-pair entails $\exists x(Cx \wedge \neg Bx)$, which is $(o,uas)(C,B)$. So $U[C,B]$ is productive, yielding the conclusion $\exists x(Cx \wedge \neg Bx)$. The premise-pair $U[B,C]$ has the same two premises as $U[C,B]$ but in the opposite order:

(3.3.3) $$\forall x(Bx \to Ax),\ \exists x(Cx \wedge \neg Ax).$$

By Definition 3.3.2 a conclusion of $U[B,C]$ would have to have subject symbol B and predicate symbol C. There is no such conclusion in $\mathcal{L}_{uas}$, and so $U[B,C]$ is sterile. (Ibn Sīnā is explicit about this; $\exists x(Cx \wedge \neg Bx)$ can't count as a conclusion of $U[B,C]$; *Qiyās* ii.4, 110.9–111.1.) So a productive premise-sequence can become sterile if we reverse the direction. See also Lemma 7.2.2 on goclenian sequences, which are not sterile but have what seems the wrong conclusion.

Definition 3.3.4 Let $<\phi_1,\ldots,\phi_n>$ $[A_0,A_n]$ be a premise-sequence, and list as $A_0,\ldots,A_n$ the relation symbols of this sequence, in the order in which they occur in the subject-predicate digraph, beginning with the minor symbol A_0.

(a) Each sentence ϕ_i has relation symbols A_{i-1} and A_i. We say that ϕ_i is *forwards* in the premise-sequence if ϕ_i has subject symbol A_{i-1} and predicate symbol A_i, and that ϕ_i is *backwards* in the premise-sequence otherwise.

(b) If A_i is a middle symbol, we say that A_i is a *switchpoint* if exactly one of ϕ_i and ϕ_{i+1} is forwards. If A_i is an extreme, we say that A_i is a *switchpoint* if the sentence containing A_i is forwards.

(c) If $<\psi_1, \ldots, \psi_m> [B_0, B_m]$ is also a premise-sequence, we say that the two premise-sequences have *the same figure* if $m = n$ and for each i with $1 \leqslant i \leqslant n$, ϕ_i is forwards in $<\phi_1, \ldots, \phi_n> [A_0, A_n]$ if and only if ψ_i is forwards in $<\psi_1, \ldots, \psi_m> [B_0, B_m]$. Clearly this is an equivalence relation on premise-sequences; the equivalence classes of this relation are called *figures*. (In Arabic 'figure' is *šakl*.)

(d) We call a premise-sequence and its figure *prograde* if all its sentences are forwards; we call a premise-sequence and its figure *retrograde* if all its sentences are backwards.

For example if $T[B, A]$ has length two, then there are four possible figures that it might have, listing from left to right (so that $\rightarrow$ is forwards and $\leftarrow$ is backwards):

(3.3.4)

$$\begin{array}{l} (1)\ \rightarrow \rightarrow. \\ (2)\ \rightarrow \leftarrow. \\ (3)\ \leftarrow \rightarrow. \\ (4)\ \leftarrow \leftarrow. \end{array}$$

In (1) there are switchpoints at the two ends; in (2) there are switchpoints at the left and in the middle; in (3) there are switchpoints at the right and in the middle; in (4) there are no switchpoints. (1) is prograde and (4) is retrograde.

Definition 3.3.5 The four figures in (3.3.4) are known respectively as the *first figure*, the *second figure*, the *third figure* and the *fourth figure*.

Definition 3.3.6 Let $T[B, A]$ and $U[D, C]$ be premise-sequences. We say that $T[B, A]$ and $U[D, C]$ *have the same mood* if there is a bijection from the subject and predicate symbols used in $T[B, A]$ to those used in $U[D, C]$, which translates $T[B, A]$ into $U[R, S]$. The relation of having the same mood is clearly an equivalence relation between premise-sequences; an equivalence class of this relation is called a *mood* (Arabic *ḍarb*).

Two premise-sequences with the same mood must have the same figure too; so each mood has a figure, and the figures can be subdivided into their moods. In practice the term 'mood' is often restricted to moods of productive premise-pairs.

Lemma 3.3.7 *Let $T[B, A]$ and $U[D, C]$ be premise-sequences in the same mood. If $T[B, A]$ is productive then so is $U[D, C]$. Moreover if $\chi(B, A)$ is a conclusion of $T[B, A]$ then $\chi(D, C)$ is a conclusion of $U[D, C]$.*

Proof. Suppose $T[B, A]$ is productive with conclusion $\chi(B, A)$. Since $T[B, A]$ and $U[D, C]$ have the same mood, there is some permutation π of the set of subject or predicate symbols in the sentences involved which takes $T[B, A]$ to $U[D, C]$; then π takes $\chi(B, A)$ to $\chi(\pi B, \pi A)$, which is $\chi(D, C)$. There may be other relation symbols beside B, A, D, C in these premise-sequences and sentences, but by Definition 3.1.1 these other relation symbols are not changed when π is applied. It follows that $U[D, C], \chi(D, C)$ comes from $T[B, A], \chi(B, A)$ by a permutation of the relation symbols, and hence any inference relations that hold or fail between sentences of $T[B, A]$, $\chi(B, A)$ also hold or fail to hold between the corresponding sentences of $U[D, C], \chi(B, A)$. □

Definition 3.3.8 We define an *instance* of a premise-sequence $T[B, A]$ in the same way as an instance of a single formal subject-predicate sentence (Definition 3.1.4 above), but using text for each relation symbol occurring as subject or predicate symbol of any sentence in $T[B, A]$; the same text is used for each occurrence of the same relation symbol. As in Definition 3.1.4, the texts are called *terms*; the term replacing the minor extreme is the *minor term*, that replacing the major extreme is the *major term*, and the remaining texts are *middle terms*. (In Arabic 'term' is *ḥadd*, 'minor term' is *al-ḥadd al-'aṣġar*, 'major term' is *al-ḥadd al-'akbar* and 'middle term' is *al-ḥadd al-'awsaṭ*.)

Productive premise-sequences of length 1 give us single-premise inferences. One case is worth noting here.

Definition 3.3.9 Let $<\phi(B, A)>\ [A, B]$ be a retrograde premise-sequence of length 1.

(a) If this premise-sequence is productive, we say that it, and the sentence $\phi(B, A)$, are *convertible* (Arabic *mun^{c}akis*), and its conclusion

$\chi(A, B)$ is a *converse* (Arabic ^{c}*aks*) of $\phi(B, A)$; if (as often) this is the unique conclusion, we call it *the converse* of $\phi(B, A)$.

(b) If $\phi(B, A)$ and $\chi(A, B)$ are each a converse of the other, we say that the premise-sequence in (a) above *converts symmetrically* (Arabic *yan^{c}akisu miṯla nafsih*). Warning: Ibn Sīnā (and not he alone) sometimes describes a sentence as 'convertible' meaning that it converts symmetrically.

(c) The process of passing from $\phi(B, A)$ to its converse is called *conversion* (also ^{c}*aks* in Arabic). When ϕ is a sentence of the form (f), we speak of (f)-*conversion*.

3.4 Exercises

3.1. Suppose T is a theory in $\mathcal{L}_{uas}$, and the Σ-structure M is a model of T. Let N be the Σ-structure got from M by adding one or more new elements to the domain of M, and requiring that $A^N = A^M$ for every relation symbol A in Σ. Show that N is again a model of T. (We will see that the analogous result fails for metathetic logic.)

Solution. From their meanings, the truth-value of any of the sentences $(a, uas)(B, A)$, $(e, uas)(B, A)$, $(i, uas)(B, A)$ and $(o, uas)(B, A)$ is never altered just by adding new elements that are not in B. (Or for a rigorous proof one can use the truth definition.)

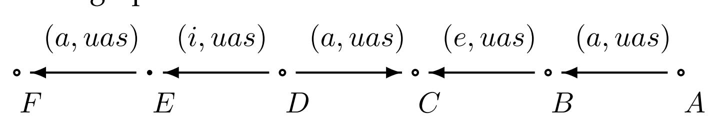
3.2. Draw the subject-predicate digraph of the following premise-sequence in unaugmented assertoric logic, with direction $[F, A]$:

$$\forall x(Ex \to Fx),\ \exists x(Dx \wedge Ex), \forall x(Dx \to Cx), \forall x(Bx \to \neg Cx), \forall x(Ax \to Bx).$$

Which nodes are switchpoints? Is the premise-sequence productive? (An informal argument will suffice.)

Solution. The digraph:

(3.4.1)



The switchpoints are D and C. The premise-sequence is productive and has conclusion $\exists x(Fx \wedge \neg Ax)$, i.e. $(o, uas)(F, A)$. We can deduce this conclusion as follows. By the second sentence there is a such that Da and Ea. By the first sentence, Fa. By the remaining sentences read from left to right, Ca, so $\neg Ba$, so $\neg Aa$. Writing $T[F, A]$ for the premise-sequence, this proves $T[F, A] \vdash (o, uas)(F, A)$. To improve the $\vdash$ to $\rhd$ we need to show that T entails no other sentence $f(F, A)$. Let M be any model of T. By the argument above, F^M is not empty. But M remains a model if we take B^M and A^M to be empty, and this shows that neither $(i, uas)(F, A)$ nor $(a, uas)(F, A)$ is deducible from T. Alternatively we can add a new element b to F^M, A^M and B^M and still have a model of T, showing that $(e, uas)(F, A)$ is not deducible either.

3.3. We consider premise-sequences $T[B, A]$ of length n.

(a) Show that the number of switchpoints of $T[B, A]$ is always an even number.

By Corollary 5.3.5(b) and BELOW, in both assertoric logic and core two-dimensional logic the figures containing productive premise-sequences are exactly those with either 0 or 2 switchpoints. Show:

(b) The total number of figures of length n is 2^n.

(c) The number of figures of length n with either 0 or 2 switchpoints is $(n^2 + n + 2)/2$.

Solution: (a) Consider the internal switchpoints, i.e. those that are not first or last term. If the number of internal switchpoints is even, then the first and last sentences are both forwards (adding 2 switchpoints) or both backwards (adding 0 switchpoints). If the number of internal switchpoints is odd, then exactly one of the first and last sentences is forwards, adding 1 switchpoint.

(b) Each of the n sentences is either forwards or backwards.

(c) The number of figures of length n with $2k$ switchpoints is $\begin{pmatrix} n+1 \\ 2k \end{pmatrix}$. When $k = 0$ this is 1; when $k = 1$ it is $(n^2 + n)/2$. [For these figures see Meredith's formulas in Lukasiewicz [73] p. 42.]

3.4. Let $\mathcal{L}$ be the logic $\mathcal{L}_{uas}$.

(a) Show that if $T[B, A]$ is a premise-sequence of $\mathcal{L}_{uas}$ in a first-order language $L(\Sigma)$, then some Σ-structure M is a model of T with exactly two elements, such that for each relation symbol R in Σ, R^M has cardinality 1. [Go by induction on the length of $T[B, A]$.]

(b) Show that for every positive natural number n there is a set of sentences of $\mathcal{L}$ which has a model of cardinality n but no model of cardinality $< n$.

History 3.4.1 3.4(a) shows that the logic $\mathcal{L}_{uas}$ doesn't have the the resources to allow us to deduce a conclusion from a formally inconsistent set of premises. The same willl apply to other logics BELOW. This agrees with Ibn Sīnā's own analysis (in Chapter BELOW) that in order to formalise reductio ad absurdum we need to move to a more complicated kind of language.

3.5. Let T be a subject-predicate theory in a language $L(\Sigma)$, such that for every sentence ϕ in T, (i) $=$ doesn't occur in ϕ and (ii) the only symbols of Σ that occur in ϕ are the subject and predicate symbols. Suppose every connected component of T has a model. Show that T has a model.

Solution: Let the distinct connected components of T be $\{T_i : i \in T\}$, and for each $i \in I$ let Σ_i be the set of symbols of Σ that occur in T_i. Then if $i \neq j$, T_i and T_j have no overlap in their subject and predicate symbols, and so by (ii) the signatures Σ_i are pairwise disjoint. Also by (ii) we can assume that Σ is the union of the signatures Σ_i and it contains no function symbols. Now each T_i has a model M_i, and we can take M_i to be a Σ_i-structure. By Lemma 2.1.16 and (i) we can assume that all the M_i have domains of the same cardinality, and hence also that all the M_i have the same domain X. Form the Σ-structure M with domain X by making its restriction to Σ_i equal to M_i. Then M is a model of each T_i and hence of T.

3.6. Let $T[D, C]$ be a productive premise-sequence of length n with conclusion χ.. Show that the following are equivalent:

(a) $T[D, C]$ has no switchpoints.

(b) $T[D, C]$ is retrograde.

and that if $n > 1$ then (a) entails

(c) The automorphism group of the underlying abstract digraph of the subject-predicate digraph of $T \cup \overline{\chi}$ is transitive on arrows.

Show that the implication $(a) \Rightarrow (c)$ fails if $n = 1$.

Solution. (a) $\Rightarrow$ (b): Assume $T[D, C]$ has no switchpoints. Then all the sentences of $T[D, C]$ have the same direction. Since the conclusion has direction $[D, C]$, all the sentences of $T[D, C]$ must be backwards, i.e. (b).

(b) $\Rightarrow$ (a), (c): If $T[D, C]$ is retrograde then the underlying abstract digraph is the circular digraph with $n + 1$ arrows, and all the arrows point anticlockwise. So there are no switchpoints. Also the automorphism grop of this digraph is the cyclic group of order $n+1$ generated by a turn taking each arrow to its immediate successor.

(c) $\Rightarrow$ (a) if $n > 1$: Assuming (c), the automorphism group of the underlying abstract digraph Γ is a cyclic subgroup of the dihedral group D_{n+1}. Since the group is transitive it has order $n + 1 > 2$, so the group is generated by a turn and not by a back-front involution. Hence all the arrows are anticlockwise and there are no switchpoints.

If $n = 1$ and $T[D, C]$ is prograde, then the digraph consists of two parallel arrows, and the automorphism group is transitive, switching the two arrows.

Chapter 4

Avicennan preliminaries

4.1 Questions and methods

When Ibn Sīnā has set up a subject-predicate logic and defined its basic notions, what questions does he reckon he needs to answer about the logic? And given those questions, what should we as mathematical logicians aim to prove about the logic?

This section is largely a checklist of requirements, but it gives us a convenient place to describe some methods for meeting those requirements.

Ibn Sīnā typically raises and attempts to answer the following questions about a logic $\mathcal{L}$.

Question One What are the logical relations between single sentences?

Question Two What are the productive premise-pairs, and for each such premise-pair, what conclusion or conclusions does it have?

Question Three If a premise-pair has a given conclusion, how can we come to know this fact?

Question Four If a premise-pair is sterile, how can we come to know this fact?

Question Five How can we determine, given a premise-pair, whether or not it is productive?

Question Six How can we tell, given a productive premise-pair, what conclusion it has?

Questions Two to Six can also be asked about premise-sequences of length greater than two. Ibn Sīnā treats these longer premise-sequences as a part of *analysis* (Arabic *taḥlīl*, i.e. the art of applying formal logic to natural language arguments). We treat it more fully in [47], but Chapter 8 below will prove some of the mathematical preliminaries.

It's worth noting that Questions Five and Six are not covered by Questions Three and Four; Ibn Sīnā gives them different answers and in different places. Questions Three and Four ask for methods of proof, while Questions Five and Six ask for methods of effective calculation. In the West the move towards effective methods of calculation in logic has sometimes been credited to the Catalan scholar Ramon Llull (for example by Leibniz REF). I thank Hans Daiber for pointing out that Llull was one of the few Scholastics who could read Arabic, and that we know he read Ibn Sīnā. There is as yet no evidence that Llull read those parts of Ibn Sīnā's logic that raise issues of effective calculation, but this is a historical question worth leaving on the table.

For **Question One** Ibn Sīnā has no special methods. He seems to reckon that the facts will be self-evident and can simply be listed. As a result he does make mistakes in a few of the more complex cases; see Section 11.4 below.

For **Question Two** a prior question is how to list the premise-pairs. There will be finitely many of them up to choice of tag, i.e. up to mood. Ibn Sīnā follows Aristotle's lead and lists the cases by figure, and then by mood within the figures. In some cases (e.g. propositional logic REF) the listing of moods is quite complex, but Ibn Sīnā is dauntlessly systematic.

The question of how to make effective lists had been raised already in the eighth century by Al-Ḳalīl in connection with the design of dictionaries; see Dichy [21]. We can see from Ibn Sīnā's listings that he took the question seriously. In his description of his proof search algorithm he raises a question about listings and says that he will answer it in the Appendices to *Qiyās*; sadly these Appendices haven't come down to us. We don't know how far Ibn Sīnā was aware of earlier work on listing.

For Question Two itself, Ibn Sīnā seems to have used bare-hands calculation—what he sometimes calls *istiḵrāj*, 'working it out'. Since all his logics can be expressed within decidable fragments of first-order logic, we could get the same information by running the list of all moods through a suitable computer program. But in this book we aim to find an explicit syntactic description of the productive premise-pairs and their conclusions, and to prove its correctness. So our answer to Question Two will also yield answers to Questions Five and Six.

There is an issue, easily missed, about where to look for conclusions. A modern logician, working in first-order logic, would expect to be looking for first-order conclusions—even if stronger second-order conclusions can be drawn from the same first-order premises. We can't transfer this kind of expectation straightforwardly to Ibn Sīnā, since he doesn't have our modern notion of a 'logic'. A fortiori he doesn't have our notion of a fragment; see Definition 9.4.2(b) below. But he does have a notion that does some of the same work. We can state it as a principle:

Definition 4.1.1 The *Genetic Principle* states that for each productive two-premise inference, each feature of the conclusion is inherited from at least one of the two premises. A premise is said to have the *ᶜibra* ('dominance', maybe) for a feature if it passes on this feature to the conclusion.

For example the conclusion will be negative only if one of the premises is negative; so a negative premise always has the dominance for quality. Ibn Sīnā does recognise occasional exceptions; for example in *Darapti* the conclusion is existential but neither of the premises is existential. But in 2D logic, if both premises are in the *dt* fragment then Ibn Sīnā will look for a conclusion that is also in this fragment. I believe he does this because of the Genetic Principle, and not because he thinks in terms of a *dt* logic. But there may well be more to be said about this. (It is discussed in [45].)

Ibn Sīnā sees **Question Three** as asking for a proof calculus. But the calculus is not in the first instance a device for establishing that certain things are the case. Rather it's a framework for describing thought processes that lead us from the premises to the conclusion. So for example Ibn Sīnā generally avoids calculations that involve reducing one logic to another one, presumably because these calculations belong more to mathematics than to everyday thinking. He always starts from the proof procedures of Aristotle's *Prior Analytics* i 5f, which he takes as a paradigm. Only when he can't find a way of adjusting Aristotle's proofs does he do something new. This does happen in several places in the *dt* fragment of his two-dimensional logic. The facts here are mathematical, and we report them in Section 10.4.

For his propositional logic PL3 he adopts a completely different approach, and gives proofs that involve a systematic reduction to assertoric logic. He sometimes speaks as if this change of approach was inevitable:

(4.1.1) It has been explained from the facts of the case that a propositional syllogism becomes informative through being completed by recombinant syllogisms. (*Qiyās* [55] ix.1, 415.9)

This approach gets him the right results, but I bet that very few of his readers have ever recognised it as a description of their own thought processes. This misfit is one of the features of PL3 that makes me suspect it is work in progress.

In [47] we try to construct a single proof calculus that meets the requirements that Ibn Sīnā mentions. For example it covers premise-sequences of any length, and pays due regard to Ibn Sīnā's remarks about the *bāl*. In this way we test how far his requirements are compatible with each other. Ibn Sīnā's logic is a good testbed for this kind of analysis, because he starts with a rather simple and straightforward logic and then adds various kinds of sophistication. The account in [47] relies on mathematical properties of these logics as drawn out in the present book.

For **Question Four** Ibn Sīnā relies on Aristotle's method for proving sterility. But it becomes clear that he doesn't understand it and is using it by rote. This gives us an excuse to set out Aristotle's method in a robust form; it will then be clear how Ibn Sīnā should have adjusted it to his new logics. We refer to the method as the *method of pseudoconclusions*.

Definition 4.1.2 Let $\mathcal{L}$ be a subject-predicate logic. A *cover* for $\mathcal{L}$ is a set $\mathcal{C}$ of forms of $\mathcal{L}$ with the property that for every sentence $f(R, S)$ of $\mathcal{L}$ there is a form $g \in \mathcal{C}$ with $g(R, S) \vdash f(R, S)$.

Of course the set of all sentence forms of $\mathcal{L}$ is a cover for $\mathcal{L}$. But since in practice the set of sentence forms of $\mathcal{L}$ is always finite, there is always a unique smallest cover for $\mathcal{L}$, namely the set of sentence forms of maximal strength. See Corollary 5.2.5 and Corollary 9.3.3 below for covers of some of Ibn Sīnā's logics.

Theorem 4.1.3 *Let $\mathcal{L}$ be a subject-predicate logic that has contradictory negations, and let $\mathcal{C}$ be a cover for $\mathcal{L}$. Suppose $T[B, A]$ is a premise-sequence of $\mathcal{L}$. Then the following are equivalent:*

(a) $T[B, A]$ is sterile.

(b) For each form $g \in \mathcal{L}$ there is a model M_g of the sentences in $T[B, A]$ which is also a model of $g(B, A)$.

Proof. (a) ⇒ (b). Suppose (b) fails; let $g \in \mathcal{C}$ be such that there is no model of $T[B,A]$ and $g(B,A)$. Then $T[B,A] \vdash \overline{g(B,A)}$, proving that $T[B,A]$ is not sterile.

(b) ⇒ (a). Suppose (a) fails. Then there is a sentence $f(B,A)$ of $\mathcal{L}$ such that $T[B,A] \vdash f(B,A)$. By the cover property there is $g \in \mathbb{C}$ such that $g(B,A) \vdash \overline{f(B,A)}$. Then no model of $T[B,A]$ is a model of $g(B,A)$, contradicting (b). □

Definition 4.1.4 When (b) holds in Theorem 4.1.3, Aristotle at *Prior Anal. i.15, 34b16* describes the sentence $g(B,A)$ as the *conclusion* of $T[B,A]$—perversely, since it is not deduced from $T[B,A]$. Alexander of Aphrodisias and Ibn Sīnā both copy this usage. We will describe $g(B,A)$ as the *pseudo-conclusion* of $T[B,A]$.

Aristotle's method for proving the sterility of $T[B,A]$, given a cover $\mathcal{C}$, was to write down for each form $g \in \mathcal{L}$ a true instance of $T[B,A]$ which has $g(B,A)$ as a pseudoconclusion. We use models rather than instances, but otherwise the method is the same.

Example 4.1.5 In unaugmented assertoric logic $\mathcal{L}_{uas}$ there is only one cover: the set of all four forms (a, uas), (e, uas), (i, uas) and (o, uas). So for example to prove the sterility of

(4.1.2) $$\forall x(Cx \to Bx),\ \ \exists x(Bx \wedge Ax)$$

by the method of pseudoconclusions we need to show that no sentence $f(C,A)$ can be deduced from these premises, by finding for each of the four possible values of f a corresponding model M_f:

- $M = M_{(a,uas)}$: domain $\{1\}$, $C^M = B^M = A^M = \{1\}$.
- $M = M_{(e,uas)}$: domain $\{1\}$, $C^M = \emptyset$, $B^M = A^M = \{1\}$.
- $M = M_{(i,uas)}$: as for (a, uas).
- $M = M_{(o,uas)}$: domain $\{1,2\}$, $C^M = \{1\}$, $B = \{1,2\}$, $A^M = \{2\}$.

Ibn Sīnā's answer to **Question Five** was to write down syntactic necessary and sufficient conditions for a premise-pair to be productive. He probably aimed to give conditions that would apply across the whole of a logic. But he didn't manage to do this, even in assertoric logic. In assertoric logic he gives separate necessary and sufficient conditions for each of the first three figures; they include clauses that apply to all three figures. In *Qiyās* he regularly begins a listing of productive premise-pairs by stating conditions for the premise-pairs in the relevant figure to be productive; he calls such a condition a *condition of productivity* (Arabic *šarṭ al-'intāj*).

Likewise he aims to state rules which define, for a given productive premise-pair, how to find the conclusion of the premise-pair. He often writes as if each feature of the conclusion (such as being universal or being negative) is inherited from one or other of the premises; so the rules can take the form of stating which of the two premises the conclusion 'follows' in respect of each feature. For this reason I call these rules the *rules of following*.

For assertoric logic, Ibn Sīnā took his conditions of productivity and his rules of following from earlier Peripatetic sources, most likely the commentary of Philoponus on the *Prior Analytics*. They were simply a compendium of observed facts. In Chapter 7 below we prove them, together with their generalisations to arbitrary premise-sequences. In later chapters we give—so far as we can—corresponding rules for Ibn Sīnā's other logics.

As noted above, we will reach these rules by a different route from Ibn Sīnā's: for each logic in question we will prove a mathematical characterisation of the productive premise-pairs and their conclusions. The gory details will take up several dozen pages below. But let me comment briefly on exactly what it is that we aim to characterise.

To match Ibn Sīnā's lists, we need a characterisation of the premise-sequences $T[B, A]$ and sentences $\chi(B, A)$ such that the inference

$$T[B, A] \;\rhd\; \chi(B, A). \tag{4.1.3}$$

holds. It will be convenient to characterise not the pairs as in (4.1.3), but their antilogisms (cf. Definition 2.1.24) using contradictory negations:

$$T \cup \{\overline{\chi(B, A)}\}. \tag{4.1.4}$$

In most of our logics the theories (4.1.4) are minimal inconsistent, which means the following:

Definition 4.1.6 A theory T is *minimal inconsistent* if T is inconsistent, and for every sentence $\phi \in T$, $T \setminus \{\phi\}$ is consistent. A Minimal Inconsistent Theory will be called an *MIT*.

Lemma 4.1.7 *Let T be an inconsistent first-order theory. Then some finite subset of T is minimal inconsistent.*

Proof. By the Compactness Theorem (Theorem 2.1.23) there is a finite n such that some set of n sentences in T is inconsistent. Take the least such n and an inconsistent set U of n sentences in T. Then U is minimal inconsistent. □

Minimal inconsistency is a powerful notion and very handy for proving things. So we will usually start by characterising the minimal inconsistent theories, and then among them we will select the antilogisms of the inferences (4.1.3). This approach gives us important new information. In the case of assertoric logic it turns out that all minimal inconsistent theories are graph-circular; this explains why Aristotle and his successors could confine their attention to premise-sequences and their conclusions. In the case of two-dimensional logic there are minimal inconsistent theories that are not graph-circular. Ibn Sīnā knew this but he seems not to have realised what it implies for logics that extend Aristotle's.

For some logics, experiment suggests that a variant of this route works better. Instead of cutting down from the minimal inconsistent sets to the antilogisms, we first cut down to a strictly smaller set of theories, namely those that are optimal inconsistent in the following sense.

Definition 4.1.8 (a) A theory U is a *weakening* of the theory T if there is a bijection $i : U \to T$ such that (i) for each $\phi \in U$, $i\phi \vdash \phi$, and (ii) for at least one $\phi \in U$, $\phi \nvdash i\phi$.

(b) A theory T is *optimally minimal inconsistent*, or for short *optimal inconsistent*, if T is minimal inconsistent and there is no consistent weakening of T.

These optimal inconsistent theories will give us the antilogisms of a proper subclass of the inferences (4.1.3); we recover the full class of inferences by allowing the premises to be strengthened, but not so much that they yield stronger conclusions. This approach works well for core two-dimensional logic, because there is a neat and simple characterisation of the optimal inconsistent theories in this case.

4.2 Logic as a science

Ibn Sīnā counts logic as a science, more precisely as a theoretical science. In several works he speaks about what kind of theoretical science logic is, and these passages together form a kind of metatheory for logic. The book [45] discusses this more fully. Here we pick out some themes that are useful for our present purposes.

Every theoretical science has a body of theorems. We can separate these theorems into two groups in two different ways, giving a total of four groups. First we can distinguish between those theorems that are self-evident or self-justifying (these are the 'principles', *mabādi'*) and those that are derived (Ibn Sīnā rather misleadingly calls these the 'questions', *masā'il*, because they answer questions posed in developing the science). And second we can distinguish those theorems whose justfication lies within the science itself from those whose justification rests on a higher science. For Ibn Sīnā there is only one science higher than logic, namely the part of metaphysics that he calls *First Philosophy*; it handles basic questions about existence, truth, concepts etc.

The main theorems of logic state that certain entailments hold; for example they state that a given mood or group of moods is productive, and that a given mood has a conclusion of such-and-such a form. The simplest case for a modern logician to understand is where these theorems are derived within the science of logic. In other words, the theorems are given logical proofs from other theorems already known. This part of Ibn Sīnā's scheme of logic corresponds closely to what we now call proof theory.

There are also theorems that are derived, but not within logic. Examples are the principles of excluded middle and non-contradiction, which Ibn Sīnā derives from underlying conceptual truths of First Philosophy. They are there within logic, but they play little formal role and we will rarely meet them below.

The most perplexing cases, at least for us, are the theorems that are self-justifying. What kind of justification can be given for a theorem that can't be derived from any prior theorem? In modern terms this is a question about the justification of the axioms of a science, as for example in Maddy [74].

When a self-justifying theorem states that a certain mood is productive and has such-and-such conclusions, Ibn Sīnā describes the mood as 'perfect' (Arabic *kāmil*). The theorem is self-justifying in the sense that the mood itself is self-evidently valid. Most of Ibn Sīnā's examples of perfect moods will raise few eyebrows; they describe particularly straightforward inferences. The exception is where Ibn Sīnā states what he considers perfect moods in alethic modal logic; the problem is that his examples include moods that many other logicians, both medieval and modern, consider invalid with obvious counterexamples.

Clearly we need some explanation of why he regards these modal moods as self-evidently valid. Some writers have constructed putative logical derivations of one or more of the problematic moods, for example from laws of S4 modal logic or from logical principles used by some Scholastic authors. But no derivation of this kind can rescue Ibn Sīnā's claim that these moods are self-justifying. If one could construct a putative derivation of a problematic mood from an unproblematic one, this could perhaps indicate that Ibn Sīnā had in mind a different formal system from the one that he presents, and that would be something that our account in this book would need to take on board. But so far as I'm aware, none of the published explanations of Ibn Sīnā's problematic modal moods make any sense within the science of logic as Ibn Sīnā sees it. Future work may alter this picture, but only if it is carried out with more awareness of Ibn Sīnā's own metatheory.

My own view is that the mathematical facts presented in this book fall into place within Ibn Sīnā's scheme of logic, and justify a quite different explanation of Ibn Sīnā's modal moods. Briefly, Ibn Sīnā regards necessity and possibility as concepts to be analysed and explained in logic and not in First Philosophy. (This does *not* include the notions of necessary existence and possible existence, which for Ibn Sīnā are theological concepts that belong in the higher parts of metaphysics.) His approach is to set up a formal and purely extensional framework in which we can describe in abstract terms how necessity works, taking temporal permanence as a paradigm. His claim is that in this way he can reach abstract laws of necessity and possibility. These laws are then criterial for whether a concept can be regarded as a 'necessity' concept or a 'possibility' concept, or more precisely whether a pair of concepts can be regarded as a 'necessity/possibility' pair. The scheme is driven by his two-dimensional logic; the justifications that he gives for the problematic 'perfect' modal moods are largely bluster and play no essential role in his formal logic. Ibn Sīnā is doing something new here, and with hindsight we can see that he doesn't always present his case in the best way. The phenomenon of *k̠abṭ* noted by Rāzī is one symptom of that.

This picture brings together the facts of two-dimensional logic with Ibn Sīnā's own explanations of the nature of logic. The present book takes care of the two-dimensional logic, and in Chapter n the relationship to the alethic modal logic is briefly spelled out. The book [45] will give a fuller account but taking the mathematics for granted.

Returning to Ibn Sīnā's general conception of a theoretical science: for him the theorems are normally universally quantified statements, where the quantification is restricted to some kind P of 'existing thing' (*mawjūd*):

(4.2.1) For all $x_1, \ldots, x_n$ in P, ...

He observes that for a typical science there is a most general P; for example in the science of arithmetic we have P = numbers. This P is called the *subject-term* (*mawḍū^c*) of the science; the *subject individuals* of the science are the existing things that fall under P.

We can see what the subject-term of logic should be if we look at typical theorems of logic, for example theorems that express the validity of a given mood. As in (3.3.1) above, we have a theorem of $\mathcal{L}_{uas}$:

(4.2.2) $$\exists x(Cx \wedge \neg Ax),\ \forall x(Bx \to Ax)\ \vdash\ \exists x(Cx \wedge \neg Bx).$$

This is a theorem of formal logic, in other words it states that all entailments of a certain form hold. Or in other words again, if we take any meaningful expressions (of the appropriate type) and put them for C, B and A in (4.2.2), we get a valid entailment between meaningful sentences. For Ibn Sīnā the syntactic forms in a language are an outward and visible sign of an inner construction of meanings; so he takes it that the quantification is really over meanings (of appropriate type) rather than over expressions that have those meanings. Thus for him, (4.2.2) should be understood as having the form

(4.2.3) For all well-defined meanings C, B and A, ...

This pattern of translation works quite generally, and it leads to the conclusion that the subject-term of logic is 'well-defined meaning'. (I am translating *ma^{c}nā ma^{c}qūl*, literally 'intellected meaning'. For Ibn Sīnā the intellect refines meanings to the point where they are suitable for rational thought.)

In this way Ibn Sīnā answers a question that is often asked: Should the 'C' and 'B' and 'A' in (4.2.2) be taken to be meaningful expressions, or to mark holes in a schema where meaningful expressions can be put? His answer is: Neither. The letters 'C' etc. should be read as (meta)variables of quantification over meanings. When we carry out calculations in formal logic, the calculations are done under an implicit universal quantification over meanings.

It will not concern us again in this book, but some readers may be interested to know what Ibn Sīnā has to say about the contents of the '...' in theorems of logic like (4.2.1) or (4.2.3). The concepts that appear in the '...' are what he calls the 'features' (*aḥwāl*) of well-defined meanings. He gives lists of these 'features' in both *Madḵal* and *Ta^clīqāt*. They include for example 'being a subject', 'being a predicate', 'being contained in', 'being incompatible', 'being universally quantified', 'being necessary'. This is as close as Ibn Sīnā ever comes to describing what he counts as a logical concept. The features are simply listed without justification and seem to be extracted from the actual practice of formal logic; there is no trace in Ibn Sīnā of any claim that they are distinguished by being 'topic-free', for example.

History 4.2.1 In a nutshell, Ibn Sīnā's characterisation of the subject term of logic answers the question 'What does it mean for logic to be formal?', and his answer bears close comparison with the views of Bolzano. His characterisation of the 'features' of the subject individuals answers a different question, namely 'What are the logical constants?'. The reader should be warned that over the centuries, Ibn Sīnā's characterisation of the subject term of logic has captured the interest of quite a number of people whose logical knowledge didn't reach to distinguishing between these two questions. As a result one often sees a confusion between the subject individuals of logic, which are arbitrary well-defined meanings, and their 'features', which are a small group of higher-order concepts. The texts collected by El-Rouayheb [25] show this confusion beginning within a couple of centuries after Ibn Sīnā.

4.3 Non-syllogistic steps

For Ibn Sīnā as for many Peripatetic logicians, the main purpose of logic was to provide a way of checking the correctness of inferences expressed in natural language. Given such an inference, one would try to analyse or paraphrase it so as to make it an instance of a valid mood. Success would mean that the inference was correct; failure would suggest that it might not be correct. A piece of text is an instance of a valid mood if it consists of terms arranged in an appropriate way; so this process of logical analysis required one to find terms in the text.

Arguments tend to consist of more than one inference. So at least the better courses in Peripatetic logic would teach the student how to break an argument down into single inferences that could be analysed and validated separately. There was no requirement that any of the terms used to validate one inference step would be the same as terms for validating other steps in the same argument. In fact Ibn Sīnā was one of a number of logicians who pointed out that when a sentence is the conclusion of one inference and a premise of the next inference, you might need to re-paraphrase the sentence and choose new terms for the second inference. Leibniz coined a phrase for this kind of move:

(4.3.1) It should also be realized that there are *valid non-syllogistic inferences* which cannot be rigorously demonstrated in any syllogism unless the terms are changed a little, and this altering of the terms is the non-syllogistic inference. (Leibniz [71] part 4 §xvii, p. 479)

So we can speak of steps that consist of *re-terming* a single sentence. In [40] I argued that this acceptance of different terms for different inferences within the same overall argument, which I called *local formalising*, is a pivotal difference between logic before the generation of Frege and Peano and logic after them.

The practice of validating arguments is a main theme of the book [47] and doesn't figure much in the present book. But here is an example of a re-terming at work, taken from Ibn Sīnā's student Bahmanyār bin al-Marzubān; see the translation by El-Rouayheb [24] p. 24.

(4.3.2)

$$\frac{\{C \equiv B\} \qquad \{B \equiv D\}}{\{C \equiv B\} \text{ and } \{B \equiv D\}}\ (\alpha)$$

$$\Big|\ (\beta)$$

$$\frac{\{B\}\ \{\text{has } C \equiv \text{it and is} \equiv D\}}{\text{Some } \{\text{line}\}\ \{\text{has } C \equiv \text{it and is} \equiv D\}}\ (\gamma)$$

$$\Big|\ (\delta)$$

$$\frac{\{C, D\} \text{ is a } \{\text{pair of lines with some line} \equiv\text{-between them}\} \qquad \text{Every } \{\text{pair of lines with some line} \equiv\text{-between them}\} \text{ is a } \{\text{pair of} \equiv \text{lines}\}}{\{C, D\} \text{ is a } \{\text{pair of} \equiv \text{lines}\}}\ (\varepsilon)$$

$$\Big|\ (\zeta)$$

$$C \equiv D$$

The task here is to apply the axiom of transitivity of equality for geometrical lines. Within assertoric logic this is complicated because each formal sentence has just one quantifier, so we have to switch between quantification over individuals and quantification over ordered pairs of individuals. My diagram uses curly brackets to show the terms. See for example the switch at (δ) from terms for lines to terms for ordered pairs of lines.

Thus George Boole, if he ever wanted to formalise an argument in English that uses both propositional and quantifier steps, would have to formalise once for a propositional step, then translate the conclusion back into

English, then re-term for the quantifier step, then back into English again. This has its comic aspects.

A number of the re-termings that we will use in this book have an important feature: the re-terming can be done uniformly from the formalised sentence without having to dig down inside the terms. Instead the new terms contain the old terms plus some extra material. Ibn Sīnā sometimes describes re-termings of this kind as 'including within the predicate'; he also speaks of them as 'reductions' (Arabic *rujū*^{c} or *radd*).

Reductions are in fact interpretations in the model-theoretic sense; the new terms are interpreted in the language of the old terms. If Boole had had re-termings like this, he could have incorporated them into his formalism and not had to slip back into English in mid flow. In this sense these re-termings, though they are non-syllogistic, mark a crucial step towards incorporating different kinds of logical step within a single formalism, and hence a move away from local formalising. Ibn Sīnā's use of them to prove some Second Figure moods in 2D logic (see History 10.4.1 below) was a historic moment. But the breakthrough might not have been visible to the Peripatetics, because almost everything was done in natural language. The only symbolism that Ibn Sīnā himself used was letters for terms.

It will be helpful to have a uniform formalism for reductions, to include both the reductions that Ibn Sīnā himself uses, and some that we will introduce for metamathematical purposes.

Definition 4.3.1 Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be logics. Then by a *paraphrase* of $\mathcal{L}_2$ into $\mathcal{L}_1$ we mean the following:

1. (a) For each signature Σ for $\mathcal{L}$, a signature Σ'' for $\mathcal{L}_1$ and a mapping $\phi \mapsto \phi'$ of sentences of $\mathcal{L}_1$ in $L(\Sigma'')$ to sentences of $\mathcal{L}_2$ in $L(\Sigma)$, such that:

(b) The mapping $\phi \mapsto \phi'$ is bijective from the sentences of $\mathcal{L}_1$ to the sentences of $\mathcal{L}_2$, up to logical equivalence.

(c) Let T be any set of sentences of $\mathcal{L}_1$, and put $T' = \{\phi' : \phi \in T\}$. Then T' is consistent if and only if T is consistent.

(d) For each sentence ϕ of $\mathcal{L}_1$,

$$\overline{\phi'} = (\overline{\phi})'. \qquad (4.3.3)$$

(We will abbreviate this statement to: *The mapping $\phi \mapsto \phi'$ respects contradictory negation*.)

(e) For each sentence $\phi(B, A)$ of T, if $\phi(B, A)'$ is $\theta(B', A')$ then $\phi(A, B)'$ is $\theta(A', B')$. (We will abbreviate this statement to: *The mapping* $\phi \mapsto \phi'$ *respects conversion*.)

In many cases the mapping $\phi \mapsto \phi'$ is a model-theoretic interpretation which gives, for each atomic formula of the language of $\mathcal{L}_1$, a formula of the language of $\mathcal{L}_2$. In these cases (d) and (e) are immediate, and by standard model theory REF any model of T' in (c) above yields a model of T. Moreover the proof of left to right in (c) then has a conceptual content: any logical law that holds in $\mathcal{L}_1$ has to apply also to the corresponding sentences of $\mathcal{L}_2$ because these sentences only say something that is already expressed in $\mathcal{L}_1$. So left to right in (c) would in principle be clear to Ibn Sīnā even without the apparatus of model-theoretic interpretations. The argument to prove the other direction of (c) is in general not so obvious; Ibn Sīnā almost always leaves it out—he would have had difficulties expressing it within his logical resources. But in all the cases that he uses, (c) does hold in both directions, so evidently some kind of intuition is operating efficiently in the background.

Together (b) and (c) show that for any sentence ψ of $\mathcal{L}_1$ and set T of sentences of $\mathcal{L}_1$,

(4.3.4) $$T \vdash \psi \;\Leftrightarrow\; T' \vdash \psi'.$$

This together with (a) shows that $\mathcal{L}_1$ and $\mathcal{L}_2$ are 'logically equivalent' in the sense that any inference in one logic can be translated into the other and vice versa. Condition (d) adds that conversions carry over too: if $\phi(B, A) \vdash \phi(A, B)$ then

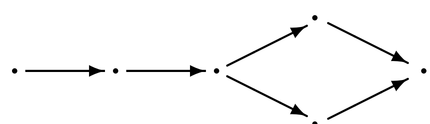
(4.3.5) $$\theta(B', A') = \phi(B, A)' \;\vdash\; \phi(A, B)' = \theta(A', B').$$

SOME OF THIS AS EXERCISE? SHOW CONVERSIONS ALSO CARRY OVER IN THE OPPOSITE DIRECTION. ALSO ADD ON MEANING POSTULATES.

History 4.3.2 Despite his use of reduction in some crucial cases, Ibn Sīnā shows a reluctance to use it freely. For example the propositional logic of his PL2 (Chapter BELOW) can be reduced to assertoric logic; both Wallis and Boole present versions of this paraphrase. But Ibn Sīnā prefers to let PL2 stand on its own two feet. I guess that his view was that in general reductions are a device of theoretical logicians and not a move that we make in everyday reasoning, even in the sciences. So his reluctance was basically for cognitive reasons. But in the case of the Second Figure moods, we see his 2D logic forcing him to adopt a more procedural approach. A century and a half later, the Persian logician Suhrawardī, who prided himself on his radicalism (and lost his life in consequence), made reduction one of the central methods of his logic. See Movahed REF.

4.4 Exercises

4.1. Give an example of a minimal inconsistent theory in unaugmented assertoric logic whose digraph has the form



Solution. $\exists x(Fx \wedge Ex)$, $\forall x(Ex \to Dx)$, $\forall x(Dx \to Cx)$, $\forall x(Cx \to Ax)$, $\forall x(Dx \to Bx)$, $\forall x(Bx \to \neg Ax)$.

4.2. The *propositional subject-predicate logic* $\mathcal{L}_{prop}$ has for its admissible signatures propositional signatures, i.e. signatures with no relation, function or individual constant symbols. There are four sentence forms (a), (e), (i) and (o), as follows.

$$\begin{array}{rcl} (a)(q,p) & : & (q \to p) \\ (e)(q,p) & : & (q \to \neg p) \\ (i)(q,p) & : & (q \wedge p) \\ (o)(q,p) & : & (q \wedge \neg p) \end{array}$$

The contradictory negation of (a) is (o), and that of (e) is (i). For each of the three digraphs Γ below, find a minimal inconsistent theory in $\mathcal{L}_{prop}$ whose subject-predicate digraph has Γ as underlying abstract digraph.

(a) The trinity digraph.

(b) The fourth figure digraph.

(c) The digraph of the form

$$\bullet \longrightarrow \bullet \longleftarrow \bullet$$

Solution. (a) $(p \to q)$, $(q \to r)$, $(p \wedge \neg r)$.
(b) $(p \to q)$, $(q \to \neg r)$, $(r \wedge p)$.
(c) $(p \wedge q)$, $(r \wedge \neg q)$.

4.3. Consider the four sentences of $\mathcal{L}_{uas}$ exhibited in Definition 3.1.5

(a) Show that if ϕ_1 and ϕ_2 are any two of them, then there is a model of ϕ_1 which is not a model of ϕ_2.

(b) Deduce from (a) that no sentence of $\mathcal{L}_{uas}$ has a weakening in $\mathcal{L}_{uas}$.

(c) Deduce from (a) that if $T(C, A)$ is a productive premise-sequence in $\mathcal{L}_{uas}$ then $T(C, A)$ has an optimal conclusion.

Solution. (a) Let M_1, M_2, M_3 and M_4 all have domain $\{0, 1\}$. Put $A^{M_2} = B^{M_2} = \emptyset$, $A^{M_1} = A^{M_3} = B^{M_3} = A^{M_4} = \{0\}$, $B^{M_4} = \{1\}$, $B^{M_1} = \{0, 1\}$.

	$(a, uas)(B, A)$	$(e, uas)(B, A)$	$(i, uas)(B, A)$	$(o, uas)(B, A)$
M_1	$false$	$false$	$true$	$true$
M_2	$true$	$true$	$false$	$false$
M_3	$true$	$false$	$true$	$false$
M_4	$false$	$true$	$false$	$true$

4.4. Show by the method of pseudo-conclusions that the unaugmented assertoric premise-pair

$$\forall x(Cx \to Bx), \ \ \exists x(Bx \wedge \neg Ax)$$

is sterile. (By Example ex:3.3.10 WHERE? this involves constructing four structures.)

4.5. (With reference to Definition 4.1.8) Suppose T is an MIT and U is a weakening of T. If U is inconsistent then U is also an MIT.

Solution. Suppose U is inconsistent but not an MIT. Then there is a proper subset U_0 of U that is an MIT. Let T_0 be the set of sentences of T that are strengthenings of sentences of U_0. Then T_0 is a proper subset of T, so it is consistent and has a model M. Since every sentence of U_0 is a weakening of one in T_0, U_0 is also a model of M, contradiction.]

4.6. Let T be a first-order theory and χ a sentence not in T. Show that the following are equivalent:

(a) $T \cup \{\chi\}$ is minimal inconsistent.

(b) T is consistent, $T \vdash \neg\chi$, and there is no proper subset T' of T such that $T' \vdash \neg\chi$.

Solution: (a) ⇒ (b): Assume (a). Then $T \cup \{\chi\}$ is inconsistent, so $T \vdash \neg\chi$. If there was a proper subset T' of T with $T' \vdash \neg\chi$, then $T' \cup \{\chi\}$ would also be inconsistent, and since χ is not in T, $T' \cup \{\chi\}$ would be a proper subset of $T \cup \{\chi\}$, contradicting the minimal inconsistency. If T was inconsistent then by similar reasoning T would be an inconsistent proper subset of $T \cup \{\chi\}$, again contradicting minimal inconsistency.

(b) ⇒ (a): Assume (b). Then since $T \vdash \neg\chi$, $T \cup \{\chi\}$ is inconsistent. Suppose for contradiction that some proper subset U of $T \cup \{\chi\}$ is inconsistent. If $U \subseteq T$ then we have a contradiction to the consistency of T. If $U \not\subseteq T$ then $U = T' \cup \{\chi\}$ for some proper subset T' of T, contradicting the third clause of (b). □

Part II

Assertoric

Chapter 5

Assertorics: the logic

This and the following three chapters are devoted to Aristotle's logic of assertoric sentences, as seen through the eyes of Ibn Sīnā. This logic is a subject-predicate logic; we write it $\mathcal{L}_{as}$.

There are three main kinds of material in Ibn Sīnā that lie behind these chapters. The first is his description of assertoric sentences, most fully in ^{c}*Ibāra* [54]. The second is his formal proof theory of assertoric sentences. He reports this in *Muḵtaṣar* [58] 49b9–53a6, *Najāt* [57] 57.1–64.3, *Qiyās* [55] ii.4, 108.12–119.8 and *Dānešnāmeh* [60] 67.5–80.2. Besides these four accounts, we also have a report in *Išārāt* [61] i.7, 142.10–153.9 ([50] 135–143) which is sketchier and mixed with modal material. In *Qiyās* vi.4, 296.1–304.4 Ibn Sīnā repeats the entire scheme in detail, but with a version of propositional logic (the PL2 of CHAPTER BELOW) in place of Aristotle's assertoric sentences. The third kind of material is his treatment of compound syllogisms in the early sections of *Qiyās* ix.

So we have a rather full account of this logic. As we noted in Section 1.3 above, nearly all Ibn Sīnā's other logics are derived from this one.

5.1 Assertoric sentence forms

Definition 5.1.1 We write $\mathcal{L}_{as}$ for *assertoric logic* (or *augmented assertoric logic* when we need to distinguish it from the unaugmented assertoric logic of Definition 3.1.1 above). The logic $\mathcal{L}_{as}$ is a subject-predicate logic with the four following sentence forms:

(5.1.1) $$(a), (e), (i), (o).$$

These are the four *assertoric sentence forms*. The domain of these sentence forms is the class of pairs (B, A) of distinct monadic relation symbols (cf. Definition 2.1.1 above), giving the following values:

$$\begin{array}{rcl} (a)(B,A) & : & (\forall x(Bx \to Ax) \land \exists x Bx) \\ (e)(B,A) & : & \forall x(Bx \to \neg Ax) \\ (i)(B,A) & : & \exists x(Bx \land Ax) \\ (o)(B,A) & : & (\exists x(Bx \land \neg Ax) \lor \forall x \neg Bx) \end{array}$$

Figure 5.1: The assertoric sentences

Definition 5.1.2 (a) The sentences $(a)(B, A)$ and $(i)(B, A)$ are said to be *affirmative*; the sentences $(e)(B, A)$ and $(o)(B, A)$ are said to be *negative*. The *simple negation* of $(a)(B, A)$ is $(e)(B, A)$, and the *simple negation* of $(i)(B, A)$ is $(o)(B, A)$ (cf. (2) in Section 3.2 above). The classification of an assertoric sentence as affirmative or negative is called its *quality* (Arabic *kayfa*).

(b) The sentences $(a)(B, A)$ and $(e)(B, A)$ are said to be *universal* (Arabic *kullī*), and the sentences $(i)(B, A)$ and $(o)(B, A)$ are said to be *existential* (Arabic *juz'ī*). The classification of an assertoric sentence as universal or negative is called its *quantity* (Arabic *kamm*).

(c) The conjunct $\exists x Bx$ in $(a)(B, A)$ is called the *existential augment*, and the disjunct $\forall x \neg Bx$ in $(o)(B, A)$ is called the *universal augment*. The augments are the only difference beetween the unaugmented and the augmented assertoric sentences.

History 5.1.3 The letters a, e, i, o were not used by the medieval Arabic logicians. Instead Ibn Sīnā has

(a):	universal affirmative
(e):	universal negative
(i):	existential affirmative
(o):	existential negative.

In fact the letters *a, e, i, o* come from some anonymous but very ingenious 13th century Latin Scholastic, who took the first two vowels of the Latin *affirmo* 'I say something affirmative' and the vowels of *nego* 'I say something negative'. Cf. pp. 66f of Kretzmann [70].

Conventional renderings of the assertoric sentences in modern English are

$$(5.1.2)\qquad \begin{array}{ll} (a)(B,A) & \text{‘Every } B \text{ is an } A\text{’.} \\ (e)(B,A) & \text{‘No } B \text{ is an } A\text{’.} \\ (i)(B,A) & \text{‘Some } B \text{ is an } A\text{’.} \\ (o)(B,A) & \text{‘Not every } B \text{ is an } A\text{’.} \end{array}$$

Nothing in this book depends on the sentences on the righthand side of (5.1.2) being accurate translations of the sentences on the left. If you find these translations inappropriate then feel free to choose others.

History 5.1.4 ‘Assertoric’ is our name; Ibn Sīnā has no settled name for the sentences of Arabic that are formalised by assertoric sentences as above. In ^{c}*Ibāra* [54] 112.5f he calls them ‘two-part or three-part’ to distinguish them from modal sentences, which are ‘four-part’. In *Qiyās* [55] i.5, 38.5f he describes them as ‘mentioned in the third book’ (i.e. ^{c}*Ibāra*) to distinguish them from the two-dimensional sentences where ‘time is taken into account’. They belong among the ‘predicative’ (*ḥamlī*) sentences because they assert or deny a predicate of a subject, and within the predicative sentences they belong to the ‘absolute’ (*muṭlaq*) sentences because they have no modality; within the absolute sentences Ibn Sīnā tends to pick them out as ‘standard’ (*mašhūr*), or as ‘absolute in the standard way’ (e.g. *Qiyās* ii.2, 89.11–13). Sometimes he says they have ‘convertible absoluteness’ (*al-’iṭlāq al- mun^{c}akis*, *Qiyās* iv.4, 214.2f). In short we have to identify them by what Ibn Sīnā says about them and does with them, not by any name that he uses for them.
The modern literature sometimes confuses ‘assertoric’ with ‘absolute’—and occasionally also with ‘actual’. This may be either a misunderstanding of Ibn Sīnā or a different but unexplained use of the word ‘assertoric’. Reader beware!

History 5.1.5 Ibn Sīnā is very clear and explicit about the augments on (a) and (o) sentences; see ^{c}*Ibāra* [54] 79.11–80.12 and discussion in [42] and [8]. He may have been the first logician to argue explicitly that sentences of the form ‘Every B is an A’ presuppose the existential augment. But he maintained that his predecessors believed this ‘except for a few hotheads’, and modern scholarship seems to be moving towards the same view. Read [86] argues that Aristotle assumed the existential augment on (a) sentences. Recently Chatti (private communication) found a passage in Al-Fārābī’s *Categories* that shows Al-Fārābī assumed the existential augment at least on a large class of affirmative sentences, over a century before Ibn Sīnā.

History 5.1.6 A fuller account of Ibn Sīnā's view of assertoric sentences would make further distinctions and add further sentence-types. (1) Ibn Sīnā distinguishes two kinds of existential proposition, namely those where the context (including the speaker's intentions) determines a particular individual and those where it doesn't. (In Arabic the distinction is between *mu^{c}ayyan* 'determinate' and *ḡayr mu^{c}ayyan* 'indeterminate'). (2) Ibn Sīnā also allows assertoric sentences where the subject term identifies an individual (these sentences are 'singular', Arabic *šaḵsī*), and assertoric sentences where the quantificational status of the subject is not made explicit (in Arabic these are *muhmal*). See further on both (1) and (2) in [45], and on (1) in [43].

5.2 At most two sentences

Lemma 5.2.1 *Every assertoric sentence has a model.* □

Lemma 5.2.2 *In assertoric logic $\mathcal{L}_{as}$ the sentences $(a)(B,A)$ and $(o)(B,A)$ are the contradictory negations of each other, and the sentences $(e)(B,A)$ and $(i)(B,A)$ are the contradictory negations of each other.* □

Lemma 5.2.3 *(a) The only entailments between assertoric sentences with tags (B,A) or (A,B) are as follows, together with those deducible from them by transitivity of entailment. (An arrow from ϕ to ψ means that $\phi \vdash \psi$.)*

$$
\begin{array}{ccccccc}
(a)(B,A) & & (a)(A,B) & \quad & (e)(B,A) & \Leftrightarrow & (e)(A,B) \\
\Downarrow & & \Downarrow & & \Downarrow & & \Downarrow \\
(i)(B,A) & \Leftrightarrow & (i)(A,B) & & (o)(B,A) & & (o)(A,B)
\end{array}
$$

(b) In the terminology of Definition 3.3.9, $(a)(B,A)$, $(e)(B,A)$ and $(i)(B,A)$ are all convertible, but only $(e)(B,A)$ and $(i)(B,A)$ convert symmetrically.

(c) If ϕ and ψ are assertoric sentences, ϕ is affirmative and ψ is negative, then neither of ϕ and ψ entails the other.

Proof. (c) follows from (a), but proving it directly will shorten the proof of (a). Let Σ be a monadic relational signature and ϕ, ψ assertoric sentences in $L(\Sigma)$. Suppose ϕ is affirmative and ψ negative. Let M be a Σ-structure such that $A^M = \text{dom}(M)$ for each relation symbol A in Σ. Then ϕ is true in

M and ψ is false in M, so $\phi \not\vdash \psi$. Let N be a Σ-structure such that $A^M = \emptyset$ for each relation symbol A in Σ. Then ψ is true in N and ϕ is false in N, so $\psi \not\vdash \phi$.

(a) The implications are all easy to check directly. By (c) there are no implications between the affirmative sentences and the negative.

If M is a structure with $A^M = \{1, 2\}$ and $B^M = \{1\}$ then $(a)(B, A)$, $(i)(B, A)$ and $(i)(A, B)$ are true in M but $(a)(A, B)$ is not. Hence none of $(a)(B, A)$, $(i)(B, A)$ and $(i)(A, B)$ entail $(a)(A, B)$. This and symmetry show that there are no further entailments on the affirmative side. We can complete the argument for the negative sentences either by a similar argument, or by applying contradictory negation to the affirmative sentences.

(b) can be read off from (a) and Definition 5.1.1. □

Example 5.2.4 As a result of Lemma 5.2.3, a minimal inconsistent set of assertoric sentences need not be optimal inconsistent (cf. Definition 4.1.8(b)). An example is

$$\{(\forall x(Bx \to Ax) \wedge \exists x Bx), \forall x(Bx \to \neg Ax)\}$$

which can be optimised by replacing the first sentence by $\exists x(Bx \wedge Ax)$.

Corollary 5.2.5 *The smallest cover (cf. Definition 4.1.1) for assertoric logic consists of the two forms* (a) *and* (e).

Proof. By Lemma 5.2.3 these are the forms of maximal strength. □

History 5.2.6 There is a diagram called the *square of opposition*, which exhibits the four assertoric sentences $(a)(B, A)$, $(e)(B, A)$, $(i)(B, A)$ and $(o)(B, A)$ as the corners of a square. Along the sides and diagonals of the square it states the logical relations between the pairs of sentences. In our sources this diagram first appears in the logic text of Apuleius ([72] p. 87ff = *Peri Hermeneias* V) in the 2nd century AD. It states the logical relations as they hold between the sentences. (But there are other readings of (a), (e), (i) and (o) that would give the same relations; this follows from Exercise 5.9 below.) It would be prudent not to talk of the square of opposition in connection with Ibn Sīnā, for two reasons. First, like Aristotle he never mentions such a diagram. And second, the relations hold in the case of assertoric sentences, but not in general. They fail quite badly for some of Ibn Sīnā's two-dimensional sentence forms; cf. Exercise 9.1 below.

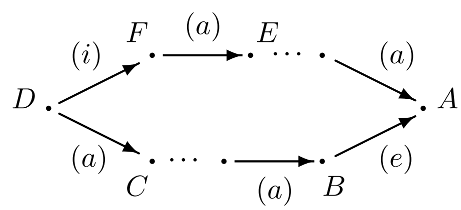
5.3 Classification of MITs

In this section we state a characterisation of the minimal inconsistent sets of assertoric sentences (i.e. the MITs). The proof of the characterisation will occupy Chapter 6 below; readers who are prepared to take the characterisation on trust can skip that chapter.

We describe theories in terms of their subject-predicate digraphs. The MITs fall into four disjoint families, and each family is subclassified by either one or two number parameters. The names of the families refer to their 'object classifier type', a notion explained in Chapter 6.

Type $(i)\uparrow$

(5.3.1)

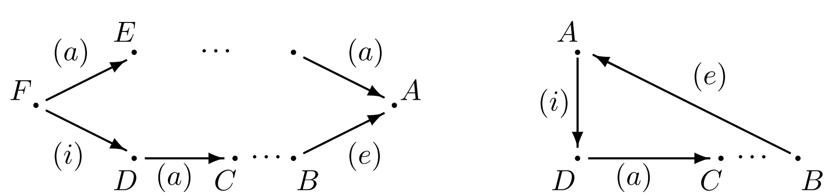


An MIT has type $(i)\uparrow\,(m,n)$ if it has a subject-predicate digraph as shown, where the upper track from D to A has length m and the lower track from D to A has length n. These two parameters can each have any value $\geqslant 1$. MITs of this type are always optimal inconsistent.

Type $(i)\downarrow$

For this type, two pictures are needed because the upper track can have length 0.

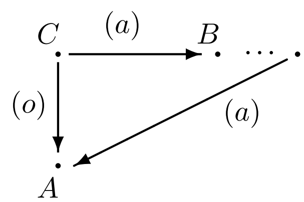
(5.3.2)



An MIT has type $(i)\downarrow\,(m,n)$ if it has a subject-predicate digraph as shown, where the upper track from D to A has length m and the lower track from D to A has length n. Here m can be any number $\geqslant 0$ and n can be any number $\geqslant 1$, subject to the requirement that $m+n \geqslant 2$. MITs of this type are always optimal inconsistent.

Type (o)

(5.3.3)

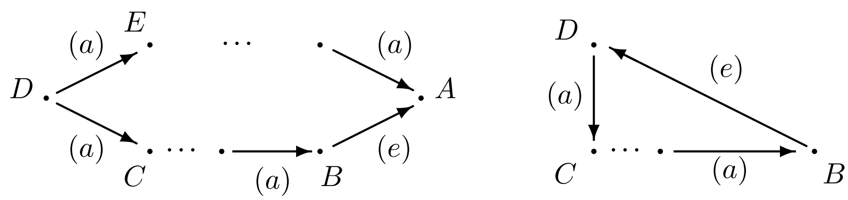


An MIT has type $(o)(m)$ if it has a subject-predicate digraph as shown, where the uppper (i.e. righthand) track from C to A has length m. Here m can be any number $\geqslant 1$. MITs of this type are always optimal inconsistent.

Type (a)

Again two pictures are needed, for the same reason as with type $(i)\downarrow$.

(5.3.4)



An MIT has type $(a)(m,n)$ if it has a subject-predicate digraph of one of the forms shown, where the upper track from D to A has length m and the lower track from D to A has length n. Here m can be any number $\geqslant 0$ (the righthand digraph applies when $m = 0$), subject to the requirement that $m+n \geqslant 2$. MITs of this type are never optimal inconsistent; weakening the first initial sentence of the lower track from (a) to (i) transforms $(a)(m,n)$ to $(i)\downarrow(m,n)$.

Theorem 5.3.1 *A set of assertoric sentences is minimal inconsistent if and only if it has a subject-predicate digraph of one of the types $(i)\uparrow$, $(i)\downarrow$, (o) or (a) as above. It is optimal inconsistent if and only if it has a subject-predicate digraph of one of the types $(i)\uparrow$, $(i)\downarrow$ or (o).*

Proof. This is Theorem 6.5.6 of Chapter 6 below. □

Corollary 5.3.2 *(a) Every assertoric MIT is graph-circular. Its digraph has at least two arrows and consists of two tracks, though one of the tracks may have length 0.*

(b) Every graph-linear assertoric theory is consistent; in particular if $T[B, A]$ is a premise-sequence then T is consistent.

Proof. (a) is from the Theorem, by inspection of the digraphs listed earlier in this section. Then (b) follows because no straight line contains a circle. □

Corollary 5.3.3 *If T is a consistent set of assertoric sentences and $\chi(B, A)$ is an assertoric sentence such that $T \vdash \chi(B, A)$, then there is a premise-sequence $U[B, A]$ with $U \subseteq T$, such that*

(a) $U \vdash \chi(B, A)$;

(b) there is an assertoric sentence $\psi(B, A)$, which is either $\chi(B, A)$ or a weakening of $\chi(B, A)$, such that $U \rhd \psi(B, A)$.

Proof. Suppose $T \vdash \chi(B, A)$. Then $T \cup \{\overline{\chi(B, A)}\}$ is inconsistent by Lemma 2.1.25, and hence by Lemma 4.1.7 some finite subset T' of $T \cup \{\overline{\chi(B, a)}\}$ is minimal inconsistent. Since T is consistent, T' contains $\overline{\chi(B, A)}$ and hence has the form $U \cup \{\overline{\chi(B, A)}\}$ for some $U \subseteq T$, and $\overline{\chi(B, A)}$ is not in U. Since T' is minimal inconsistent, it is graph-circular by Corollary 5.3.2, so that U is graph-linear with extremes B and A. Also $U \vdash \chi(B, A)$ by Lemma 2.1.25 again. □

Corollary 5.3.3 shows that all questions of entailment between assertoric sentences are reducible to questions about the productivity and conclusions of premise-sequences. So the restriction to this very special kind of inference in discussions of assertoric logic by Aristotle and other Peripatetic logicians is not an oversight.

Corollary 5.3.4 *Suppose $T[B, A]$ is a productive assertoric premise-sequence with conclusion $\phi(B, A)$. Then $\overline{\phi(B, A)}$ is not in T, and $T \cup \{\overline{\phi(B, A)}\}$ is an assertoric MIT.*

Proof. Since T has the extremes B and A, the only way that $\overline{\phi(B, A)}$ can be in T is for it to be the whole of T. But there is no assertoric sentence ψ such that $\{\overline{\psi}\} \vdash \psi$, by Lemma 5.2.1. So $\overline{\phi(B, A)}$ is not in T.

Now by Lemma 2.1.25, $T \cup \{\overline{\phi(B,A)}\}$ is inconsistent. Since T is graph-linear with extremes B and A, and it doesn't contain $\overline{\phi(B,A)}$, the theory $T \cup \{\overline{\phi(B,A)}\}$ is graph-circular. No proper subset of a graph-circular theory is graph-circular, and so by Corollary 5.3.2, no proper subset of a graph-circular theory is inconsistent. Hence $T \cup \{\overline{\phi(B,A)}\}$ is an MIT. □

Corollary 5.3.5 *(a) The subject-predicate digraph of an assertoric MIT consists of at most two tracks of positiive length.*

(b) A productive assertoric premise-sequence has either 0 or 2 switchpoints.

Proof. (a) is by inspection of the digraphs. For (b), let $T[B.A]$ be a productive assertoric premise-sequence, $\chi(B,A)$ the conclusion of $T[B,A]$, U the theory $T \cup \{\overline{\chi(B,A)}\}$. which is an MIT by Corollary 5.3.4, and Γ the subject-predicate digraph of U. If Γ has only one track then there are no internal switchpoints in T, and the initial and final sentences of $T[B,A]$ are both backwards, so $T[B,A]$ has no switchpoints. Suppose then that Γ has two tracks of positive length. We consider the ways that $\overline{\phi(B,A)}$ can lie in the two tracks.

Case One: $\overline{\phi(B,A)}$ is the whole of one track. Then there are no internal switchpoints in $T[B,A]$, and the two end sentences of $T[B,A]$ are both forwards, so that both B and A are switchpoints.

Case Two: $\overline{\phi(B,A)}$ is at one end of a track of length $\geqslant 2$. Then $T[B,A]$ has one internal switchpoint, one of its end sentences is forwards and one is backwards; so again $T[B,A]$ has two switchpoints.

Case Three: $\overline{\phi(B,A)}$ is not at either end of its track. In this case $T[B,A]$ has two internal switchpoints, and its end sentences are both backwards so that neither B nor A is a switchpoint. □

Corollary 5.3.6 *Let T be an assertoric MIT. Then:*

(a) [Rule of Quality] Exactly one sentence in T is negative.

(b) The unique negative sentence is the final sentence in one track of the subject-predicate digraph of T.

The *proof* is by inspection of the digraphs. □

Corollary 5.3.7 *Let $T[B,A]$ be a productive assertoric premise-sequence with conclusion $\phi(B,A)$. Then:*

(a) $\phi(B, A)$ *is negative if and only if some sentence in* T *is negative.*

(b) $\phi(B, A)$ *is the unique conclusion of* $T[B, A]$.

Proof. (a) follows from Corollary 5.3.4(a) and Corollary 5.3.6(a).

(b) Suppose also $\psi(B, A)$ is a conclusion of $T[B, A]$. Then by Definition 3.3.2(b), neither of $\phi(B, A)$ and $\psi(B, A)$ is a strengthening of the other. Also by (a), $\phi(B, A)$ and $\psi(B, A)$ are both affirmative or both negative. It follows from Lemma 5.2.3(a) that $\phi(B, A)$ and $\psi(B, A)$ are the same sentence. □

Corollary 5.3.8 *Let* T *be an assertoric MIT. Then:*

(a) [Rule of Quantity] At most one sentence in T *is existential.*

(b) There is an existential sentence in T *if and only if* T *is optimal inconsistent.*

(c) If there is an existential sentence in T *then it is the initial sentence in one track of the subject-predicate digraph of* T.

The *proof* is by inspection of the digraphs. □

Definition 5.3.9 In the light of Corollaries 5.3.6 and 5.3.8, we adopt certain conventions for writing subject-predicate digraphs of MITs. We call the track containing the unique negative sentence the *lower track*, and the other track (which may have length 0) the *upper track*. The lower track has positive length; we always orient the digraph so that this track runs anticlockwise. Let ν be the first node of the lower track; we put ν at the lefthand side of the digraph. If there is an existential sentence, its arrow has source ν; so far as convenient, we make this the only vertical arrow of the digraph.

Corollary 5.3.10 *Let* T *be an assertoric MIT that is not optimal inconsistent.*

(a) If T *has type* $(a)(0, n)$ *then* T *has exactly one inconsistent weakening, namely where the initial sentence of the lower track is changed from* (a) *to* (i).

(b) If T *has type* $(a)(m, 1)$ *then* T *has exactly two inconsistent weakenings, namely where the initial sentence of the upper track is changed from* (a) *to* (i)*, and where the initial sentence of the lower track is changed from* (e) *to* (o).

(c) *If T has type $(a)(m,n)$ with $m,n \geqslant 1$ then T has exactly two inconsistent weakenings, namely where the initial sentence of one of the two tracks is changed from (a) to (i).*

Proof. In assertoric logic a weakening of a sentence is always from (a) to (i) or from (e) to (o). In both cases this puts an existential sentence into the theory. An inconsistent weakening of an MIT is also an MIT, so by Corollary 5.3.8(c), the sentence weakened must be initial. Nor can both initial sentences be weakened, by Corollary 5.3.8(a).

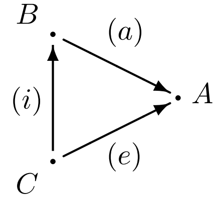
It remains to check that the sentence weakenings described in the Corollary do yield inconsistent weakenings of T. In case (a), weakening in the lower track changes the type to $(i)\downarrow(0,n)$. In case (b), weakening in the upper track changes the type to $(i)\uparrow(m,1)$, and weakening in the lower track changes the type to $(o)(m)$. In case (c), weakening changes the type to $(i)\downarrow(m,n)$ or $(i)\uparrow(m,n)$ according as the weakening is in the lower or upper track. □

5.4 Productive two-premise moods

We can find the productive assertoric two-premise moods by taking the assertoric MITs of Section 5.3 and selecting any one $\phi(B,A)$ of the sentences; the contradictory negation $\chi(B,A)$ of this sentence is entailed by the other two sentences, and these other two form the premise-sequence $T[B,A]$. We also require that no strengthening of $\chi B,A$ is also entailed by T. This requirement is equivalent to requiring that no weakening of $\phi(B,A)$ yields an inconsistent weakening of the original MIT; the conditions for this can be read off from Corollary 5.3.10 above.

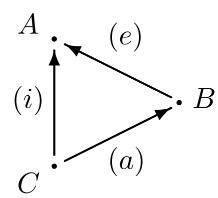
Each mood below is given a number of the form $i.j$; for example *Darii* is 1.3. The number means that the mood is the j-th mood in the i-th figure. These numbers are standard, though Ibn Sīnā uses only the numbering of the first three figures.

$(i)\uparrow(2,1)$:



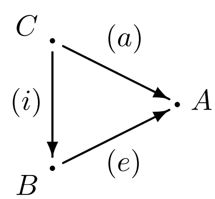
mood	minor	major	conclusion
1.3 *Darii*	$(i)(C,B)$	$(a)(B,A)$	$(i)(C,A)$
2.2 *Camestres*	$(e)(C,A)$	$(a)(B,A)$	$(e)(C,B)$
3.6 *Ferison*	$(i)(C,B)$	$(e)(C,A)$	$(o)(B,A)$

$(i)\!\uparrow(1,2)$:



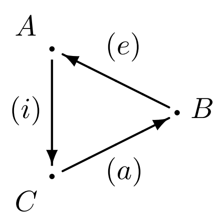
mood	minor	major	conclusion
1.2 *Celarent*	$(a)(C,B)$	$(e)(B,A)$	$(e)(C,A)$
2.3 *Festino*	$(i)(C,A)$	$(e)(B,A)$	$(o)(C,B)$
3.4 *Disamis*	$(a)(C,B)$	$(i)(C,A)$	$(i)(B,A)$

$(i)\!\downarrow(1,2)$:



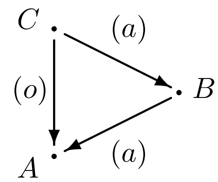
mood	minor	major	conclusion
1.4 *Ferio*	$(i)(C,B)$	$(e)(B,A)$	$(o)(C,A)$
2.1 *Cesare*	$(a)(C,A)$	$(e)(B,A)$	$(i)(C,B)$
3.3 *Datisi*	$(i)(C,B)$	$(a)(C,A)$	$(i)(B,A)$

$(i)\!\downarrow(0,3)$:



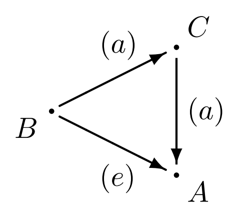
mood	minor	major	conclusion
4.5 *Fresison*	$(i)(A,C)$	$(e)(B,A)$	$(o)(C,B)$
4.3 *Calemes*	$(e)(B,A)$	$(a)(C,B)$	$(e)(A,C)$
4.2 *Dimatis*	$(a)(C,B)$	$(i)(A,C)$	$(i)(B,A)$

$(o)(2,1)$:



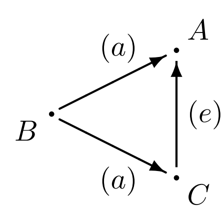
mood	minor	major	conclusion
1.1 *Barbara*	$(a)(C,B)$	$(a)(B,A)$	$(a)(C,A)$
2.4 *Baroco*	$(o)(C,A)$	$(a)(B,A)$	$(o)(C,B)$
3.5 *Bocardo*	$(o)(C,A)$	$(a)(C,B)$	$(o)(B,A)$

$(a)(2,1)$:



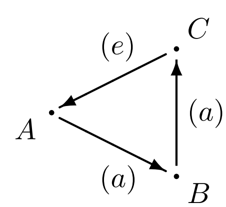
mood	minor	major	conclusion
3.2 *Felapton*	$(a)(B,C)$	$(e)(B,A)$	$(o)(C,A)$

$(a)(1,2)$:



mood	minor	major	conclusion
3.1 *Darapti*	$(a)(B,C)$	$(a)(B,A)$	$(i)(C,A)$

$(a)(0,3)$:



mood	minor	major	conclusion
4.1 *Bamalip*	$(a)(B,C)$	$(a)(A,B)$	$(i)(C,A)$
4.4 *Fesapo*	$(a)(A,B)$	$(e)(C,A)$	$(o)(B,C)$.

5.5 Exercises

5.1. Let T be a set of assertoric sentences.

(a) Show that if all the sentences in T are affirmative then T has a model.

(b) Show that if all the sentences in T are negative then T has a model.

(This strengthens Lemma 5.2.1 above.)

Solution: Let Σ be any monadic relational signature. (a) Take any nonempty set X, and construct a Σ-structure with domain X by taking $A^M = X$ for each relation symbol A in Σ. Then every affirmative assertoric sentence of $L(\Sigma)$ is true in M. (b) Do the same but taking $A^M = \emptyset$ for each relation symbol A in Σ. Then every negative assertoric sentence of $L(\Sigma)$ is true in M.

5.2. Two sentences ϕ, ψ are called *contraries* (Arabic *ḍidd*) if $\{\phi,\psi\}$ has no model, and *subcontraries* (Arabic *dāḵil taḥta al-taḍādd*) if $\{\overline{\phi},\overline{\psi}$ has no model. Show, from the definitions of the sentence forms, that $(a)(B,A)$ and $(e)(B,A)$ are contraries, and $(i)(B,A)$ and $(o)(B,A)$ are subcontraries.

Solution. Let Σ be a monadic relational signature containing the symbols A and B, and M a Σ-structure. We show first that $(a)(B,A)$ and $(e)(B,A)$ are not both true in M. Suppose $(a)(B,A)$ is true in M. Then there is an object a of M such that

$$M \models (\forall x(Bx \to Ax) \land Ba) \tag{5.5.1}$$

so also $M \models Aa$. But then $M \models (Ba \land Aa)$, so that M is not a model of $(e)(B,A)$.

We show secondly that at least one of $(i)(B,A)$ and $(o)(B,A)$ is true in M. Suppose $(o)(B,A)$ is false in M. Then there is an object a of M such that $M \models Ba$, but there is no object b of M such that $M \models (Bb \land \neg Ab)$, Hence $M \models (Ba \land Aa)$, in other words $(i)(B,A)$ is true in M.

5.3. Determine all the inconsistent sets consisting of two assertoric sentences. [Use Exercise 3.5 and Corollary 5.3.7 to restrict attention to pairs of sentences that have the same relation symbols.]

Solution. Suppose $T = \{\phi, \psi\}$ is an inconsistent set of assertoric sentences. If ϕ and ψ have no relation symbols in common, then each of $\{phi\}$ and $\{\psi\}$ is a connected component of T, and moreover each of these sets is consistent by Lemma 5.2.1; so by Exercise 3.5, T is consistent. We deduce that ϕ and ψ have at least one relation symbol in common. If they have only one relation symbol in common then T is graph-linear, and so by Corollary 5.3.7 T is consistent. We deduce that ϕ and ψ must have both their relation symbols in common, say A and B.

Now if T is inconsistent then $\phi \vdash \overline{\psi}$. By Lemma 5.2.3(c) it follows that one of ϕ and ψ is affirmative and the other is negative; let ψ be the affirmative, so that $\overline{\psi}$ is also affirmative. Then by Lemma 5.2.3(a) the only possibilities are:

(5.5.2) $\phi(B, A)$ is $(a)(B, A)$ and $\overline{\psi}$ is $(a)(B, A)$ or $(i)(B, A)$ or $(i)(A, B)$;
or $\phi(B, A)$ is $(i)(B, A)$ and $\overline{\psi}$ is $(i)(A, B)$ or $(i)(A, B)$.

or the same with A and B transposed. So the only inconsistent pairs are

$$\begin{array}{c} \{(a)(B, A), (o)(B, A)\}, \\ \{(a)(B, A), (e)(B, A)\}, \\ \{(a)(B, A), (e)(A, B)\}, \\ \{(i)(B, A), (e)(B, A)\}, \\ \{(i)(B, A), (e)(A, B)\} \end{array} \tag{5.5.3}$$

up to choice of relation symbols.

5.4. Use Aristotle's method of pseudoconclusions (cf. Definition 4.1.4 above) to prove that each of the following premise-sequences is sterile.

(a) $((a)(C, B), (i)(B, A))$.

(b) $((e)(C, B), (i)(A, B))$.

(c) $((o)(B, C), (e)(B, A))$.

Solution. (a) With pseudoconclusion $(a)(C, A)$: $A^M = B^M = C^M = \{1\}$.
With pseudoconclusion $(e)(C, A)$: $A^M = \{1\}, B^M = \{1, 2\}.C^M = \{2\}$.

(b) With pseudoconclusion $(a)(C, A)$: $A^M = \{1, 2\}$, $B^M = \{1\}$, $C^M = \{2\}$.
With pseudoconclusion $(e)(C, A)$: $A^M = B^M = \{1\}$, $C^M = \{2\}$.
(c) With pseudoconclusion $(a)(C, A)$: $A^M = C^M = \{1\}$, $B^M = \{2\}$.
With pseudoconclusion $(e)(C, A)$: $A^M = \{1\}$, $B^M = \{2\}$, $C^M = \{3\}$.

5.5. Use Corollary 5.3.7 to show the following variant of Theorem 4.1.3 for assertoric logic:

Theorem 5.5.1 *Suppose $T[B, A]$ is a premise-sequence in assertoric logic. Then the following are equivalent:*

(a) $T[B, A]$ is sterile.

(b) Either T contains a negative sentence and there is a model of T which is also a model of $(a)(B, A)$, or T doesn't contain a negative sentence and there is a model of T which is also a model of $(e)(B, A)$.

Solution. (a) $\Rightarrow$ (b) is as for Theorem 4.1.3. In the other direction, suppose (a) fails, and let $\phi(B, A)$ be the conclusion of $T[B, A]$. Then by Corollary 5.3.7, ϕ is negative if and only if some sentence in T is negative. In the case where T does contain a negative sentence, suppose for contradiction that there is a model M of $T \cup \{(a)(B, A)\}$. The sentence $\phi(B, A)$ must also be true in M, and it is negative, hence it is either $(e)(B, A)$ or $(o)(B, A)$. But this is impossible since $(a)(B, A)$ is not consistent with either of these sentences.

History 5.5.2 In practical terms, using Theorem 5.5.1 instead of Theorem 4.1.3 reduces the effort of proving sterility by about half. Aristotle can be excused for not using it, because he had to build up the entire subject from first principles. But one wonders why Ibn Sīnā never mentioned it. This adds to the evidence that Ibn Sīnā wasn't really on top of the theory of sterility.

5.6. Show that for any natural number n there is an assertoric theory that is consistent but has only models of cardinality $\geqslant n$.

Solution. Take a signature with relation symbols $A_1, \ldots, A_n, B$, and the sentences

$$\begin{array}{ll} (e)(A_i, A_j) & \text{whenever } 1 \leqslant i < j \leqslant n; \\ (i)(A_i, B) & \text{whenever } 1 \leqslant i \leqslant n. \end{array}$$

5.7. Show that if $\phi(A, B)$ is any assertoric sentence then ϕ has a model in which A and B have the same interpretation. Show that this is not so for unaugmented assertoric sentences in general.

Solution. Models with the required property are in the solution to Exercise 5.1 above. In unaugmented assertoric logic the sentence $(o, uas)(A, B)$ has no models M in which $A^M = B^M$.

5.8. Let A and B be two distinct monadic relation symbols.

(a) Show that up to logical equivalence there are exactly 25 boolean combinations of assertoric sentences with tag (B, A) or (A, B). [Open but vague question: Find a description of these 16 combinations, up to logical equivalence, that makes the set of them in any way intuitively natural.]

(b) For comparison, show that up to logical equivalence the number of boolean combinations of the sentences $\exists x(Ax \wedge Bx)$, $\exists x(Ax \wedge \neg Bx)$, $\exists x(\neg Ax \wedge Bx)$, $\exists x(\neg Ax \wedge \neg Bx)$ is $2^{16} = 65,536$.

Solution: (a) By considering contradictory negations, the strongest boolean combinations of sentences with tag (B, A) are

$$(a)(B,A)\wedge(e)(B,A),\ (a)(B,A)\wedge(i)(B,A),\ (o)(B,A)\wedge(e)(B,A),\ (o)(B,A)\wedge(i)(B,A).$$

which simplify to

$$\bot,\ (a)(B,A),\ (e)(B,A),\ (o)(B,A)\wedge(i)(B,A).$$

So the strongest boolean combinations using both (B, A) and (A, B) are $\bot$ and the nine combinations

$(a)(B,A) \wedge (a)(A,B)$	$(a)(B,A) \wedge (e)(A,B)$	$(a)(B,A) \wedge ((o)(A,B) \wedge (i)(A,B)$
$(e)(B,A) \wedge (a)(A,B)$	$(e)(B,A) \wedge (e)(A,B)$	$(e)(B,A) \wedge ((o)(A,B) \wedge (i)(A,B)$
$((o)(B,A) \wedge (i)(B,A))$ $\wedge(a)(A,B)$	$(((o)(B,A) \wedge (i)(B,A))$ $\wedge(e)(A,B)$	$((o)(B,A) \wedge (i)(B,A))$ $\wedge((o)(A,B) \wedge (i)(A,B))$

which simplify to

$(a)(B,A) \wedge (a)(A,B)$	$\bot$	$(a)(B,A) \wedge (o)(A,B)$
$\bot$	$(e)(B,A)$	$\bot$
$(o)(B,A) \wedge (a)(A,B)$	$\bot$	$(o)(B,A) \wedge (i)(B,A) \wedge (o)(A,B)$

This provides 5 pairwise inconsistent strongest consistent boolean combinations, so there are $5^2 = 25$ pairwise inequivalent disjunctions of these 5.

(b) The number of pairwise inconsistent strongest consistent boolean combinations is $2^4 = 16$.

5.9. Let Σ be a monadic relational signature and T a set of assertoric sentences of $L(\Sigma)$. Let T' be the unaugmented assertoric theory consisting of all the sentences $(f, uas)(B, A)$ such that $(f)(B, A) \in T$. Show that the following are equivalent:

(a) T has a model.

(b) T has a model M in which for every relation symbol A in Σ, $A^M \neq \emptyset$.

(c) T' has a model M in which for every relation symbol A in Σ, $A^M \neq \emptyset$.

[So assertoric logic is isomorphic to the logic where the sentences are the unaugmented assertoric sentences but the structures are required to have all relations non-empty.]

Solution: (a) $\Rightarrow$ (b): Given a model N of T, let $\mathcal{R}$ be the set of relation symbols B in Σ such that some sentence of the form $(o)(B, A)$ is true in N. Let M be an extension of N got by adding to the domain of N a distinct new element a_B for each $B \in \mathcal{R}$, putting $A^M = A^N \cup \{a_B\}$. (b) $\Rightarrow$ (c): The only sentences that need checking are those of the form $(o, uas)(B, A)$, since $(o)(B, A) \nvdash (o, uas)(B, A)$. But if $M \vdash (o)(B, A)$ and B^M is not empty then $M \vdash (o, uas)(B, A)$. (c) $\Rightarrow$ (a): The only sentences that need checking are those of the form $(a)(B, A)$, since $(a, uas)(B, A) \nvdash (a)(B, A)$. But if $M \vdash (a, uas)(B, A)$ and B^M is not empty then $M \vdash (a)(B, A)$.

5.10. Show that every graph-acyclic assertoric theory is consistent.

Solution. Let T be an assertoric theory with acyclic subject-predicate digraph Γ. By Exercise 3.5 above, to show that T is consistent it suffices to find a model for each connected component T_0 of T. The subject-predicate digraph Γ_0 of T_0 will still be acyclic. We begin with a claim.

> **Claim**. Consider a structure M in a monadic relational signature containing distinct relation symbols A and B; suppose the domain of M is $\{1, 2\}$ and $A^M = \{1\}$. Let ϕ be an assertoric sentence with relation symbols A and B. If $B^M = \{1\}$ and ϕ is affirmative then $M \models \phi$. If $B^M = \{2\}$ and ϕ is negative then $M \models \phi$.

Proof of claim. This is best checked directly for each of the eight possible forms of ϕ. □ Claim.

Now let $f : n \to Nodes$ be a listing of the nodes of Γ_0 as in (2.3.8) of Lemma 2.3.18, and let Σ be a monadic relational signature consisting of the n relation symbols A_i ($i < n$) that occur in T_0. We construct a Σ-structure M with domain $\{1, 2\}$. We define M^{A_i} for each i, by induction on i. Take M^{A_0} to be $\{1\}$. Suppose M^{A_i} has been defined. There is $j < i + 1$ such that the labels on the nodes $f(j)$ and $f(i + 1)$ are the relation symbols of a (necessarily unique) sentence ϕ_i of T_0. By the Claim we can choose $M^{A_{i+1}}$ so as to make ϕ true in M. When $M^{A_{n-1}}$ has been defined, M is a model of all the $n - 1$ sentences in T_0. $\square$

Chapter 6

Skolem theory of assertorics

The chapter break at this point marks a shift to a different style of working. The previous chapter put some of Ibn Sīnā's own logic into modern notation. The present chapter develops methods for answering questions about this logic, with no claim that we are using concepts that Ibn Sīnā himself could handle. The main outcomes—that this or that premise-sequence is or is not productive, with such-and-such a conclusion—are ones that he could easily calculate for himself, but he had no tools for proving general metatheorems.

6.1 The Skolem sentences

Definition 6.1.1 It will be convenient in this chapter to take both the augmented and the unaugmented assertoric sentences together. So we introduce a logic $\mathcal{L}_{\text{as+uas}}$ whose sentence forms are all six of the forms (a), (a, uas), (e), (i), (o) and (o, uas) from Definitions 3.1.5 and 5.1.1 above. Generally in this chapter we will abbreviate $\mathcal{L}_{\text{as+uas}}$ to $\mathcal{L}$, and then spell out just which sentence forms we are dealing with at each point.

We begin by choosing skolemisations for the sentences of $\mathcal{L}$; see Definition 2.2.6(a) above.

Definition 6.1.2 In four of the six sentence forms of $\mathcal{L}$, a direct skolemising of the sentence ϕ gives a conjunction, and we split the conjunction into separate sentences. These separate sentences are called the *Skolem pieces* of ϕ; we write $\phi_{[n]}$ for the n-th Skolem piece of ϕ, in the order given in the relevant line of the figure below. A Skolem piece of ϕ is *primary* if it contains the predicate symbol of ϕ, and *secondary* otherwise. Every assertoric ϕ has just one primary Skolem piece, and we list this piece as $\phi_{[1]}$.

sentence	primary Skolem piece	secondary Skolem piece
$(a)(B,A)$	$\forall x(Bx \to Ax)$	Ba_1
$(a,uas)(B,A)$	$\forall x(Bx \to Ax)$	
$(e)(B,A)$	$\forall x(Bx \to \neg Ax)$	
$(i)(B,A)$	Aa_2	Ba_2
$(o)(B,A)$	$\forall x(Bx \to \neg Aa_3)$	$\forall x(Bx \to Ba_3)$
$(o,uas)(B,A)$	$\neg Aa_4$	Ba_4

Figure 6.1: Skolem pieces for assertoric sentences

Each numbered constant in Figure 6.1 is a Skolem constant. The numbering is to indicate which constants have to be distinct, as in Definition 2.2.6(b). The symbols are shown as for a two-dimensional sentence with subject and predicate symbols B and A respectively; so for example the constant a_2 should be read as if it was $a_{2,B,A}$. For the assertoric sentence $(i)(D,C)$ the corresponding constant is $a_{2,D,C}$, so the two constants are distinct.

We will need some vocabulary for discussing these Skolem pieces. (In the 2D logic of Chapter 9 below the sentences are more complicated, but their Skolem pieces are not much more complicated than those above, so that the definitions given below can carry over to 2D logic too.)

Definition 6.1.3 (a) A Skolem piece is said to be *n-part* if it has exactly n atomic subformulas. Here n can be 1 or 2.

(b) The *head subformula* of a 1-part Skolem piece is its atomic subformula; the *head subformula* of a 2-part Skolem piece is its second atomic subformula.

(c) If a Skolem piece ψ of ϕ is 2-part, we describe the relation symbol in the first atomic subformula of ψ as the *subject symbol* of ψ and the relation symbol in the second atomic subformula of ψ as the *predicate symbol* of ψ.

(d) A Skolem piece is called *negative* if it contains a negation sign, and *affirmative* otherwise.

The following classification of Skolem sentences will become crucial in later sections.

Definition 6.1.4 Let ϕ be a Skolem piece. For this definition we assume that the head subformula of ϕ is At, where t is a term.

(a) We say that ϕ is *fixed at* t if t is a constant.

(b) We say that ϕ is *fixed* if ϕ is fixed at some c.

(c) We say that ϕ is *echoic* if ϕ is 2-part and t is a variable that occurs in both atomic subformulas, both of them inside the scope of the same universal quantifier $\forall t$.

Example 6.1.5 We refer to the listing of Skolem pieces in Figure 6.1. The sentence $(a)(B, A)$ has its primary Skolem piece echoic, but its secondary Skolem piece is fixed at a_1. The sole Skolem piece of $(a, uas)(B, A)$ is echoic, and likewise the sole Skolem piece of $(e)(B, A)$. Both Skolem pieces of $(i)(B, A)$ are fixed at a_2. Both Skolem pieces of $(o)(B, A)$ are fixed at a_3.

The Skolem pieces of $(o)(B, A)$ are tiresomely complicated. But it turns out that for our purposes in this chapter, the sentence form (o) is completely avoidable. The following lemma shows how.

Lemma 6.1.6 *Let T be a theory in $\mathcal{L}$. Let T' be the theory constructed from T as follows:*

(6.1.1) *For each sentence of the form $(o)(B, A)$ in T, suppose that there is no sentence $(a, uas)(B, C)$ in T with the same subject symbol as $(o)(B, A)$. Then replace $(o)(B, A)$ in T by $(o, uas)(B, A)$. Keep all other sentences of T unchanged.*

Then T is consistent if and only if T' is consistent.

Proof. Since $(o)(B, A)$ is a weakening of $(o, uas)(B, A)$, the consistency of T' implies the consistency of T.

In the other direction, suppose T is in the language $L(\Sigma)$, and the Σ-structure M is a model of T. We construct a Σ-structure N as follows. The domain of N is the same as that of M. Let A be any monadic relation symbol in Σ such that A^M is empty and there is no sentence of the form $(a, uas)(A, C)$ in T. Choose a new element a, add it to the domain of N and put $A^N = \{a\}$. Apart from the symbols A treated in this way, and the new elements added to the domain, N is the same as M.

We claim that N is a model of T'. Suppose first that a sentence $(a)(B, D)$ is in T and hence true in M. Then B^M and D^M are not empty, so $B^N = B^M$ and $D^N = D^M$; it follows that $(a)(B, D)$ is true in N.

Suppose that $(a, uas)(B, D)$ is in T. If B^M is not empty then neither is D^M, and the argument of the previous paragraph applies. If B^M is empty then B^N is still empty because (a, uas) is in T; so again $(a, uas)(B, D)$ is true in N.

If $(e)(B, D)$ was true in T then it remains true in N since we added no element that is in both B^N and D^N.

If either $(i)(B, D)$ or $(o, uas)(B, D)$ was true in M then it remains true in N, by the model-theoretic fact that adding new elements never changes a prenex sentence with only existential quantifiers from true to false.

There remains the case that T contains a sentence $(o)(B, D)$. We have to consider three subcases. The first subcase is that there is an element b in M which is in B^M but not in D^M; since b is in N too, $(o, uas)(B, D)$ is true in N. The second subcase is that B^M is empty in M, and there is no sentence $(a, uas)(B, C)$ in T. In this subcase we added in N an element that is in B^N and not in D^N, so that (o, uas) is true in N. The third subcase is that B^M is empty and there is a sentence $(a, uas)(B, C)$ in T. In this subcase B^N remains empty, making $(o)(B, D)$ true in N; but also we kept the sentence $(o)(B, D)$ in T instead of changing it to $(o, uas)(B, D)$.

So in all cases N is a model of T', as required. □

Corollary 6.1.7 *An assertoric theory T is consistent if and only if T' is consistent, where T' comes from T by removing the universal augments from (o) sentences.* □

6.2 Refutations

Definition 6.2.1 Let Σ be a monadic relational signature. By a *Skolem Σ-track*, or *Skolem track* for short, we mean a track

(6.2.1)
$$\overset{\bullet}{\nu_1} \longrightarrow \overset{\bullet}{\nu_2} \quad \cdots \quad \overset{\bullet}{\nu_{n-1}} \longrightarrow \overset{\bullet}{\nu_n}$$

where in addition each node carries two labels, its *theory label* and its *Skolem label*, satisfying the following conditions.

(a) The *theory label* of ν is an assertoric sentence of $L(\Sigma)$.

(b) The *Skolem label* of ν is a Skolem piece σ of the theory label of ν and is also in $L(\Sigma)$; σ is 1-part if ν is the initial node and 2-part otherwise.

Definition 6.2.2 (a) Let ν be a node. The node ν is said to be *affirmative* or *negative* according as its Skolem label is affirmative or negative. Also we apply the classifications of Definition 6.1.4 to ν according to its Skolem label; for example ν is *fixed at* t if its Skolem label is fixed at t.

(b) We say that the node ν is *primary* if it carries the predicate symbol of its theory label, and *secondary* otherwise.

(c) Let ν be a node and R a relation symbol. If R is the relation symbol of the head subformula of the Skolem label of ν, then we say that R is the *relation symbol at* ν, or that ν *carries* R.

Definition 6.2.3 Let ρ be a Skolem track. The first node in the track is the only node that is fixed; all later nodes are echoic. We call the first node the *nominator* of the track. (In 2D logic it is no longer true that all successor nodes are echoic, and the present definition will have to be made seriously more complicated.)

Definition 6.2.4 Let Σ be a monadic relational signature.

(a) By a *Skolem* Σ*-diagram*, or simply a *diagram* when the context allows, we mean an ordered pair ρ of tracks as in Definition 2.3.6(b), distinguished as the *upper track* and the *lower track* of ρ, such that each node carries two labels, its *theory label* and its *Skolem label*, satisfying the following condition.

(1) Each track is a Skolem track.

The set of theory labels of ρ is called the *theory* of ρ.

(b) We now give a further three conditions that a Skolem Σ-diagram may or may not satisfy. In these conditions, we describe the relation symbol of the head subformula of the Skolem label of a node μ as the relation symbol *at* μ, or *carried by* μ. Note that the notions of 'constrained', 'nominator', 'relation symbol carried' etc. make sense in the present context because they are defined in terms of the Skolem labels of the nodes. We say that a Skolem Σ-diagram ρ is a *Skolem Σ-refutation* (or when the context allows, just a *refutation*) if it satisfies the following three conditions (3)–(5):

(2) (First Propositional Condition) For any nonterminal node μ of ρ, the relation symbol at μ is equal to the subject relation symbol of the Skolem label of μ^+.

(3) (Second Propositional Condition) The relation symbols at the two terminal nodes are the same, but the upper terminal node is affirmative and the lower terminal node is negative.

By (3) there is only one negative node in ρ, namely the lower terminal node (cf. Exercise 5.1?? below); we call it the *negative node* of ρ.

(4) (Constraint Condition) The nominators of the two tracks are fixed at the same constant.

Theorem 6.2.5 (Characterisation Theorem) *Let T be an assertoric theory in a language $L(\Sigma)$, and Sk(T) its skolemisation in $L(\Sigma^{sk})$. Assume that Σ^{sk} has at least one individual constant. Then the following are equivalent:*

(a) T is inconsistent.

(b) There is a Skolem Σ^{sk}-refutation whose theory is included in T.

Proof. For both parts of the proof we define a *Herbrand sequence* of T to be a sequence

$$\phi_0, (\phi_0 \to \phi_1), \ldots, (\phi_{n-1} \to \phi_n) \tag{6.2.2}$$

where the sentences ϕ_0 and $(\phi_i \to \phi_{i+1})$ are all in the Herbrand theory Hr(T) of T in the language $L(\Sigma^{sk})$. So the sentences ϕ_i are literals, and all except perhaps ϕ_n are atomic. We call ϕ_n the *outcome* of this sequence, and we note that Hr(T) $\vdash \phi_n$.

(a) $\Rightarrow$ (b). Assume (a). Let Φ be the set of all outcomes of Herbrand sequences of T.

Claim One. For some atomic sentence ψ, Φ contains both ψ and $\neg\psi$.

Proof of Claim. Suppose the Claim fails. Then by the Canonical Model Lemma (Lemma 2.1.15) there is a Σ^{sk}-structure M such that for every atomic sentence θ of $L(\Sigma^{sk})$,

(6.2.3) $$M \models \theta \;\Leftrightarrow\; \theta \in \Phi.$$

We show that M is a model of Hr(T), so that Hr(T) is consistent.

Suppose first that θ is a literal in Hr(T). Then $\theta \in \Phi$, since it is the outcome of a one-sentence Herbrand sequence. If θ is atomic, we have $M \models \theta$ by (6.2.3). If θ is $\neg\eta$ with η atomic, then by the failure of the Claim, η is not in Φ, so $M \not\models \eta$ by (6.2.3) and hence $M \models \theta$.

Suppose secondly that $(\theta \to \chi)$ is a sentence in Hr(T). If $M \models \neg\theta$ then $M \models (\theta \to \chi)$. Suppose then that $M \models \theta$. Since θ is atomic, we have that $\theta \in \Phi$ by (6.2.3), and so there is a Herbrand sequence σ with outcome θ. We can add $(\theta \to \chi)$ at the end of σ to get a new Herbrand sequence with outcome χ; this shows that $\chi \in \Phi$. It follows that $M \models \chi$ as in the preceding paragraph.

So Hr(T) is consistent, and it follows by Lemma 2.2.9 that T is consistent, contradiction. Hence the Claim holds. □ Claim One.

So by the claim there are a Herbrand sequence σ with an outcome θ and another Herbrand sequence τ with an outcome $\neg\theta$, say with lengths m and n respectively. Form a Skolem refutation as follows. Draw a diagram of two tracks, the upper of length m and the lower of length n. Correlate the Herbrand sentences of σ and τ to the nodes of the upper and lower tracks respectively, preserving the order of the sentences. Give each node μ a Skolem label, which is the Skolem sentence that gave rise to the Herbrand sentence correlated to μ; give μ a theory label which is the sentence of T that gave rise to the Skolem label at μ. (It can be checked that these two labels are both determined by the Herbrand sentence.) Most of the conditions for a Skolem refutation are readily checked. We check the Constraint Condition. Let μ and ν be respectively the upper and lower nominators, fixed respectively at constants c and d. By the construction of the Herbrand sequences, c is the constant in the head subformula of each of the Herbrand sentences in the upper track from the sentence at μ to the terminal node, and likewise d is the constant from ν onwards in the lower track. So c and d are respectively the constants in the outcomes of the two sequences. But one of the outcomes is the negation of the other, and hence $c = d$. This proves (b).

(b) $\Rightarrow$ (a). By (b), let ρ be a Skolem refutation with $Th(\rho) \subseteq T$. We show that $Th(\rho)$ is inconsistent.

For each sector S of either track of ρ, let $c(S)$ be the constant at which the nominator of the sector is fixed. By the Constraint Condition, if S_1 and S_2 are the two terminal sectors, then $c(S_1) = c(S_2)$. Add to each node μ of ρ a third label, its *Herbrand label*, as follows. If μ is the first node of the track, then by Definition 6.2.4(a)(2) the Skolem label of μ is 1-part, and by inspection of the Skolem pieces we see that every 1-part Skolem label is a sentence in the Herbrand theory; we take it to be the Herbrand label of μ. If μ is a successor node, the Herbrand label of μ is got from the Skolem label of μ by removing any universal quantifier and then replacing any free variable by $c(S)$, where S is the sector containing μ^-. By this construction each of these Herbrand labels is a sentence in Hr(T).

If we list in order the Herbrand labels of the nodes in a track, they form a sequence

$$\phi_0, (\phi'_0 \to \phi_1), (\phi'_1 \to \phi_2), \ldots, (\phi'_{n-2} \to \phi_{n-1}). \tag{6.2.4}$$

correlated to the n successive nodes $\mu_0, \ldots, \mu_{n-1}$ of the track.

Claim Two. For each $i < n$, ϕ_i and ϕ'_i are the same sentence.

Proof of Claim. By the First Propositional Condition, ϕ_i and ϕ'_i carry the same relation symbol. Suppose $i < n-1$; we must show that ϕ_i and ϕ'_i have the same constant. There are two cases according as μ_i is or is not fixed.

Suppose μ_i is fixed at a constant c_i. Then μ_i is the nominator of a sector S, so $c_i = c(S)$, and by construction the constant substituted for variables in the Skolem label of μ_{i+1} is $c(S)$. Hence ϕ'_i has the same constant c_i as ϕ_i, and also the same relation symbol; so ϕ_i is ϕ'_i.

Suppose μ_i is not fixed; then it is echoic (hence a successor node), and so by construction ϕ_i has the same constant as ϕ'_{i-1}, namely $c(S)$ where S is the sector containing μ_{i-1}. But since μ_i is echoic, it also lies in S, and so by construction $c(S)$ is the constant in ϕ'_i. So again ϕ_i is ϕ'_i. □ Claim Two.

By Claim Two, the sequence (6.2.4) is a Herbrand sequence with outcome ϕ_{n-1}. Write ϕ for the outcome from the upper track and ϕ' for the outcome from the lower track. By the Second Propositional Condition, ϕ is affirmative and ϕ' is negative, and they have the same relation symbol. By the Constraint Condition and arguments like those of the proof of Claim Two, the constants in ϕ and ϕ' are the same. So ϕ' is $\neg\phi$. Since Hr(T) entails the outcomes of all Herbrand sequences, it entails both ϕ and $\neg\phi$, proving that Hr(T) is inconsistent. So by Lemma 2.2.9, T is inconsistent. □

The requirement that Σ^{sk} contains an individual constant is innocent. As noted in Definition 2.1.19, adding symbols to the signature has no effect on the validity of sequents. In any case we see from the Skolem pieces in Figure 5.1.1 that Σ^{sk} always contains individual constants unless T consists entirely of sentences of the form (e).

Corollary 6.2.6 *For every minimal inconsistent theory T there is a Skolem refutation ρ whose theory is T.*

Proof. By Theorem 6.2.5 there is a Skolem refutation ρ with Th(ρ) $\subseteq T$. But also by Theorem 6.2.5, Th(ρ) is inconsistent. Since T is minimal inconsistent it follows that Th(ρ) $= T$. □

6.3 Example and notation

Example 6.3.1 Consider the following set of assertoric sentences:

(6.3.1) $$(a)(C,B),(a)(B,A),(e)(C,A).$$

What does a Skolem refutation of this set look like? We first form its skolemisation using Figure 5.1.1:

(6.3.2) $$\forall x(Cx \to Bx),\ Ca,\ \forall x(Bx \to Ax),\ Bb,\ \forall x(Cx \to \neg Ax).$$

So the Herbrand theory is

(6.3.3) $$\begin{array}{l} Ca, Bb, \\ (Ca \to Ba),\ (Cb \to Bb),\ (Ba \to Aa), \\ (Bb \to Ab),\ (Ca \to \neg Aa), (Cb \to \neg Bb). \end{array}$$

We find the following Herbrand sequences:

(6.3.4) $$\begin{array}{l} Ca,\ (Ca \to Ba),\ (Ba \to Aa). \\ Ca,\ (Ca \to \neg Aa). \\ Bb,\ (Bb \to Ab). \end{array}$$

together with all initial segments of these sequences. The resulting outcomes are:

(6.3.5) $$Ca, Ba, Aa, \neg Aa, Bb, Ab.$$

The list contains both Aa and $\neg Aa$. Selecting Herbrand sequences with these two outcomes gives us two tracks; we write the theory label of a node above it and the Skolem label below it:

(6.3.6)
$$\begin{array}{ccccc}
(a)(C,B) & & (a)(C,B) & & (a)(B,A) \\
\bullet & \longrightarrow & \bullet & \longrightarrow & \bullet \\
Ca & & \forall x(Cx \to Bx) & & \forall x(Bx \to Ax) \\
(a)(C,B) & & (e)(C,A) & & \\
\bullet & \longrightarrow & \bullet & & \\
Ca & & \forall x(Cx \to \neg Ax) & &
\end{array}$$

We can improve the notation of (6.3.6). First, it suffices to give at each node the theory label and an indication of which Skolem piece is the Skolem label. So above each node we can write $\phi_{[n]}$ where ϕ is the theory label and the Skolem label is the n-th of its Skolem pieces as listed in Figure 5.1.1. This leaves the space below each node free for us to record the relation symbol carried by that node, which is a useful guide to the eye:

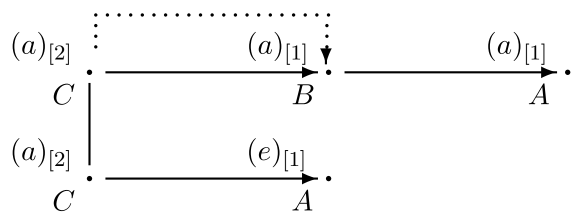
(6.3.7)
$$\begin{array}{ccccc}
(a)(C,B)_{[2]} & & (a)(C,B)_{[1]} & & (a)(B,A)_{[1]} \\
\bullet & \longrightarrow & \bullet & \longrightarrow & \bullet \\
C & & B & & A \\
(a)(C,B)_{[2]} & & (e)(C,A)_{[1]} & & \\
\bullet & \longrightarrow & \bullet & & \\
C & & A & &
\end{array}$$

There is some redundancy in (6.3.7). At top right the relation symbols below the nodes show that the subject and predicate symbols of the theory label of the terminal node are respectively B and A, so we can write simply $(a)_{[1]}$ in place of $(a)(B,A)_{[1]}$. The same applies at all the primary nodes, i.e. the nodes with first Skolem piece. For the secondary nodes we can't simply leave off the tag, because it is needed to show the predicate symbol. However, each secondary node μ in this example has the same theory label as some primary node ν, so the required information can be given by showing the connection between μ and ν.

For this purpose we introduce two new pieces of symbolism. First, if the i-th node in the upper track has the same theory and Skolem labels as the i-th node in the lower track, we draw a solid vertical line connecting these two nodes. In (6.3.7) there will be such a line between the first nodes of the two tracks. And second, where a secondary node ν has the same theory label as a primary node μ, and this is not otherwise indicated, we can indicate it by drawing a dotted arrow from μ to ν. Note the direction of the arrow: it will go from a node with the subject symbol to a node with the predicate symbol, exactly as the solid arrows between primary nodes.

Adding these vertical lines and dotted arrows, and then removing the tags from the sentences, changes (6.3.7) to the following:

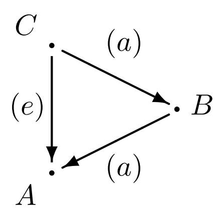
(6.3.8)



(6.3.8) will be our preferred way of writing the Skolem refutation (6.3.6).

In (6.3.8), every relation symbol used in the theory (6.3.1) is carried by at least one node, and every sentence of (6.3.1) is the theory label of some primary node with an arrow pointing to it. This tells us that (6.3.8) contains everything needed for constructing the subject-predicate digraph of (6.3.1). For this we need to remove some duplications. Thus the symbol A appears at both the terminal nodes; these nodes must be identified. Likewise a pair of nodes at the lefthand side carrying the symbol C must be shrunk to a single node. The end result is the digraph

(6.3.9)



Definition 6.3.2 Let ρ be a Skolem refutation.

(a) Suppose that for some i, the i-th node in the upper track of ρ has the same theory and Skolem labels as the i-th node in the lower track. Then we say that these two nodes are *twins*.

(b) There is a set N nodes of ρ that is maximal with the properties

(i) all the nodes in N are twins, and

(ii) the nodes of N in each track are an initial segment of that track.

We call N the *tail* of ρ; the *length* of the tail is the number of nodes in the upper track that are in the tail. The *final* nodes of the tail are the two nodes that are rightmost in the parts of the tail in the upper and lower tracks. If ρ has no twins we say that ρ has a *tail of length* 0, or (abusing language) that ρ *has no tail*.

Warning: These are definitions for the tail of a refutation. We will see that subject-predicate graphs can have tails too, and for these the definitions are not quite the same (though they are related).

Thus the refutation (6.3.8) has a tail of length 1.

Definition 6.3.3 The subject-predicate digraph (6.3.9) has a unique node that is the target of an arrow whose label is a negative sentence. It follows from the Rule of Quality (Corollary 5.3.6) that the subject-predicate digraph of an assertoric MIT always has a unique node with this property. We call this node the *contradictory node*. (The name refers to the fact that this node always represents the two terminal nodes of a refutation.)

6.4 Terse refutations

Could there be twins that are not in the tail? Yes there could be, but if this happens it betrays a redundancy in the refutation. If μ and ν are twins, then we can always replace the lower track up to ν by a copy of the upper track up to μ, so that the nodes up to μ and ν fall into the tail.

This is a cue to us to simplify refutations by removing redundancies. For this we will copy a standard formalism. We will introduce a well-founded relation $\prec$ on refutations. If $\rho' \prec \rho$, this indicates that ρ' contains fewer redundancies than ρ. For example if the theory of the refutation ρ' is a proper subset of the theory of the refutation ρ then $\rho' \prec \rho$. (Of course Gentzen and Prawitz are in the background here, but our simplifications are not special cases of theirs.)

Definition 6.4.1 Let ρ be a refutation.

(a) We write Th(ρ) for the theory of ρ.

(b) We write $NS(\rho)$ for the *node size* of ρ, i.e. the number of nodes of ρ counting a pair of twins in the tail as two.

(c) We write $TL(\rho)$ for the *tail length* of ρ, i.e. the number of nodes in the tail of ρ counting a pair of twins in the tail as one.

(d) We write $RK(\rho)$ for the *rank* of ρ, meaning $NS(\rho) - TL(\rho)$; this equals the number of nodes of ρ counting a pair of twins in the tail as one.

Definition 6.4.2 (a) We define a binary relation $\prec^\star$ on the class of refutations, as follows. Let ρ and ρ' be refutations. Then $\rho' \prec^\star \rho$ if and only if

- Either Th(ρ') is a proper subset of Th(ρ),
- or Th(ρ') = Th(ρ) and $RK(\rho') < RK(\rho)$.

(b) We define a binary relation $\prec$ on the class of refutations, as follows. Let ρ and ρ' be refutations. Then $\rho' \prec \rho$ if and only if there is a finite sequence of refutations $\rho_1, \ldots, \rho_n$ ($n \geqslant 2$) such that

$$\rho' = \rho_1 \prec^\star \ldots \prec^\star \rho_n = \rho.$$

Lemma 6.4.3 *(a) The relation $\prec$ is irreflexive, and it is the transitive closure of the relation $\prec^\star$.*

(b) $\rho' \prec^\star \rho \;\Rightarrow\; \rho' \prec \rho \;\Rightarrow\;$ Th(ρ') $\subseteq$ Th(ρ).

(c) The relation $\prec$ is well-founded, i.e. for every nonempty class $\mathbb{K}$ of refutations there is a refutation $\rho \in \mathbb{K}$ such that for no $\rho' \in \mathbb{K}$ do we have $\rho' \prec \rho$.

Proof. (a) and (b) are left to the reader.

(c) Suppose for contradiction that $\mathbb{K}$ is a nonempty class of refutations, and for every refutation ρ in $\mathbb{K}$ there is a refutation ρ' in $\mathbb{K}$ with $\rho' \prec \rho$. Then by the axiom of choice there is a sequence $(\rho_n : n \in \omega)$ of refutations in $\mathbb{K}$ such that for each n, $\rho_{n+1} \prec \rho_n$. So by the definition of $\prec$ there is a sequence $(\rho'_n : n \in \omega)$ of refutations, not necessarily in $\mathbb{K}$, such that for each n, $\rho'_{n+1} \prec^\star \rho'_n$.

Then by (b), for each n, Th(ρ'_{n+1}) $\subseteq$ Th(ρ'_n). Since each refutation has a finite theory, there must be some n_0 and some theory T such that for each $n \geqslant n_0$, Th(ρ'_n) $= T$. Then for each $n \geqslant n_0$, $RK(\rho'_{n+1}) < RK(\rho'_n)$. But this is impossible since $RK(\rho'_{n_0})$ is finite. This contradiction proves (b). □

Definition 6.4.4 We say that a refutation ρ is *terse* if there is no refutation ρ' with $\rho' \prec \rho$.

Theorem 6.4.5 *(a) If ρ is a terse refutation then the theory of ρ is minimal inconsistent.*

(b) Every minimal inconsistent assertoric theory is the theory of some terse refutation. Moreover for every refutation ρ with minimal inconsistent theory, either ρ is terse or there is a terse $\rho' \prec \rho$ with $Th(\rho') = Th(\rho)$.

(c) An assertoric theory is minimal inconsistent if and only if it is the theory of some terse refutation.

Proof. (a) Let ρ be a terse refutation and T the theory of ρ. By Theorem 6.2.5 above, T is inconsistent. Suppose for contradiction that T is not minimal inconsistent. Then there is a proper subset T' of T which is also inconsistent, and by Theorem 6.2.5 again, there is a refutation ρ' whose theory $\subseteq T'$. Then $\rho' \prec \rho$, contradicting the terseness of ρ.

(b) We prove the second sentence first. Let T be a minimal inconsistent theory and ρ a refutation with theory T. Let $\mathbb{K}$ be the class of all refutations ρ' such that (1) T is the theory of ρ' and (2) either ρ' is ρ or $\rho' \prec \rho$. By assumption $\mathbb{K}$ is not empty. So by Lemma 6.4.3(b) there is a refutation $\rho' \in \mathbb{K}$ such that $\rho'' \not\prec \rho'$ for every $\rho'' \in \mathbb{K}$. We show that ρ' is terse. If it is not, then there is a refutation ρ'' such that $\rho'' \prec^{\star} \rho'$. By Lemma 6.4.3(c), the theory of ρ'' is a subset of that of ρ', which is T. But T is minimal inconsistent, so the theory of ρ'' is T and hence $\rho'' \in \mathbb{K}$, contradicting the choice of ρ'. This proves the second sentence of (b). Now if T is any minimal inconsistent theory, then T is the theory of some refutation ρ by Corollary 6.2.6, and so we can apply the second sentence to ρ.

(c) follows from (a) and (b). □

By Theorem 6.4.5, an assertoric theory is minimal inconsistent if and only if it is the theory of some terse refutation. We haven't shown that the terse refutation is always unique, but nevertheless we will see in the next few sections that the classification of terse refutations is useful for classifying minimal inconsistent theories.

We will use the notion of terse refutations to prove that every inconsistent theory in assertoric logic has a refutation with no relation symbol occurring at two affirmative nodes. This will not be true for two-dimensional logic. The calculations of the invariants are no doubt avoidable in this logic, but they will transfer exactly to two-dimensional refutations and save us some effort later.

Definition 6.4.6 The following definition is for $\mathcal{L}_{\text{as+uas}}$, and will need to be revised when we move to 2D logic. Let ρ be a refutation, μ a node of ρ and c a constant.

(a) We say that μ is *left constrained to* c if the nominator of the track of μ is fixed at c, and *right constrained to* c if the nominator of the other track is fixed at c.

(b) We say that another node ν of ρ is *left constrained away from* μ if ν is left constrained to a constant and μ is right constrained to a different constant.

Our first application deals with the case where a relation symbol B occurs twice in a single track, and the later occurrence is outside the tail.

Lemma 6.4.7 *Let ρ be a terse refutation, and μ and ν nodes of ρ such that μ and ν are respectively the i-th and j-th nodes of the same track, with $i < j$. Suppose that μ and ν are affirmative and the same relation symbol is at both nodes. Then ν is in the tail.*

Proof. We proceed by contradiction. Assume that ν is not in the tail. Let ρ' be the diagram formed from ρ by deleting the nodes after μ and up to ν inclusive. Then ρ' is a diagram in the sense of Definition 6.2.4(a); we will show that it is also a refutation in the sense of Definition 6.2.4. Then we will show that $\rho' \prec \rho$. This will contradict the terseness of ρ.

The Propositional Conditions require that the relation symbol B at μ in ρ' is the same as the subject symbol A of the Skolem label of ν^+ in ρ'. By the Propositional Conditions in ρ, A is the same as the relation symbol at ν, and by assumption this symbol is B. The other requirements of the Propositional Conditions also hold in ρ' because they held in ρ.

The Constraint Condition holds in ρ' because it held in ρ and the two diagrams have the same nominators.

Since we have proved that conditions (2)–(4) of Definition 6.2.4 hold in ρ', it follows that ρ' is a refutation.

We show that $\rho' \prec \rho$ by proving the stronger statement that $\rho' \prec^\star \rho$. Since every node of ρ' was already a node of ρ, the theory of ρ' is included in that of ρ. So we look next at the invariant RK. Since ν is not in the tail,

$j > TL(\rho)$, so we have
(6.4.1)

$$\begin{array}{rcl} NS(\rho') & = & NS(\rho) - (j-i); \\ TL(\rho') & \geqslant & \min\{TL(\rho), i\}. \\ RK(\rho') & = & NS(\rho') - TL(\rho') \\ & \leqslant & (NS(\rho) - (j-i)) - \min\{TL(\rho), i\} \\ & = & \max\{NS(\rho) - (j-i) - TL(\rho), NS(\rho) - (j-i) - i\} \\ & = & \max\{RK(\rho) - (j-i), RK(\rho) - (j - TL(\rho))\} \\ & < & RK(\rho). \end{array}$$

Hence $\rho' \prec^\star \rho$ as required. □

For our second application we turn to some situations where an initial segment of a refutation can be removed, using an existential augment from a later theory label.

Definition 6.4.8 Let ρ be a refutation and μ, ν affirmative nodes in different tracks of ρ but with the same relation symbol A. By a *support* for μ and ν we mean a sentence of the form (a) which occurs as theory label either of a node later than μ in its track, or later than ν in its track, and has A as its subject symbol. We say that μ and ν are *supported* if they have a support.

Lemma 6.4.9 *Let ρ be a terse refutation and μ, ν nodes in different tracks. Suppose both μ and ν carry the same relation symbol, and μ and ν are supported. Then the following hold:*

(i) μ and ν are twins in the tail.

(ii) μ and ν are the two initial nodes.

(iii) The theory label of μ and ν is also the theory label of some successor node.

Proof. By assumption there is a support ψ for μ and ν. Let ρ' be the diagram formed from ρ by removing all the nodes up to and including μ and ν, and replacing them with new initial nodes μ°, ν° which both carry the Skolem label $\psi_{[2]}$.

The Propositional Conditions hold in ρ' because they held in ρ, and μ° and ν° replace nodes that had in ρ the same relation symbol as μ° and ν° have in ρ'. The Constraint Condition holds in ρ' because its nominators, namely μ° and ν°, have the same Skolem label and hence are fixed at the same constant. It follows that ρ' is a refutation.

We show that if any of (i)–(iii) of the lemma fail then we have $\rho' \prec^\star \rho$ and hence $\rho' \prec \rho$.. Suppose μ and ν are respectively the i-th and j-th nodes in their tracks of ρ, with $i \leqslant j$.

If (i) fails then μ and ν are not twins in the tail of ρ. Then $TL(\rho) < i$ and so we calculate

$$\begin{array}{rcl} NS(\rho_\psi) & = & NS(\rho) - i - j + 2; \\ TL(\rho_\psi) & \geqslant & 1. \\ RK(\rho_\psi) & = & NS(\rho_\psi) - TL(\rho_\psi) \\ & \leqslant & (NS(\rho) - i - j + 2) - 1 \\ & = & (NS(\rho) - TL(\rho)) + (TL(\rho) - i - j + 1) \\ & < & RK(\rho) + (i - i - j + 1) \\ & \leqslant & RK(\rho). \end{array} \tag{6.4.2}$$

So $\rho_\psi \prec \rho$.

If (i) holds but (ii) fails, then μ and ν are twins in the tail but are not initial. So $i = j > 1$ and $TL(\rho_\psi) = TL(\rho) - i + 1$, and we calculate

$$\begin{array}{rcl} NS(\rho_\psi) & = & NS(\rho) - 2i + 2; \\ TL(\rho_\psi) & = & TL(\rho) - i + 1. \\ RK(\rho_\psi) & = & NS(\rho_\psi) - TL(\rho_\psi) \\ & \leqslant & (NS(\rho) - 2i + 2) - (TL(\rho) - i + 1) \\ & = & (NS(\rho) - TL(\rho)) + (-2i + 2 + i - 1) \\ & = & RK(\rho) + (1 - i) \\ & < & RK(\rho). \end{array} \tag{6.4.3}$$

So again $\rho_\psi \prec \rho$.

If (iii) fails then the theory of ρ' is a proper subset of that of ρ, so again $\rho' \prec \rho$.

In each case we have contradicted the terseness of ρ. This proves the lemma. □

Lemma 6.4.9 depends for its applications on there being supports in the appropriate places. The following lemma is useful for finding these supports.

Lemma 6.4.10 *Let ρ be a Skolem refutation in $\mathcal{L}$. Suppose μ and ν are affirmative nodes in the upper and lower tracks of ρ respectively, and the same relation symbol A is at both μ and ν. Then there is an affirmative node which is either later than μ in its track, or later than ν in its track, and carries the relation symbol A.*

Proof. Let ξ be the rightmost affirmative node of the upper track such that the relation symbol at each of μ, ξ and the nodes between them is A; let ζ be defined likewise from ν in the lower track. Since ζ is affirmative, it is not the terminal node of the lower track.

If ζ^+ is not the terminal node of the lower track, then the Skolem label of ζ^+ is affirmative and has subject symbol A and predicate symbol distinct from A. A Skolem piece with these features must be the first Skolem piece of a sentence of the form (a) or (a, uas). So the theory sentence of ζ^+ must be of the form (a) or (a, uas) with subject symbol A as required.

On the other hand if ζ^+ is the terminal node of the lower track, then again it carries a relation symbol B distinct from A, and so the terminal symbol of the upper track also carries B. Therefore ξ is not the terminal node of the upper track, and we can use the Skolem label of ξ^+ by an argument similar to that of the previous paragraph. □

6.5 Describing the assertoric MITs

We will describe the assertoric MITs by describing their terse refutations, and then checking whether the same assertoric MIT can have more than one terse refutation.

As above, we put $\mathcal{L} = \mathcal{L}_{\text{as+uas}}$. We say '$\mathcal{L}$ without f' for the logic which has the six sentence forms of $\mathcal{L}$ except for the form or forms f.

Lemma 6.5.1 *Let ρ be a refutation in $\mathcal{L}$ without (o). Then:*

(a) The negative node has a theory label of the form (o, uas) if it is initial, and (e) if it is a successor.

(b) Apart from the negative node, every successor node has a theory label of the form (a) or (a, uas).

(c) If there are no sentences of the form (a, uas) in the theory of ρ, then only one negative sentence appears in the theory of ρ.

Proof. The calculations are made from the Skolem pieces in Figure 6.1

(a) If the negative node is initial then its theory label has a 1-part primary Skolem piece and hence has the form (o, uas). If the negative node is a successor then its theory label has a 2-part primary Skolem piece; since we have excluded the form (o), the theory label must have the form (e).

(b) Every affirmative successor node has a 2-part affirmative Skolem label. Since we have excluded sentences of the form (o), these nodes must have theory labels of the form (a) or (a, uas).

If (c) fails, then some negative sentence ψ is theory label of some node ν, but its negative Skolem piece is not the Skolem label of any node. Hence ψ has a secondary Skolem piece, and so ψ has the form (o, uas). The secondary piece is 1-part, so ν is initial. But both initials are fixed at the same constant, so ν is in the tail, and its twin μ in the tail also has the Skolem label $\psi_{[2]}$. But ψ is not also the theory label of any successor node, since it has no 2-part Skolem pieces. So by Lemma 6.4.9, there is no support for μ and ν. The affirmative node provided by Lemma 6.4.10 is a successor node and so of the form either (a) or (a, uas); so by the assumption in (c), it has the form (a). So by Definition 6.4.8, the theory label of this node provides a support for μ and ν. Therefore by Lemma 6.4.9, the theory label ψ on μ and ν is also the theory label of a successor node, contradicting that ψ has the form (o, uas). □

To progress further we need to partition the refutations in some meaningful way. The following lemma indicates how.

Lemma 6.5.2 *Let ρ be a refutation in $\mathcal{L}$ without (o), and let μ and ν be the nominator nodes of ρ. Then μ and ν carry the same theory label.*

Proof. Since the two terminal nodes are fixed at the same constant, and the Skolem pieces of distinct theory labels never carry the same constant, the theory labels of μ and ν must be the same. □

Definition 6.5.3 Let ρ be a refutation in $\mathcal{L}$ without (o). The theory label of the two nominator nodes of ρ is called the *nominator sentence* of rho, and the form of the nominator sentence is called the *nominator type* of ρ.

We can now complete the description of the terse refutations in the case of $\mathcal{L}$ without (a, uas) or (o). Recall from Corollary 6.1.7 that the classification of the MITs for $\mathcal{L}$ without (a, uas) or (o) exactly matches the classification of the MITs for $\mathcal{L}_{\text{as}}$.

Theorem 6.5.4 *Let ρ be a refutation in $\mathcal{L}$ without (o). Then exactly one of the following holds.*

Case (**i**) *Exactly one of the sentences in the theory of ρ has the form (i), and this sentence is the nominator sentence.*

Case (**o**) *Exactly one of the sentences in the theory of ρ has the form (o, uas), and this sentence is the nominator sentence and the theory label of the negative node.*

Case (**a**) *The theory of ρ contains no sentence of either of the forms (i) and (o, uas), and the nominator sentence has the form (a).*

Proof. The only sentences with Skolem pieces that are fixed at a constant are those of the forms (a), (i) and (o, uas), so only sentences of these forms can be nominator sentences. We have already noted that the nominator nodes must be the initial nodes. So the Skolem labels of the initial nodes must be from the same sentence χ of one of the three forms (a), (i) and (o, uas). If χ has the form (i), this leaves no node that can carry a theory label of the form (o, uas), and vice versa. So at most one of the sentences in the theory of ρ has the form (i) or (o, uas), and any such sentence is the nominator sentence. □

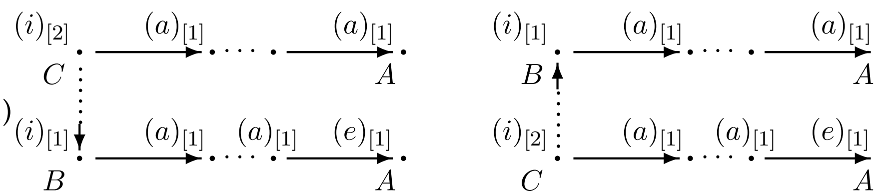
The Theorem shows that the nominator type of the refutation ρ is determined by the theory of ρ. So the classification by nominator types applies directly to the MITs themselves.

We can complete the details of the refutations not using (a, uas) or (o) in short order.

Case (**i**). In this case the Skolem labels on the two initial nodes are Skolem pieces of the nominator sentence χ of the form (i). We show that each Skolem piece is the Skolem label of one of the initial nodes. If not, then the two initial nodes carry the same relation symbol, and so by Lemma 6.4.10 and the absence of sentences of the form (a, uas), they are supported. But then by Lemma 6.4.9, χ is also the theory label of some successor node, which is impossible since χ has no 2-part Skolem pieces. Hence each of the Skolem pieces of χ is the Skolem label of one of the initial nodes. Examples show that the primary initial node can be either the lower as in $(i)\downarrow$, or the upper as in $(i)\uparrow$. The negative node carries a label of the form (e) by Lemma 6.5.1(a); so the lower track has at least two nodes.

In this case the theory is optimal inconsistent. Any weakening would replace an (a) sentence by an (i) sentence, or an (e) sentence by an (o, uas) sentence, and either of these replacements would violate Theorem 6.5.4.

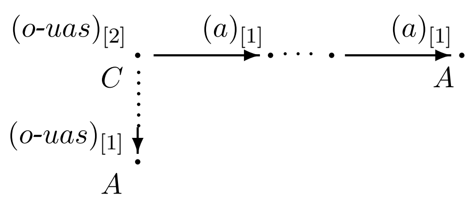
Using the conventions of Section 6.3, we can write the refutation in one of the two forms

(6.5.1)


Case (**o**). In this case the Skolem labels on the two initial nodes are Skolem pieces of the nominator sentence χ of the form (o, uas). The first Skolem piece is negative and so can only be the Skolem label of the negative node. But could both initial nodes have as Skolem label the second Skolem piece of χ? The answer is No by the same argument as for the case of (i). So the Skolem labels of the initial nodes are the second Skolem piece on the upper node and the first Skolem piece on the lower node; and the lower node is the negative node, so that the lower track has just one node.

The theory in this case is optimal inconsistent for the same reason as in Case (i).

Using the conventions of Section 6.3, we can write the refutation in the form

(6.5.2)


Case (**a**). This case results when no sentence in the theory of ρ is existential. The nominator sentence χ has the form (a) with only one 1-part Skolem piece. This piece must be the Skolem label of both initial nodes, so that these initial nodes μ and ν are twins in the tail and carry the same relation symbol (namely the subject symbol of χ). By Lemma 6.4.10, μ and ν are supported. So by Lemma 6.4.9, χ is also the theory label of the second node of at least one of the two tracks. Then χ is not the theory label of the second node of the other track too, since otherwise the two second nodes would have the same Skolem label too and hence be twins; then another application of Lemmas 6.4.10 and 6.4.9 would show that these two second nodes are initial, a contradiction. So the tail has length 1. Since no sentence of the form (o, uas) occurs, the theory label on the negative node has the form (e) and so the lower track has at least two nodes.

The theory in this case is never optimal inconsistent; weakening the first sentence of the upper track to an (i) sentence converts to Case (i). The refutation (6.3.8) was a typical example of this case.

Lemma 6.5.5 *If ρ is a terse refutation in $\mathcal{L}$ without (a, uas) or (o), then no two distinct affirmative sentences in the theory of ρ have the same predicate symbol.*

Proof. Suppose two affirmative sentences $(f)(B, A)$ and $(g)(C, A)$ occur in the theory of ρ. These sentences must be the theory labels of two affirmative nodes μ, ν carrying the same relation symbol A. By Lemma 6.4.9, μ and ν must both be in the tail; from the cases listed above it follows that μ and ν are the two initial nodes. But we have seen that in all cases the two initial nodes carry the same theory label. □

EXPAND THE FOLLOWING TO AN EXPLICIT ALGORITHM.
The Lemma shows how we can reconstruct the subject-predicate digraph from the refutation ρ. The predicate symbol of the unique negative sentence is the relation symbol carried by the two terminal nodes; let μ be the upper terminal node, carrying the relation symbol A. There is at most one affirmative universal sentence $(a)(B, A)$ in the theory of ρ; if there is such a sentence then μ^- exists and carries the relation symbol B, and so on. Let ν be the lower terminal node, say with theory label $(e)(C, A)$; then ν^- exists and carries the relation symbol C, and so we can trace backwards from C in the same way. This process spells out the upper and lower tracks until we come to initial nodes. In Case (a) we eventually reach the same relation symbol in both tracks, and it is the symbol carried by the unique initial node of the subject-predicate digraph. In Cases (i) and (o) the theory of ρ contains a sentence whose subject and predicate symbols are the symbols carried by the two initial nodes of ρ; in the subject-predicate digraph this sentence provides an arrow from the subject symbol to the predicate symbol. (In source and target this arrow matches the dotted arrow that we added to indicate the secondary and primary nodes.)

In sum:

Theorem 6.5.6 *A set of assertoric sentences is minimal inconsistent if and only if it has a subject-predicate digraph of one of the types* $(i){\uparrow}$, $(i){\downarrow}$, (o) *or* (a) *as above. It is optimal inconsistent if and only if it has a subject-predicate digraph of one of the types* $(i){\uparrow}$, $(i){\downarrow}$ *or* (o).

6.6 Exercises

6.0. For each theory T in $\mathcal{L}_{\dashv\int}$, let $T^{\star}$ be the theory in $\mathcal{L}_{\sqcap\dashv\int}$ got by replacing (a) by (a, uas) and (o) by (o, uas). Show that T is consistent if and only if $T^{\star}$ has a model M in which all relations A^M are nonempty. [This gives an alternative reading of Aristotle's assertoric sentences, as for example in Strawson REF.]

6.1. Show that in a Skolem track, at most the terminal node is negative.

6.2. Prove the existence and properties of canonical models as defined in Lemma 2.1.15. [Referring to Lemma 2.1.15, assume Σ has at least one individual constant. (If not, show that one can be added and the same result derived.) Let D be the set of all closed terms of Σ. Construct M with domain D. For each individual constant c put $c^M = c$, and for each function symbol f, say 1-ary, define f^M by putting $f^M(c) = f(c)$ for each individual constant c. Use a similar idea for the relation symbols.]

6.2. Find sectors in some Skolem tracks, and zones in some refutations.

6.3. Find refutations of some inconsistent sets.

6.4. One of the other miminising cases.

6.5. Show that in a terse refutation all twins lie in the tail. (This will follow from BELOW, but the exercise is to prove it directly from WHAT.)

6.6. Show that in a Skolem refutation an affirmative node and a negative node never have the same theory label.

6.7. Putting these steps into a computer program is very straightforward. If the initial set of sentences is consistent, this will show up when the outcomes are listed; there will be no atomic sentence such that both it and its negation are outcomes. In this case the Canonical Model Theorem will give a model of the theorem. We can bound the size of this model: if there are n sentences then the Herbrand universe will have cardinality at most n. Also the number of relation symbols will be at most $2n$, so the total number of literals to be considered is $\leqslant 2 \times 2n \times n = 4n^2$. So we have here a reasonably practical decision procedure for assertoric logic.

6.8. Verify the skolemisations in Figure 6.1.

6.9. Test your understanding of the proof of Lemma le:3.4.9 by proving the following, which is almost a generalisation of the lemma. Let T and U be theories of $\mathcal{L}_{uas}$, both in the language $L(\Sigma)$. Suppose that T is consistent, $T \subset U$, and every circular subgraph of $\Gamma(U)$ is a subgraph of $\Gamma(T)$. Prove that U is consistent, and that it has a model satisfying $(\exists x Ax \wedge \exists x \neg Ax)$ for every relation symbol A in Σ. (By Exercise 3.8 we can't put a finite bound on the cardinality of the model.)

6.10. Give necessary and sufficient conditions for T to be extendable to a minimal inconsistent theory.

6.11. Show that an assertoric refutation is terse if and only if it satisfies the condition $(\star)$ in one of Theorems th:6.1.5, th:6.1.5 and th:6.1.6. [Let T be the theory of ρ. Suppose for example that ρ is a refutation as in Theorem th:6.1.5, but is not terse. Then there is a refutation $\rho' \prec^\star \rho$. By the definition of $\prec^\star$, either the theory of ρ' omits some sentence in T, or Th($\rho' = T$ but $RK(\rho') < RK(\rho)$. The first violates the fact that every assertoric refutation is graph-circular. The second violates the fact (using Theorem th:6.1.5) that either ρ' or a terse refutation $\prec \rho'$ has rank $k+1$ where k is the number of distinct relation symbols used in T.]

6.12. Something about moods where the contradictory node has only a single arrow to it.

Chapter 7

Productivity conditions

7.1 Distribution

Definition 7.1.1 We describe a relation symbol occurring in an assertoric sentence ϕ as either *distributed in* ϕ or *undistributed in* ϕ, according to the following chart:

(7.1.1)

ϕ	B	A
$(\forall x(Bx \to Ax) \wedge \exists x Bx)$	distributed	undistributed
$\forall x(Bx \to \neg Ax)$	distributed	distributed
$\exists x(Bx \wedge Ax)$	undistributed	undistributed
$(\exists x(Bx \wedge \neg Ax) \vee \forall x \neg Bx)$	undistributed	distributed

These notions are simply a rearrangement of notions that we have already introduced. The following consequences are immediate:

Lemma 7.1.2 *Let ϕ be an assertoric sentence and C a relation symbol occurring in ϕ.*

(a) If C is subject symbol in ϕ, then C is distributed in ϕ if ϕ is universal and undistributed in ϕ if ϕ is existential.

(b) If C is predicate symbol in ϕ, then C is distributed in ϕ if ϕ is negative and undistributed in ϕ if ϕ is affirmative.

□

Corollary 7.1.3 *A symbol is distributed in the contradictory negation of ϕ if and only if it is undistributed in ϕ.* □

Corollary 7.1.4 *Let each of the words X and Y be either 'distributed' or 'undistributed'. Then there is a unique assertoric sentence $\phi(B, A)$ such that B is X in ϕ and A is Y in ϕ.* □

The notions of distributed and undistributed carry over to the ends of a premise-sequence as follows.

Definition 7.1.5 Let $T[B, A]$ be an assertoric premise-sequence.

(a) Since B occurs in only one sentence of $T[B, A]$, we can define B to be *distributed in* $T[B, A]$ if it is distributed in the sentence of T containing it, and *undistributed in* $T[B, A]$ otherwise. Similarly with A.

(b) By Corollary 7.1.4 there is a unique assertoric sentence $\chi(B, A)$ such that B is distributed in $\chi(B, A)$ if and only if it is distributed in $T[B, A]$, and A is distributed in $\chi(B, A)$ if and only if it is distributed in $T[B, A]$. We call this sentence $\chi(B, A)$ the *mate* of $T[B, A]$.

History 7.1.6 Nothing like the notion of distribution appears anywhere in Ibn Sīnā's logic. This fact is part of the evidence that he lacked any coherent notion of positive and negative occurrences. A related notion seems to have first appeared in Scholastic logic around the beginning of the 13th century as a classification of kinds of quantification. By the time of Buridan it had become clear that one could usefully extend the classification to predicates too. In fact Buridan's definition of *suppositio confusa et distributiva* ([15] 4.3.6, cf. Read [85]) yields the table of Definition 7.1.1; translated into our terminology, Buridan makes the symbol A distributed in the assertoric sentence ϕ if and only if for every model M of ϕ, any structure N got from M by taking A^N to be a singleton subset of A^M, leaving the rest the same, is again a model of ϕ. Though Buridan's definition agrees with ours on assertoric sentences, it leaves us in the dark about why distribution should have the properties that it does have. Some scholars (for example Colwyn Williamson in a talk) understandably asked whether the logical properties of distribution might be just a fluke. The question was a good one, but we have a negative answer in Lyndon's Interpolation Theorem together with a better definition of distribution; see [39].

We proceed to put these notions to work.

Theorem 7.1.7 *Let T be a finite set of assertoric sentences. Then (a) and (b) below are equivalent:*

(a) T is optimal inconsistent.

(b) (i) T is graph-circular.

(ii) *T contains exactly one negative sentence.*

(iii) *Every relation symbol occurring in two sentence of T is distributed in exactly one of the two sentences.*

Proof. Let Γ be the subject-predicate digraph of T.

(a) $\Rightarrow$ (b): Assume (a). Then (i) and (ii) hold by Corollaries 5.3.2 and 5.3.6(a).

For (iii) we consider two adjacent sentences ϕ, ψ in Γ, sharing a relation symbol C.

Case One: C is not at one of the two ends of a track. In this case the two sentences are in the same track, say ϕ before ψ in the track. Then (iii) says that if ϕ is affirmative then ψ is not universal, and that if ϕ is negative then ψ is not existential. But by Corollary 5.3.8(b) an existential sentence is always first in its track, and by Corollary 5.3.6(b) a negative sentence is always last in its track.

Case Two: Both tracks have positive length and C is initial in both tracks. Then (iii) says that the initial sentences of the two tracks are not both universal and not both existential. The case of both universal would have to be in type (a), which is not optimal inconsistent. The case of both existential contradicts Corollary 5.3.8(a).

Case Three: Both tracks have positive length and C is final in both tracks. Then (iii) says that the final sentences of the two tracks are not both affirmative and not both negative. The case of both affirmative contradicts Corollary 5.3.6(b), and the case of both negative contradicts Corollary 5.3.6(a).

Case Four: The upper track has length 0 and its only node has label C. Since T is optimal inconsistent, we are in type $(i)\downarrow$. Clause (iii) says that we don't have either (1) that the initial sentence of the lower track is existential and the final sentence of the lower track is affirmative, or that (2) the initial sentence of the lower track is universal and its the final sentence is negative. But (1) contradicts Corollary 5.3.6(b), and (2) contradicts the fact that the lower track in type $(i)\downarrow$ begins with an existential sentence.

(b) $\Rightarrow$ (a): We verify that if (i), (ii) and (iii) hold then T has type $(i)\uparrow$, $(i)\downarrow$ or (o). We do a calculation of distributivity. There are, say, N sentences in T, and hence $2N$ occurrences of relation symbols in the sentences. Of these occurrences, exactly N are distributed, by (iii). Now by (ii), $N-1$ of the sentences are affirmative, so their predicate symbols are undistributed, while 1 is negative and has distributed predicate symbol. It follows that

$N-1$ of the subject symbol occurrences are distributed and 1 is not; in other words, there is exactly one existentially quantified sentence. All the sentences except the negative and the existentially quantified (which may be the same sentence) are (a) sentences.

Now by (i) the digraph Γ is circular. Since the subject and predicate symbols of an assertoric sentence are distinct, T contains at least two sentences.

Suppose first that T contains just two sentences. Then by (ii) just one of the sentences is negative. If it is $(e)(B,A)$ then both subject and predicate are distributed, so by (iii) the other sentence is either $(i)(B,A)$ or $(i)(A,B)$; in the first case T has type $(i)\uparrow(1.1)$ and in the second case it has type $(i)\downarrow(0,2)$. If the negative sentence is $(o)(B,A)$ then its subject is undistributed and its predicate is distributed; since the other sentence is affirmative, by (iii) it has to be $(a)(B,A)$ and the type is $(o)(1)$.

Secondly, suppose that T has at least three sentences, so that there is at least one sentence of the form (a). Consider a maximal segment Δ of the digraph circle consisting entirely of (a) sentences. Each (a) sentence has distributed subject symbol and undistributed predicate symbol, so by (iii) the arrows in Δ all point the same way around the circle, and the source of the first arrow of Δ is distributed at the subject of its sentence, and the target of the last arrow of Δ is undistributed at the predicate of its sentence. By (iii) again, the segment has to be matched up at each end against sentences with the opposite distributivities.

One subcase is that T contains an (i) sentence and an (e) sentence. Then the source of the (a) segment must join up with the (i) sentence and the target of the (a) segment must join up with the (e) sentence. If there is no (a) sentence outside Δ, then the remaining two ends of the (i) sentence and the (e) sentence match, giving one of the types $(i)\uparrow(1,n)$, $(i)\uparrow(m,1)$ or retrograde $(i)\downarrow(0,n)$. On the other hand if there is another segment of (a) sentences, it must join up the remaining two ends of the (i) and (e) sentences, and again we are in type $(i)\uparrow$ or $(i)\downarrow$.

The remaining subcase is that T contains an (o) sentence but no (i) sentence or (e) sentence. Then the only possibility is that the initial node of Δ joins up with the undistributed subject symbol of the (o) sentence, and the final node of Δ joins up with the distributed predicate symbol of the (o) sentence, putting us in type (o). □

Theorem 7.1.8 *Let T be a finite set of augmented assertoric sentences. Then the following are equivalent:*

(a) T is minimal inconsistent.

(b) (i) *T is graph-circular.*

(ii) *T contains exactly one negative sentence.*

(iii) *Every relation symbol occurring in a sentence of T has at least one distributed occurrence.*

Proof. The proof is close to that of the previous theorem.

(a) ⇒ (b): Assume (a). Then by Theorem 5.3.1, T has one of the types $(i)\uparrow$, $(i)\downarrow$, (o) and (a). Given the proof of the previous theorem, we need only verify (iii) in the case of (a). The proof is exactly as before except in Cases Two and Four. In Case Two it can now happen that both the initial sentences are universal, so that the relation symbol C is distributed at both occurrences. In Case Four it can now happen that the initial sentence is universal and the final sentence is negative, again making C distributed at both occurrences.

(b) ⇒ (a): Assume (i), (ii) and (iii). The distributivity calculation now shows that either N or $N-1$ of the subject symbol occurrences are distributed, according as T doesn't or does contain an existential sentence. If there is an existential sentence then the previous calculation applies and puts us in type $(i)\uparrow$, $(i)\downarrow$ or (o) as before.

Suppose then that T has no existential sentence. Then every sentence is either an (a) sentence or an (e) sentence, and by (ii) there is just one (e) sentence.

If T has just two sentences, then one has the form $(e)(B, A)$ and the other can have either the form $(a)(B, A)$ or the form $(a)(A, B)$; the type of T is then either $(a)(1, 1)$ or $(a)(0, 2)$ respectively.

If T has more than two sentences, then again we consider a maximal segment Δ of the circle, consisting of arrows labelled with (a) sentences. As before, all the arrows point the same way round the circle. There must be fewer than three such segments, or two would meet at their undistributed final nodes and violate (iii). If there are two, their undistributed final nodes must be source and target of the (e) arrow. If one, then its two ends attach to the ends of the (e) arrow. In either of these cases we are in type (a). □

History 7.1.9 The results of this section are updated versions of classical results dating back to the late Middle Ages. See Keynes [67] p. 291 for Theorem 7.1.8 in the case where T has size 3, and Thom [99] Theorems on pages 181 and 184 for the theorem in general; Thom gives a proof-theoretic proof.

7.2 Productivity and conclusions

As we saw in Section 3.3 above, Ibn Sīnā addresses two questions to any given premise-sequence. He asks first 'Are you productive?', and second, 'If you are productive then what is your conclusion?' The previous section gives us tools for calculating the answers to both questions, for any assertoric premise-sequence. In Theorem 7.2.6 we use those tools to state necessary and sufficient conditions for an assertoric premise-sequence to be productive, and in Theorem 7.2.5 we give a rule for calculating the consequence of a productive assertoric premise-sequence.

But first we need to identify and disarm a troublesome case, as follows.

Definition 7.2.1 An assertoric premise-sequence of the form

(7.2.1) $$<(a)(A_1, A_0),\ (a)(A_2, A_1), \ldots, (a)(A_n, A_{n-1})> [A_0, A_n],$$

with $n \geqslant 1$, is said to be *goclenian*.

Every goclenian premise-sequence is retrograde. There is a unique goclenian premise-pair, namely that of *Bamalip*.

Lemma 7.2.2 *Let $T[A_0, A_n]$ be the goclenian sequence (7.2.1). Then $T[A_0, A_n]$ is productive and has conclusion $(i)(A_0, A_n)$.*

Proof. Clearly T entails $(a)(A_n, A_0)$, and hence also $(i)(A_0, A_n)$ by (a)-conversion. The only strengthening of $(i)(A_0, A_n)$ is $(a)(A_0, A_n)$, but T doesn't entail $(a)(A_0, A_n)$. (Take a structure M with $A_i^M = \{0\}$ for all i, $0 < i \leqslant n$, and $A_0^M = \{0, 1\}$.) □

The following characterisation of goclenian premise-sequences will be useful.

Lemma 7.2.3 *Let $T[B, A]$ be an assertoric premise-sequence. Then $T[B, A]$ is goclenian if and only if it satisfies the following three conditions:*

(i) Every relation symbol occurring in two sentences of $T[B, A]$ is distributed in at least one of the occurrences.

(ii) Every sentence in $T[B, A]$ is affirmative.

(iii) A is distributed in $T[B, A]$.

Proof. Inspection shows that a goclenian premise-sequence has all of the properties (i)–(iii). For the converse, suppose that (i)—(iii) hold and $T[B, A]$ has the form

(7.2.2) $$(\phi_0, \ldots, \phi_{n-1})$$

where for each ϕ_i the relation symbols in ϕ_i are A_i and A_{i+1}, so that A_0 is B and A_n is A.

Claim. For each i ($0 < i \leqslant n$), A_i is distributed in ϕ_{i-1}, and ϕ_{i-1} is $(a)(A_i, A_{i-1})$.

Proof of Claim. We prove the Claim by downwards induction on i. The base case is where $i = n$. Here A_n is A and hence is distributed in ϕ_{n-1} by (iii). Since ϕ_{n-1} is affirmative by (ii), it follows that ϕ_{n-1} is $(a)(A_n, A_{n-1})$. This proves the Claim when $i = n$.

Next we show that if the Claim holds for a given i with $1 < i$, then

(7.2.3) A_{i-1} is distributed in ϕ_{i-2}, and ϕ_{i-2} is $(a)(A_{i-1}, A_{i-2})$.

Since ϕ_{i-1} is $(a)(A_i, A_{i-1})$ by induction hypothesis, A_{i-1} is undistributed in ϕ_{i-1} and so by (i) it is distributed in ϕ_{i-2}, giving the first part of (7.2.3). The second part is given by the same argument as for the base case. □ Claim.

By the Claim, $T[B, A]$ is goclenian. □

History 7.2.4 In the post-medieval Aristotelian tradition, arguments whose listed premises are as in (7.2.1), with $n \geqslant 3$, went by the name of *goclenian sorites*. The conclusion was taken to be $(a)(A_n, A_0)$. The goclenian sorites is named after Rudolf Göckel (1547–1628), who held a chair of philosophy at the University of Marburg; see p. 257 of his [32]. (He is also credited with inventing the word 'psychology'.)

As promised, we can now find the conclusion of any productive assertoric premise-sequence.

Theorem 7.2.5 *Let $T[B, A]$ be a productive assertoric premise-sequence. Then:*

(a) If $T[B, A]$ is not goclenian then its conclusion is its mate (Definition 7.1.5(b)).

(b) If $T[B, A]$ is goclenian then its conclusion is $(i)(B, A)$, which is not its mate.

Proof. Since $T[B, A]$ is productive, it has a conclusion $\chi(B, A)$, and by Corollary 5.3.4 the antilogism

$$T[B, A] \cup \{\overline{\chi(B, A)}\} \tag{7.2.4}$$

is an assertoric MIT. It follows by Theorem 7.1.8 that if B is undistributed in $T[B, A]$ then it is distributed in $\overline{\chi(B, A)}$ and hence (by Corollary 7.1.3) undistributed in $\chi(B, A)$. Likewise if A is undistributed in $T[B, A]$ then it is undistributed in $\chi(B, A)$.

Claim One. B is distributed in $T[B, A]$ if and only if B is distributed in $\chi(B, A)$.

Proof of Claim. Right to left in the Claim has already been shown. For the converse, assume B is distributed in $T[B, A]$. Then Theorem 7.1.8 allows B to be either distributed or undistributed in $\chi(B, A)$. If χ is affirmative, this means that for (7.2.4) to be minimal inconsistent, χ can be either $(a)(B, A)$ or $(i)(B, A)$. But $(a)(B, A) \vdash (i)(B, A)$ and not conversely; so by Definition 3.3.2(b), $(a)(B, A)$ is a conclusion of $T[B, A]$ and $(i)[B, A]$ is not. Similarly if χ is negative and a conclusion of $T[B, A]$ then χ is $(e)(B, A)$. Either way, the conclusion $\chi(B, A)$ is universal, so that B is distributed in it.

□ Claim One.

Claim Two. If A is undistributed in $T[B, A]$ then it is undistributed in $\chi(B, A)$. If A is distributed in $T[B, A]$ but undistributed in $\chi(B, A)$, then $T[B, A]$ is goclenian.

Proof of Claim Two. The first sentence was proved in the first paragraph. For the second sentence, assume A is distributed in $T[B, A]$ but not in $\chi(B, A)$. Then A is distributed in $\overline{\chi(B, A)}$, so that $\overline{\chi(B, A)}$ is negative. Since (7.2.4) is minimal inconsistent, it follows from the Rule of Quality (Corollary 5.3.6(a)) that every sentence in $T[B, A]$ is affirmative. Also it follows from Theorem 7.1.8 that every relation symbol occurring in two sentences of $T[B, A]$ is distributed in at least one occurrence. So by Lemma 7.2.3, $T[B, A]$ is goclenian. □ Claim Two.

We deduce the Theorem as follows. For (a), if $T[B, A]$ is not goclenian, then it follows by the two Claims that $\chi(B, A)$ is the mate of $T[B, A]$. For (b), by Lemma 7.2.2 the conclusion is $(i)(B, A)$. This is not the mate of $T[B, A]$, since A is distributed in $T[B, A]$ but undistributed in $(i)(B, A)$. □

We can also find a necessary and sufficient condition for an assertoric premise-sequence to be productive.

Theorem 7.2.6 *Let $T[B, A]$ be an assertoric premise-sequence. Then the following are equivalent:*

(a) $T[B, A]$ is productive.

(b) The following both hold:

(i) Every relation symbol that occurs in two sentences of T is distributed in at least one of them, and

(ii) $T[B, A]$ contains at most one negative sentence; and if it contains a negative sentence then A is distributed in $T[B, A]$.

Proof. (a) ⇒ (b): Assume $T[B, A]$ is productive, so that it has a conclusion $\chi(B, A)$. Then the theory

$$T \cup \{\overline{\chi(B, A)}\} \tag{7.2.5}$$

is minimal inconsistent by Corollary 5.3.4, and so Theorem 7.1.8 gives (i) and the first clause of (ii). For the second clause, suppose $T[B, A]$ contains a negative sentence, and hence is not goclenian. By the Rule of Quality (Corollary 5.3.6(a)) $\overline{\chi(B, A)}$ is affirmative, so $\chi(B, A)$ is negative, and hence A is distributed in $\chi(B, A)$. By Theorem 7.2.5(a), $\chi(B, A)$ is the mate of $T[B, A]$, and so A is distributed in $T[B, A]$. This proves (b).

(b) ⇒ (a): Assume (b). Since goclenian premise-sequences are productive by Lemma 7.2.2, we can also assume that $T[B, A]$ is not goclenian. Let $\chi(B, A)$ be the mate of $T[B, A]$, and consider the theory

$$T \cup \{\overline{\chi(B, A)}\}. \tag{7.2.6}$$

We will show that (7.2.6) is an MIT. This will prove that (7.2.6) is inconsistent, and hence that $T \vdash \chi(B, A)$, which implies that T is productive.

Since $T[B, A]$ is graph-linear, (7.2.6) is graph-circular. By (i) and the fact that $\chi(B, A)$ is the mate of $T[B, A]$, every relation symbol occurring in (7.2.6) is distributed in at least one of its occurrences. So by Theorem 7.1.8, to show that (7.2.6) is an MIT, it suffices to show that (7.2.6) contains exactly one negative sentence.

Suppose for contradiction that (7.2.6) contains either (1) no negative sentences or (2) at least two negative sentences.

In case (1), $\overline{\chi(B, A)}$ is affirmative, so $\chi(B, A)$ is negative and A is distributed in $T[B, A]$. Then by Lemma 7.2.3, $T[B, A]$ is goclenian, contrary to assumption.

In case (2), since by (ii) T contains at most one negative sentence, $\overline{\chi(B, A)}$ must be negative, so that $\chi(B, A)$ is affirmative, and A is undistributed in $\chi(B, A)$ and hence also in $T[B, A]$ by choice of $\chi(B, A)$. But also by (ii), since T contains a negative sentence, A is distributed in $T[B, A]$, which is a contradiction. □

7.3 The Philoponus rules

In the previous section we gave necessary and sufficient conditions for productivity of an assertoric premise-sequence, and a rule for finding the conclusion of a productive assertoric premise-sequence. Ibn Sīnā stated his own versions of these conditions and rule at the beginning of his study of assertoric syllogisms, in all his major logical works from *Muḵtaṣar* onwards (except that the account of the rule in *Išārāt* is curtailed, like much else in *Išārāt*). Stating them at the start of the list of moods seems to have been his own innovation, which he justifies in *Qiyās* [55] i.2. See [45] for more on his likely motivation, and [47] for a translation of *Qiyās* i.2.

So this is one place where we can make a direct comparison between Ibn Sīnā's account and the logical facts. Ibn Sīnā based his conditions and rule on earlier work reported by the 6th century Alexandrian logician John Philoponus, and we will include Philoponus' versions in the comparison.

Both Philoponus and Ibn Sīnā confine their statements of the conditions and rule to the first three figures. The earliest Arabic account that we have of the fourth figure is by Rāzī's teacher al-Jīlī in the mid 12th century; see Exercise 7.6.

We begin with the conditions of productivity.

Theorem 7.3.1 *Let $T[C, A]$ be an assertoric premise-pair consisting of a premise ϕ_1 whose relation symbols are C and B, and a premise ϕ_2 whose relation symbols are B and A. (So ϕ_1 is the minor premise and ϕ_2 the major, C the minor extreme, B the middle term and A the major extreme.) Then $T[C, A]$ is productive if and only if it meets the following three conditions:*

(α) ϕ_1 and ϕ_2 are not both negative.

(β) ϕ_1 and ϕ_2 are not both existential.

(γ) *Further requirements according to the figure of $T[C, A]$:*

First figure: *ϕ_1 is affirmative and ϕ_2 is universal.*

Second figure: *At least one of ϕ_1 and ϕ_2 is negative, and ϕ_2 is universal.*

Third figure: *ϕ_1 is affirmative.*

Fourth figure: *If ϕ_1 is existential then ϕ_2 is negative; and if ϕ_2 is existential then both premises are affirmative.*

Proof. We show that these conditions are equivalent to conditions (i), (ii) in Theorem 7.2.6.

First assume $T[C, A]$ is productive, so that (i) and (ii) of Theorem 7.2.6 hold. We will show that (α), (β) and (γ) hold. (α) is immediate from (ii). By (i), at least one occurrence of B in T is distributed. If (β) fails then both premises have undistributed subject symbols, and by (α) at least one of them has undistributed predicate symbol, so the distributed occurrence of B must be the predicate symbol of the other premise, making the other premise negative; but then by (ii), the occurrence of A is distributed too, which is arithmetically impossible. So (β) holds.

In first figure, if ϕ_1 is negative then by the second part of (ii), ϕ_2 is also negative, contradicting the first part of (i). So ϕ_1 is affirmative. If ϕ_2 is existential then B is undistributed in ϕ_2 and hence distributed in ϕ_1 by (i); this makes ϕ_1 negative and hence is impossible. So ϕ_2 is universal.

In second figure, if both premises are affirmative then B is undistributed at both occurrences, contradicting (i). Hence at least one premise is negative, and so the occurrence of A in ϕ_2 is distributed by (ii), making ϕ_2 universal.

In third figure, if ϕ_1 is negative then the occurrence of A in ϕ_2 is distributed, making ϕ_2 negative too and hence contradicting (α).

In fourth figure, if ϕ_1 is existential then B is undistributed in ϕ_1, so B is distributed in ϕ_2 by (i), making ϕ_2 negative. If ϕ_2 is existential then A is undistributed in ϕ_2, and hence by (ii) both premises are affirmative.

Second, assume that the conditions (α), (β) and (γ) hold; we deduce (i) and (ii) of Theorem 7.2.6, so that by that theorem $T[C, A]$ is productive.

The first part of (ii) is immediate from (α). We prove (i) by examining the four figures. In first figure, ϕ_2 is universal and hence B is distributed in ϕ_2. In second figure at least one premise is negative and hence at least one occurrence of B is distributed. In third figure at least one premise is universal by (β), and hence at least one occurrence of B is distributed. In fourth figure either ϕ_1 is universal or ϕ_2 is negative, and hence at least one occurrence of B is distributed. This proves (i).

It remains to prove the second part of (ii). Suppose this fails, so that there is a negative premise but the occurrence of A in ϕ_2 is undistributed. In figures one and three it follows that ϕ_2 is affirmative; but in these two figures the conditions tell us that ϕ_1 is affirmative too, contradicting that there is a negative premise. In figures two and four the condition on A implies that ϕ_2 is existential; by the figure conditions, this is impossible in figure two, and in figure four it impllies that both premises are affirmative, contradicting that there is a negative premise. So the second part of (ii) must be true too. □

History 7.3.2 The conditions of productivity in Theorem 7.3.1 for the first three figures are exactly as stated by Philoponus [80] 70.1–21. (Philoponus counts Ibn Sīnā's fourth figure as a special case of the first figure and gives no separate conditions for it.) Ibn Sīnā states the same conditions as Philoponus, but in *Muḵtaṣar*, *Najāt* and *Qiyās* he adds a further general condition: 'It is not the case that ϕ_1 is negative and ϕ_2 is existential'. This condition follows from the others (see Exercise 7.4 below), so it is correct but redundant.

History 7.3.3 Philoponus says that the productivity conditions were assembled from remarks of Aristotle, and the purpose of assembling them was to help in the task of counting the number of valid moods—this was a typical preoccupation of Roman Empire logicians. The remarks of Aristotle that Philoponus has in mind are places where Aristotle summarises the results of sterility proofs. This enables Ibn Sīnā, having stated the conditions at the outset of his account of assertoric moods, to spend no more time on questions of sterility. Abū al-Barakāt—one of the very few logicians capable of applying independent judgment to Ibn Sīnā's logic before the generation of Rāzī and Suhrawardi—took issue with Ibn Sīnā at this point and insisted on giving Aristotle-style sterility proofs for the sterile assertoric moods. Cf. [10], for example p. 140 on sterility in Second Figure.

We turn to the rule for determining the conclusion of a productive assertoric premise-pair. In Theorem 7.2.5 we stated the rule using distribution. Ibn Sīnā didn't have this notion. But the following corollary translates the rule into terms that he would have understood at once.

Corollary 7.3.4 *Let* $<\phi_1, \phi_2>$ $[C, A]$ *be a productive premise-pair in first, second or third figure, and* $\chi(C, A)$ *the conclusion of this premise-pair. Then* $\chi(C, A)$ *is identified by the following facts:*

(a) $\chi(C, A)$ *is universal if and only if (in first and second figures)* ϕ_1 *is univer-*

sal, or (in third figure) ϕ_1 *is negative;*

(b) $\chi(C, A)$ *is negative if and only if (in first and third figures)* ϕ_2 *is negative, or (in second figure)* ϕ_2 *is universal.*

The *proof* is read off from Lemma 7.1.2 and the definition of the figures. □

Philoponus ([80] 123.12–20) tells us that Aristotle and his pupils thought of the question as: From which premise does the conclusion inherit each of its properties (quantity, quality, modality)? So to say, which parent does the baby get its ears from? Answers were given in terms of which premise the conclusion 'follows' (in Arabic *yatba*^{c}*u*). Hence I speak of 'rules of following', though this is my phrase, not one found in Ibn Sīnā. The book [45] discusses the history more fully, including Ibn Sīnā's notion of 'dominance' (^{c}*ibra*). Here I confine myself to the version that Ibn Sīnā gives in his discussions of assertoric logic (again excluding *Išārāt* which has only a partial version).

Definition 7.3.5 The *Peiorem Rule* states (falsely) that if $T[C, A]$ is a productive assertoric premise-pair, then its optimal conclusion $\chi[C, A]$ has the following properties:

(a) χ is negative if and only if at least one premise in T is negative.

(b) χ is existential if and only if at least one premise in T is existential.

It can also be stated as: In quality and quantity, χ follows the worse of the two premises, where negative is worse than affirmative and existential is worse than universal.

The Peiorem rule is false, but remarkably it's false for only two of the two-premise assertoric moods, namely *Darapti* and *Felapton* which both have universal premises and existential conclusion. Ibn Sīnā was well aware of this exception; he calls attention to it in *Najāt* [57] 64.1–3 and in a smaller work ^{c}*Uyūn al-ḥikma* [63] 50.2f. This is tiresome for the historian, because it implies that in reading Ibn Sīnā we have to factor in the possibility that he is being careless about cases that he considers marginal. There are two different explanations of this particular piece of carelessness. He may reckon that a correction of the Peiorem rule wouldn't repay the cost of the complication. Or he may hope that his readers will overlook the difference between 'exactly as bad as the worst' and 'no better than the worst'. Either way, the Peiorem rule stands as a classic example of spotting the wrong pattern and then sticking with it in spite of the evidence.

History 7.3.6 'Worse' is *'akass* in Arabic, and *peior* in Latin. Ibn Sīnā states the rule at *Muktaṣar* 49b8, *Najāt* 53.13–15 and *Qiyās* 108.9. (The Cairo edition of *Qiyās*, [55], reports that the manuscripts have 'better' (*'aḥsan*) rather than 'worse' at this point. This can only be a miscopying, though it's surprising that a mistake of this kind, which must have been made very early in the transmission of the book, survived so many copyings without being picked up.) In *Išārāt* the Peiorem rule is apparently downgraded, being stated at 145.3 only for first figure syllogisms—for which it is correct. In the *Muktaṣar*, *Najāt* and *Išārāt* passages Ibn Sīnā is explicit that he is asserting the rule for quantity and quality.

7.4 Exercises

7.1. Let ϕ be an assertoric sentence and ϕ^- the unaugmented assertoric sentence that results from removing the augment (if any) in ϕ. Suppose the relation symbol A occurs in ϕ. Show that the following are equivalent:

(a) A is distributed (resp. undistributed) in ϕ.

(b) A has only negative (resp. positive) occurrences in ϕ^-.

COMPLETE THIS.

7.2. Exercise on Buridan's definition of distribution.

7.3. Show that it is not true that if $T[B, A]$ is an assertoric premise-sequence and each segment of length 2 in $T[B, A]$ is productive, then $T[B, A]$ is productive.

Solution: Take for example $< (e)(D, C), (a)(B, C), (o)(B, A) >$ $[D, A]$. The first segment of length 2 is *Camestres* and the second is *Bocardo*, but the premise-sequence is sterile since it contains two negative sentences.

7.4. Verify that the conditions of productivity in Theorem 7.3.1 entail the statement 'It is not the case that ϕ_1 is negative and ϕ_2 is existential'. (Cf. History 7.3.2. In fact the statement follows from the special figure conditions (γ) without needing (α) and (β).)

Solution. We quote the conditions in (γ) of Theorem 7.3.1. In First and Second Figures the statement follows from the condition that ϕ_2 is universal. In Third Figure it follows from the condition that ϕ_1 is affirmative. In Fourth Figure it follows from the condition that if ϕ_2 is existential then both premises are affirmative.

7.5. William of Sherwood [94] p. 76 quotes the Philoponus conditions of productivity, but leaving out the condition that in Second Figure the major premise must be universal.

(a) Show that this omission is an error. More precisely, give two examples of sterile assertoric premise-pairs in Second Figure, $<\phi_1(C,B), \phi_2(A,B))> [C,A]$, which satisfy ($\alpha$) and ($\beta$) of Theorem 7.3.1 and have exactly one negative premise.

(b) Show that if $T[C,A]$ is a sterile premise-pair as in (a), then $T[A,C]$ is a productive premise-pair in Second Figure. [This could have misled Sherwood.]

Solution. (a) $<(a)(C,B),(o)(A,B)> [C,A]$ and $<(e)(C,B),(i)(A,B)> [C,A]$ are examples.

(b) The premise-pair $T[A,C]$ is again in Second Figure, and it satisfies (α) and (β) since they are symmetric between the two premises. Similarly it has exactly one negative premise. Since $T[C,A]$ was sterile, ϕ_2 was existential and so ϕ_1 was universal. Since $\phi_1(A,B)$ becomes the major premise in $T[A,C]$, all the conditions for $T[A,C]$ to be productive are met.

7.6. In the 12th century Al-Jīlī gave the following productivity conditions for the assertoric fourth figure ([27] p. 221; I thank Asadollah Fallahi for making this paper available):

(7.4.1)

(i) Neither premise is of the form (o).

(ii) If the minor premise is of the form (e) then the major premise is of the form (a).

(iii) If the minor premise is of the form (i) then the major premise is not of the form (a).

Infer from Theorem 7.3.1 that these conditions are necessary but not jointly sufficient, and that they can be made sufficient by adding that there is at most one existential premise.

Solution. We restrict attention to the Fourth Figure. We show that the conditions (i)–(iii) are necessary. (i) If ϕ_1 has the form (o) then it is existential, so by (γ) ϕ_2 is negative, hence by (α) ϕ_1 is affirmative, contradiction. If ϕ_2 has the form (o) then it is existential, so by (γ) both premises are affirmative, contradiction. (ii) If ϕ_1 has the form (e) then it is negative, so ϕ_2 is universal by (γ) and affirmative by (α). (iii) If ϕ_1 has the form (i) then ϕ_2 is negative by (γ), and hence not of the form (a).

The conditions (i)–(iii) are not sufficient for productivity. They are all met if both premises are of the form (i), but this violates (β).

We show that if we add (β) to (i)–(iii), all the conditions of Theorem 7.3.1 are derivable. If both premises are negative, then by (i) ϕ_1 is not of the form (o), and by (ii) ϕ_1 is not of the form (e); this proves (α). (β) is assumed. For the condition in (γ), first suppose that ϕ_1 is existential, so that by (i), ϕ_1 has the form (i) and hence by (iii), ϕ_2 is not of the form (a). By (β), ϕ_2 is universal, and so ϕ_2 must be of the form (e), which is negative. Second, suppose ϕ_2 is existential. Then by (ii), ϕ_1 is not of the form (e), and by (i), ϕ_1 is not of the form (o), so ϕ_1 is affirmative; also by (i), ϕ_2 is affirmative.

7.7. Show:

(a) If the Peiorem rule for assertoric logic is taken to refer to productive premise-sequences of arbitrarily length, then it is equivalent to the claim:

> In an assertoric MIT, an UNWEAKENABLE sentence ϕ is (1) affirmative if and only if at least one other sentence in the MIT is negative, and (2) universal if and only if at least one other sentence in the MIT is existential.

(b) In the form spelled out in (a), the rule is true of an assertoric MIT T if and only if the nominator type of T is not (a).

Chapter 8

Ibn Sīnā's assertoric proof theory

Ibn Sīnā's own proof theory for assertoric sentences serves a different purpose from ours in Chapter 6 above. Our aim was to characterise the minimal inconsistent sets, using a method that we can extend to two-dimensional sentences. Ibn Sīnā's aim is to describe a process that leads us step by step from a commitment (Arabic *taslīm*) to the premises in a premise-sequence to a commitment to the conclusion.

Ibn Sīnā splits his proof theory into two parts, according to how fine-grained the steps are. One part takes the two-premise moods as steps, and shows how they can be used to derive conclusions from arbitrary productive premise-sequences. Ibn Sīnā expounds this in *Qiyās* [55] ix.3–6; he assumes throughout that the sentences he is dealing with are assertoric, though he makes a remark that the process can be extended to all recombinant syllogisms. The other part takes as steps the perfect two-premise moods together with some conversions and ectheses; in effect this part shows how to make derivations by reasoning from first principles.

8.1 Derivations

We are studying how one deduces a single assertoric sentence from a premise-sequence. We begin with the account in *Qiyās* ix.3. Here Ibn Sīnā describes adjustments that are made to the premise-sequence until only a single sentence is left; this sentence is the conclusion.

Definition 8.1.1 (a) By a *rule-book* we mean a set of rules of the form

(8.1.1) $$T[B, A] \;\Rightarrow\; U[B, A]$$

where $T[B, A]$ and $U[B, A]$ are premise-sequences with the same minor extreme and the same major extreme.

(b) Let $\mathbb{R}$ be a rule-book. By an $\mathbb{R}$-*derivation* we mean a finite sequence of premise-sequences such that (i) the final premise-sequence is forwards and has length 1, and (ii) if $T[B, A]$ and $U[D, C]$ are two consecutive premise-sequences in the sequence, then the rule $T[B, A] \Rightarrow U[D, C]$ is in $\mathbb{R}$. (It follows that $B = D$ and $A = C$.) The *posit* (Arabic *waḍ*^{c}) of the $\mathbb{R}$-derivation is its first premise-sequence. The *conclusion* (Arabic *natīja*) of the $\mathbb{R}$-derivation is the sentence in its final premise-sequence.

(c) We write

(8.1.2) $$T[B, A] \vdash_{\mathbb{R}} \phi$$

to mean that there is an $\mathbb{R}$-derivation whose posit and conclusion are respectively $T[B, A]$ and ϕ.

(d) A *derivation* is an $\mathbb{R}$-derivation for some rule-book $\mathbb{R}$ (which will usually be determined by the context). The derivation is said to be *from* (Arabic *min*) its posit and *of* (Arabic ^{c}*alā*) its conclusion.

(e) If a derivation is the sequence $(T_1, \ldots, T_n)$ where each T_i is a premise-sequence, then we call T_i the i-th *line* of the derivation. The *length* of the derivation is n.

I give the definition above without claiming—at least not in this book—that Ibn Sīnā himself wrote derivations as sequences of premise-sequences, or that he wrote each line of a derivation as a literal 'line'. In [47] we assess the limited evidence about how he did think and write derivations.

Definition 8.1.2 Let $\mathbb{P}$ be a class of premise-sequences, and $\mathbb{R}$ a rule-book.

(a) We say that $\mathbb{R}$ is *sound for* $\mathbb{P}$ if for all premise-sequences $T[B, A]$ in $\mathbb{P}$ and all sentences ϕ, if

(8.1.3) $$T[B, A] \vdash_{\mathbb{R}} \phi$$

then

(8.1.4) $$T[B,A] \;\rhd\; \phi$$

(cf. Definition 3.3.2(b) for $\rhd$).

(b) We say that $\mathbb{R}$ is *complete for* $\mathbb{P}$ if for all premise-sequences $T[B,A]$ in $\mathbb{P}$ and all sentences ϕ, if (8.1.4) then (8.1.3).

The next few definitions and lemmas collect up what we need for proving soundness and completeness.

Lemma 8.1.3 $<\phi(B,A)>[B,A] \;\rhd\; \phi(B,A)$.

□

Definition 8.1.4 Suppose $T[B,A]$ and $U[B,A]$ are productive premise-sequences with the same minor extreme B and the same major extreme A. We write

(8.1.5) $$T[B,A] \,\nabla\, U[B,A]$$

to mean that for all sentences $\phi(B,A)$,

(8.1.6) $$\begin{array}{l} T[B,A] \rhd \phi(B,A) \text{ if and only if} \\ U[B,A] \rhd \phi(B,A). \end{array}$$

So ∇ is an equivalence relation on the class of productive premise-sequences.

Lemma 8.1.5 *Let $T[B,A]$ be an assertoric productive premise-sequence. Then*

(8.1.7) $$T[B,A] \,\nabla\, <\phi(B,A)>[B,A]$$

if and only if

(8.1.8) $$T[B,A] \;\rhd\; \phi(B,A).$$

Proof. The direction $\Rightarrow$ follows at once from Lemma 8.1.3. In the other direction, (8.1.8) and Lemma 8.1.3 imply that both $T[B,A]$ and $<\phi(B,A)>[B,A]$ have the consequence $\phi(B,A)$. But by Corollary 5.3.7(b), in assertoric logic conclusions are unique. □

The next definition helps us to define rules.

Definition 8.1.6 (a) Let $V[D, C]$ be a premise-sequence. We recall from Definition 3.3.1 that this means V is a graph-linear theory with a direction defined by $[D, C]$. A *segment* of $V[D, C]$ consists of a graph-linear theory T got by removing zero or more sentences at each end of V, and choosing the terms $[B, A]$ to give T the direction inherited from V. So the resulting segment of $V[D, C]$ is a premise-sequence $T[B, A]$. We call $T[B, A]$ an *initial segment* of $V[D, C]$ if T contains the initial sentence of V, a *final segment* of $V[D, C]$ if T contains the final sentence of V, and an *internal segment* of $V[D, C]$ if it is neither initial nor final.

(b) We write $V[D, C]$ as

(8.1.9) $$<\ldots\ T[B, A]\ \ldots>[D, C]$$

to indicate that $T[B, A]$ is a segment of $V[D, C]$. Then we write

(8.1.10) $$<\ldots\ U[B, A]\ \ldots>[D, C]$$

to mean the premise-sequence that results from $V[D, C]$ if the segment $T[B, A]$ is replaced by $U[B, A]$. We call (8.1.10) *the result of replacing $T[B, A]$ by $U[B, A]$ in $V[D, C]$*.

(c) If the first '...' is missing in (8.1.9) and (8.1.10), this indicates that $T[B, A]$ is an initial segment of $V[D, C]$; if the second '...' is missing then $T[B, A]$ is a final segment of $V[D, C]$.

Lemma 8.1.7 *Let $V[D, C]$ be a productive assertoric premise-sequence and $T[B, A]$ a segment of $V[D, C]$. Then $T[B, A]$ is also productive.*

Proof. We use the criterion of productivity stated in Theorem 7.2.6 above. Condition (i) automatically passes from $V[D, C]$ to the segment $T[B, A]$, and so does the first part of (ii).

For the second part of (ii), suppose for contradiction that $T[B, A]$ contains a negative sentence ϕ, but A is undistributed in $T[B, A]$. Since ϕ is also in V, (ii) in $V[D, C]$ implies that C is distributed in $V[D, C]$. So ϕ is not the final sentence of $V[D, C]$. Let $U[E, C]$ be the final segment of $V[D, C]$ that starts at the sentence immediately after ϕ. Then the conditions (i)–(iii) of Lemma 7.2.3 are satisfied by $U[E, C]$, so by that Lemma, $U[E, C]$ is goclenian. The term A is one of the terms of $U[E, C]$, but not its final (major) extreme. Write ψ for the second sentence containing A; then ψ has the form (a) and A is its predicate term. So A is undistributed in ψ, and hence by

(i) of Theorem 7.2.6, A is distributed in the sentence immediately before ψ. But this means that A is distributed in $T[B, A]$, contradiction. □

Lemma 8.1.8 *Let $T[B, A]$ be a productive premise-sequence. Then $T[B, A]$ contains a negative sentence if and only if $T[B, A]$ is not goclenian and A is distributed in T.*

Proof. $\Rightarrow$: If $T[B, A]$ contains a negative sentence then A is distributed in T by Theorem 7.2.6, and certainly $T[B, A]$ is not goclenian. In the other direction, suppose $T[B, A]$ is not goclenian and A is distributed in T. Since $T[B, A]$ is productive, each internal term of $T[B, A]$ is distributed in at least one occurrence. It follows by Lemma 7.2.3 that $T[B, A]$ contains a negative sentence. □

Lemma 8.1.9 *Let $T[B, A]$ and $U[B, A]$ be productive assertoric premise-sequences. Then $T[B, A]\nabla U[B, A]$ if and only if one of the following holds:*

(a) Neither $T[B, A]$ nor $U[B, A]$ is goclenian, and they have the same distributivity at A and at B.

(b) Both are goclenian.

(c) One (say $T[B, A]$) is goclenian and the other is not; and in $U[B, A]$, both B and A are undistributed.

Proof. This is read off from Theorem 7.2.5. □

Lemma 8.1.10 *Let $T[B, A]$ and $U[B, A]$ be productive premise-sequences such that $T[B, A]\nabla U[B, A]$ and $U[B, A]$ is not goclenian. Let $V[D, C]$ be a productive premise-sequence which has $T[B, A]$ as a segment, and let $W[D, C]$ be the premise-sequence that results from $V[D, C]$ by replacing $T[B, A]$ by $U[B, A]$. Suppose also that if at least one of $T[B, A]$ and $U[B, A]$ is goclenian, then $V[D, C]$ is goclenian and $T[B, A]$ is a final segment of $V[D, C]$. Then $W[D, C]$ is productive and $V[D, C]\nabla W[D, C]$.*

Proof. We take the three cases of the previous lemma.

Case (a): neither $T[B, A]$ nor $U[B, A]$ is goclenian, and both have the same distributivity at A and the same distributivity at B. Then neither $V[D, C]$ nor $W[D, C]$ is goclenian, and both have the same distributivity at D, and the same distributivity at C. By Lemma 8.1.8, $V[D, C]$ and $W[D, C]$ both or neither contain a negative sentence; so $W[D, C]$ is productive by Theorem 7.2.6, and $V[D, C]\nabla W[D, C]$ by Lemma 8.1.8 again.

Case (b): both $T[B, A]$ and $U[B, A]$ are goclenian.. Then as in case (a), $V[D, C]$ has the same distributivity as $W[D, C]$, both at D and at C. Also $V[D, C]$ is goclenian if and only if $W[D, C]$ is goclenian. If neither is goclenian, then again $W[D, C]$ is productive by Theorem 7.2.6, and $V[D, C]\nabla W[D, C]$ by Lemma 8.1.8. If both are goclenian then both are productive and have the conclusion $(i)(D, C)$.

Case (c): suppose first that $T[B, A]$ is goclenian and $U[B, A]$ is not, and in $U[B, A]$ both extremes are undistributed. By the extra condition in this case, $V[D, C]$ is goclenian and both D and C are undistributed in $W[D, C]$. Then by the LEMMA, no negative sentence appears in $W[D, C]$. The distributivity condition is met at internal terms, since the change of distributivity is at the end. So $W[D, C]$ is productive, and then again by the Lemma, $V[D, C]\nabla W[D, C]$.

What about the remaining case, that $U[B, A]$ is goclenian and $T[B, A]$ is not. How do we show $W[D, C]$ is productive? It has the requirement on internal terms. Suppose $W[D, C[$ contains a negative, which neceesarily is not in $U[B, A]$. Then that negative is in $V[D, C]$ too, so by assumption that V is productive, C is distributed in V. This contradicts that $T[B, A]$ is not goclenian, since then it would have C undistributed. MAYBE NOT NEED THIS CASE. □

8.2 Compound from simple

Definition 8.2.1 The rule-book $\mathbb{R}1$ is defined as follows.

(a) Let $T[B, A]$ be any non-goclenian productive assertoric premise-pair, and $\phi(B, A)$ its conclusion. Then $\mathbb{R}1$ contains all the rules of the form

$$<\ldots, T[B, A], \ldots> [D, C] \;\Rightarrow$$
$$<\ldots, <\phi(B, A)> [B, A], \ldots> [D, C].$$

(So the rule takes a line of the derivation to the line got by replacing the segment $T[B, A]$ by the segment $<\phi(B, A)> [B, A]$.)

(b) Let $T[B, C]$ be the goclenian productive assertoric premise-pair (i.e. the premises of the mood *Bamalip*), and $\phi(B, C)$ its conclusion. Then $\mathbb{R}1$ contains all the rules of the form

$$<\ldots, T[B, C]> [D, C] \;\Rightarrow$$
$$<\ldots, <\phi(B, C)> [B, C]> [D, C].$$

(So *Bamalip* is applied only when its premises are a final segment of the line of the derivation.)

There are no other rules in $\mathbb{R}1$.

Lemma 8.2.2 *Let $V[D,C] \Rightarrow W[D,C]$ be a rule in $\mathbb{R}1$. Then $W[D,C]$ is productive and $V[D,C]\nabla W[D,C]$.*

Proof. This is covered by Lemma 8.1.10. □

Theorem 8.2.3 *The rule-book $\mathbb{R}1$ is sound and complete for the class of all assertoric premise-sequences.*

Proof. First we prove soundness, by induction on the length of the derivation. Let $\mathbb{D}$ be an $\mathbb{R}1$-derivation of length n.

Suppose first that $n = 1$. Then the one line of $\mathbb{D}$ must consist of a premise-sequence of length 1, say $<\phi(B,A)>$ $[B,A]$. In this case soundness states that

(8.2.1) $$<\phi(B,A)> [B,A] \;\rhd\; \phi(B,A)$$

This holds by Lemma 8.1.3.

Next suppose that $n > 1$. Let $V[D,C]$ and $W[D,C]$ be the first and second lines of $\mathbb{D}$, and let $\mathbb{D}'$ be the derivation got from $\mathbb{D}$ by removing the first line. By Lemma 8.2.2, $V[D,C]\;\nabla\;W[D,C]$.

Now let $\chi(D,C)$ be the conclusion of both $\mathbb{D}$ and $\mathbb{D}'$. By induction hypothesis $W[D,C] \;\rhd\; \chi(D,C)$. By the definition of ∇ it follows that $V[D,C] \rhd \chi(D,C)$ as required. This proves soundness.

Next we prove completeness. Let $V[D,C]$ be a productive assertoric premise-sequence of length n, with conclusion $\chi(D,C)$. We prove that

(8.2.2) $$V[D,C] \vdash_{\mathbb{R}1} \chi(D,C)$$

by induction on n.

Assume first that $n = 1$. Then by Lemma 8.1.3, the $\mathbb{R}1$-derivation consisting of the single line

(8.2.3) $$<\chi(D,C)> [D,C]$$

has posit $V[D,C]$ and conclusion $\chi(D,C)$ as required.

Next suppose that $n > 1$.

> **Claim**. There is a premise-sequence $W[D, C]$ of length $n - 1$ such that $V[D, C] \Rightarrow W[D, C]$ is a rule in $\mathbb{R}1$.

Proof of claim. There is a segment $T[B, A]$ of $V[D, C]$ of length 2 which either is not goclenian or is a final segment. (If all segments of length 2 are goclenian, take the last one.) Choose such a $T[B, A]$. Since $V[D, C]$ is productive, by Lemma 8.1.7 $T[B, A]$ is also productive, say with conclusion $\phi(B, A)$. So by Definition 8.2.1, $\mathbb{R}1$ contains a rule

$$\begin{array}{l} <\ldots, T[B, A], \ldots> [D, C] \;\Rightarrow \\ <\ldots, <\phi(B, A)> [B, A], \ldots> [D, C]. \end{array}$$

where the first premise-sequence in the rule is $V[D, C]$; take $W[D, C]$ to be the second premise-sequence of the rule. Then $W[D, C]$ has length $n - 1$. □ Claim.

By induction hypothesis there is an $\mathbb{R}1$-derivation $\mathbb{E}$ with posit $W[D, C]$ and conclusion $\chi(D, C)$. Form the derivation $\mathbb{D}$ from $\mathbb{E}$ by adding $V[D, C]$ in front of the first line of $W[D, C]$. Then $\mathbb{D}$ is an $\mathbb{R}1$-derivation by the Claim, and $\mathbb{D}$ has posit $V[D, C]$ and conclusion $\chi(D, C)$ as required. □

Bearing in mind that Ibn Sīnā rejected fourth-figure syllogisms as unnatural, we should check what happens if they are excluded from the rule-book.

Definition 8.2.4 We define the rule-book $\mathbb{R}2$ exactly as $\mathbb{R}1$, except that in clause (b) of Definition 8.2.1 we add the requirement that $T(B, A)$ is not retrograde.

Theorem 8.2.5 *The rule-book $\mathbb{R}2$ is sound and complete for the class of all non-retrograde assertoric premise-sequences.*

Proof. Soundness is a special case of the soundness of $\mathbb{R}1$ proved in Theorem 8.2.3 above. For completeness the proof is the same as for Theorem 8.2.3, except that we have to check that if $V[D, C]$ is not retrograde then it has a segment of length 2 that is not retrograde. This is trivially true. □

Since Ibn Sīnā rejected fourth-figure syllogisms as unnatural, he might have considered some non-retrograde premise-sequences of lengh > 2 unnatural too. In [47] we consider the evidence on this. All we can say from

Theorem 8.2.5 is that if he did accept all non-retrograde assertoric premise-sequences as natural, then his proof methods are adequate for them. Both Theorem 8.2.3 and Theorem 8.2.5 in the general case allow many different derivations of the same conclusion from the same posit, because there will be many different ways of choosing segments. There is strong evidence that Ibn Sīnā required, or at least preferred, derivations meeting some requirements on the order (Arabic *tartīb*) in which the premises are taken. This is discussed further in [47].

Let me call attention to a feature of $\mathbb{R}1$ and $\mathbb{R}2$ that might easily go unnoticed as trivial.

Definition 8.2.6 Let $\mathbb{R}$ be a rule-book. We say that $\mathbb{R}$ is *cartesian* if for every rule

$$V[D, C] \;\Rightarrow\; W[D, C] \tag{8.2.4}$$

in $\mathbb{R}$, $V \vdash \psi$ for each sentence ψ in W.

> **History 8.2.7** To summarise part of Descartes' Rule Three for the Direction of the Mind [20]: By intuition we see clearly and with certainty that a given conclusion is a necessary consequence of given premises. In deduction we recollect with certainty that we have passed through a continuous and uninterrupted sequence of steps of reasoning, each of which gave us certainty through intuition.

Lemma 8.2.8 *Every rule in* $\mathbb{R}1$ *(and hence also every rule in* $\mathbb{R}2$*) is cartesian.* □

INCORPORATE ABOVE:

Lemma 8.2.9 *Let* $T[B, A] \Rightarrow U[B, A]$ *be a rule in* $\mathbb{R}1$*. If* $T[B, A]$ *is sterile then so is* $U[B, A]$*.*

Proof. We use Theorem 7.2.6. Assume that $T[B, A]$ is sterile. Then one of the cases (1)–(3) below holds for $T[B, A]$. We show that it holds also for $U[B, A]$.

(1) Some relation symbol C that occurs in two sentences of $T[B, A]$ is undistributed in both. There are several subcases here.

(1.1) C is not a term of the simple syllogism applied in the rule. Then C passes down into $U[B, A]$ with the same distributivities that it had in $T[B, A]$.

(1.2) C is the middle term of the simple syllogism applied in the rule. This is impossible since the simple syllogism has a productive premise-pair.

(1.3) C is an extreme of the simple syllogism applied in the rule. Since C occurs in two sentences of $T[B, A]$, it is not A, and hence by Definition 8.2.1, if the simple syllogism is *Bamalip* then C is the minor term of this simple syllogism. But the major term of *Bamalip* is the only term in a productive simple syllogism which has a different distributivity in the conclusion from what it had in the premises (by Theorem 7.2.5). So C carries the same distributivities in $U[B, A]$ as it had in $T[B, A]$.

(2) $T[B, A]$ contains more than one negative sentence. Inspection shows that all rules in $\mathbb{R}1$ preserve the number of negative sentences.

(3) $T[B, A]$ contains exactly one negative sentence and A is undistributed in $T[B, A]$. In this case $U[B, A]$ also contains exactly one negative sentence, for the same reason as in (2). We have to show that A is undistributed in $U[B, A]$. By Theorem 7.2.5 again, every rule in $\mathbb{R}1$ preserves the distributivities of its terms, except for the case of the major extreme in *Bamalip*. But the rule using *Bamalip* can't apply here, because the major extreme of *Bamalip* is A by Definition 8.2.1, and in the premises of *Bamalip* this extreme is distributed. □

In [47] we examine what happens if one applies the procedure described in the proof of Theorem 8.2.3 to a sterile premise-sequence. We show that eventually one reaches a premise-sequence of length > 1 in which no segment of length 2 is the premise-sequence of a productive simple mood apart from *Bamalip*, and the final segment of length 2 is not the premise-sequence of *Bamalip*. At this point the Claim in the proof of Theorem 8.2.3 fails. It follows that the procedure of Theorem 8.2.3 is not just a sound and complete proof procedure; it is also a decision procedure for productivity.

In his proof search algorithm of *Qiyās* ix.6, Ibn Sīnā calls attention to places where the algorithm hits a failure of productivity. But these are not necessarily places where the algorithm halts, because there may be possibilities of backtracking. For this reason we are not in a position to say outright that Ibn Sīnā in *Qiyās* ix.6 intended to give a decision procedure for productivity, though there are some positive pointers in this direction.

8.3 Deriving from first principles

When he turns to the issue of proving from first principles of logic, Ibn Sīnā adopts a rule-book $\mathbb{R}3$ that is very different from $\mathbb{R}1$. At first sight his construction of $\mathbb{R}3$ simply borrowed from Aristotle—and there is no denying that most of it does come from Aristotle. But Ibn Sīnā has the further aim of finding a rule-book that he can generalise to two-dimensional logic, if possible. So it will pay to look at the details. We begin with a subset $\mathbb{R}3^-$ that is completely unproblematic.

Definition 8.3.1 The rule-book $\mathbb{R}3^-$ is defined as follows, using the format of Definition 8.1.6 with the following values for $T[B,A]$ and $U[B,A]$:

	$T[B,A]$	$U[B,A]$
1.	$<(e)(A,B)>[B,A]$	$<(e)(B,A)>[B,A]$
2.	$<(i)(A,B)>[B,A]$	$<(i)(B,A)>[B,A]$
3.	$<(a)(A,B)>[B,A]$	$<(i)(B,A)>[B,A]$

together with all cases where $T[B,A]$ is a productive first-figure premise-sequence of length 2, and $U[B,A]$ is $<\phi(B,A)>[B,A]$ where $\phi B, A$ is the conclusion of $T[B,A]$.

Lemma 8.3.2 *For each rule* $V[D,C] \Rightarrow W[D,C]$ *in* $\mathbb{R}3^-$ *we have* $V[D,C] \nabla W[D,C]$.

Proof. The first-figure rules are covered already by the proof of Lemma 8.1.10. For the new rules, by Lemma 8.1.10 it suffices to show that $T[B,A] \nabla U[B,A]$ for each rule defined in Definition 8.3.1. In cases 1 and 2 this is trivial since $(e)(A,B)$ is logically equivalent to $(e)(B,A)$ and likewise with (i). For 3 we note that the only assertoric sentence $\phi(B,A)$ that follows from $(a)(A,B)$ is $(i)(B,A)$. □

It follows from Lemma 8.3.2 that $\mathbb{R}3^-$ is sound for the class of assertoric premise-sequences. To prove its completeness it would suffice to show that it is complete for the class of assertoric premise-sequences of length 2, by Theorem 8.2.3. In fact it is not, but we can show the following:

Lemma 8.3.3 *The rule-book* $\mathbb{R}3^-$ *is complete for the class of premise-sequences in the following moods:*

Barbara, Celarent, Darii, Ferio, Cesare, Festino, Darapti, Felapton, Datisi, Ferison, Fesapo, Fresison.

Proof. *Barbara*, *Celarent*, *Darii* and *Ferio* are given directly.

For the remaining eight moods, application of 1, 2 or 3 to either one or both of the premises changes backwards sentences to forwards, and hence provides a segment in first figure. The fourth figure moods *Fesapo* and *Fresison* need two conversions. We illustrate with *Fesapo*:

1.	$<(a)(B,C),(e)(A,B)>[C,A]$	posit
2.	$<(i)(C,B),(e)(A,B)>[C,A]$	by a-conversion
3.	$<(i)(C,B),(e)(B,A)>[C,A]$	by e-conversion
4.	$<(o)(C,A)>[C,A]$	by *Ferio*

□

There remain *Camestres*, *Baroco*, *Disamis*, *Bocardo*, *Bamalip*, *Dimatis* and *Calemes*. For all of these except *Baroco* and *Bocardo*, Ibn Sīnā adds two further kinds of rule to the rule-book, giving $\mathbb{R}3^{=}$ as follows.

Definition 8.3.4 The rule-book $\mathbb{R}3^{=}$ consists of the rules in $\mathbb{R}3^{-}$ together with the following cases of Definition 8.2.1(a):

5. $T[C,A]$ is $<(e)(B,C),(a)(A,B)>[C,A]$ and $U[C,A]$ is $<(e)(A,C)>[C,A]$.
6. $T[C,A]$ is $<(a)(B,C),(i)(A,B)>[C,A]$ and $U[C,A]$ is $<(i)(A,C)>[C,A]$.

Lemma 8.3.5 *The rule-book* $\mathrm{R}3^{=}$ *is complete for the class of premise-sequences of the moods in Lemma 8.3.3 together with* Camestres, Disamis, Bamalip, Dimatis *and* Calemes.

Proof. In all the new cases we need to use a conversion as last step. Some of these cases also need one or more conversions before the rules introduced in Definition 8.3.4 are invoked. We illustrate with the most complicated case, which is *Bamalip*:

1.	$<(a)(B,C),(a)(A,B)>[C,A]$	posit
2.	$<(a)(B,C),(i)(B,A)>[C,A]$	by a-conversion
3.	$<(a)(B,C),(i)(A,B)>[C,A]$	by i-conversion
4.	$<(i)(A,C)>[C,A]$	by 6 in Def 8.3.4
5.	$<(i)(C,A)>[C,A]$	by i-conversion

□

History 8.3.6 Both the rules in Definition 8.3.4 can be described as taking Fourth Figure moods (specifically *Calemes* and *Dimatis*) and then converting the conclusion. Ibn Sīnā gives a different description: swap the two premises and apply a First Figure mood. Obviously this description is better if he wants to avoid fourth-figure moods. But still I think he owes us an explanation. If he is prepared to justfy *Camestres* and *Disamis* on the strength of taking the rules in $\mathbb{R}3^{=}$ as basic, then why is he not prepared to accept *Bamalip* and *Calemes* on exactly the same basis? This is not a formal problem; it's a problem about the coherence of his reasons for finding the Fourth Figure 'unnatural'.

There remain *Baroco* and *Bocardo*. Both of these require us to handle a sentence of the form $(o)(B, A)$. Ibn Sīnā's approach is to introduce a new term $(B \setminus A)$ by definition. This allows him to split the sentence into two parts, so that he can apply other rules to one of these parts and an adjacent sentence. This is the one case where his assertoric rule-books include a rule that increases the length of a premise-sequence.

Definition 8.3.7 (a) Given distinct terms B and A, the term $(B \setminus A)$ is defined by

(8.3.1) $$\forall x\ ((B \setminus A)x \leftrightarrow (Bx \wedge \neg Ax)).$$

(b) The rule-book $\mathbb{R}3$ consists of the rules in $\mathbb{R}3^{=}$ together with one new kind of rule: in the format of Definition 8.1.6,

7. $T[B, A]$ is $<(o)(B, A)>$ $[B, A]$
   and $U[B, A]$ is $<(a)(B, (B \setminus A)), (e)((B \setminus A), A)>$ $[B, A]$.

(c) We call a rule of this kind an *ecthesis* (without claiming that this is what Aristotle meant by the term).

History 8.3.8 Ecthesis in Arabic is *iftirāḍ*. Ibn Sīnā also calls this device of introducing terms by definition *ta*^{c}*yīn*, literally 'making definite'. Within Ibn Sīnā's logic it seems to play a similar role to Definitions in Frege's *Begriffsschrift*: a concept is introduced by definition, and the definition is then allowed to be used as a premise in further deductions. Ibn Sīnā's formal language is not suited for giving explicit definitions as Frege does; he merely writes down some assertoric sentences that would follow from an explicit definition. The extension of this device to two-dimensional logic will play a crucial role in that logic.

Lemma 8.3.9 *The rule-book* $\mathbb{R}3$ *is complete for all two-premise assertoric moods.*

Lemma 8.3.10 *For every rule* $V[D, C] \Rightarrow W[D, C]$ *in* $\mathbb{R}3$*,* $V[D, C] \nabla W[D, C]$*.*

Proof. The only rules needing further treatment are those introduced in Definition 8.3.7(b). It suffices to check the distributions. The term $(B \setminus A)$ is undistributed in $(a)(B,(B \setminus A))$ and distributed in $(e)((B \setminus A),A)$. The term B is undistributed in both $(o)(B,A)$ and $(a)(B,(B \setminus A))$. The term A is distributed in both $(o)(B,A)$ and $(e)((B \setminus A),A)$. Now use Lemmas 8.1.9 and 8.1.10. □

Proof. It remains only to derive *Baroco* and *Bocardo*. We do these as follows, beginning with *Baroco*:

1.	$<(o)(C,B),(a)(A,B)>[C,A]$	posit
2.	$<(a)(C,(C\backslash B)),(e)((C\backslash B),B),(a)(A,B)>[C,A]$	by 7 in Def 8.3.7
3.	$<(a)(C,(C\backslash B)),(e)(B,(C\backslash B)),(a)(A,B)>[C,A]$	by e-conversion
4.	$<(a)(C,(C\backslash B)),(e)(A,(C\backslash B))>[C,A]$	by 5 in Def 8.3.4
5.	$<(a)(C,(C\backslash B)),(e)((C\backslash B),A)>[C,A]$	by e-conversion
6.	$<(i)((C\backslash B),C),(e)((C\backslash B),A)>[C,A]$	by a-conversion
7.	$<(i)(C,(C\backslash B)),(e)((C\backslash B),A)>[C,A]$	by i-conversion
8.	$<(o)(C,(C\backslash B))>[C,A]$	by *Ferio*

Bocardo is a little simpler, as follows:

1.	$<(a)(B,C),(o)(B,A)>[C,A]$	posit
2.	$<(a)(B,C),(a)(B,(B\backslash A)),(e)((B\backslash A),A)>[C,A]$	by 7 in Def 8.3.7
3.	$<(i)(C,B),(a)(B,(B\backslash A)),(e)((B\backslash A),A)>[C,A]$	by i-conversion
4.	$<(i)(C,(B\backslash A)),(e)((B\backslash A),A)>[C,A]$	by *Darii*
5.	$<(o)(C,A)>[C,A]$	by *Ferio*

Theorem 8.3.11 *$\mathbb{R}3^{=}$ is cartesian, but $\mathbb{R}3$ is not cartesian.*

Proof. I leave the fact that $\mathbb{R}3^{=}$ is cartesian to the reader.

The rule

$$<(o)(B,A)>[B,A] \Rightarrow <(a)(B,(B\backslash A)),(e)((B\backslash A),A)>[B,A] \tag{8.3.2}$$

is not cartesian, because $(a)(B,(B\backslash A))$ entails $\exists x Bx$ but $(o)(B,A)$ is true when $\exists x Bx$ is false. □

It might be argued that this Theorem is superficial. We have used the new rule only to derive *Baroco* and *Bocardo*, and so we could equally well have given rules that add to the new rule the other sentence of *Baroco* or *Bocardo*—for example in the case of *Bocardo*

$$\begin{array}{l} <(a)(B,C),(o)(B,A)>[C,A] \Rightarrow \\ <(a)(B,C),(a)(B,(B\backslash A)),(e)((B\backslash A),A)>[C,A]. \end{array} \tag{8.3.3}$$

This rule is cartesian; we can derive $\exists x Bx$ from the first sentence $(a)(B, C)$.

But this argument fails for *Baroco*. The rule taking line 1 to line 2 in the derivation of *Baroco* above is not cartesian, since line 2 implies $\exists x Cx$ and line 1 doesn't.

Theorem 8.3.12 *The rule-book $\mathbb{R}3$ is sound and complete for the class of all assertoric premise-sequences.*

Proof. For completeness, first prove by $\mathbb{R}1$ and then use $\mathbb{R}3$ to remove the non-first-figure syllogisms. □

Observe again that there are no fourth-figure syllogisms in $\mathbb{R}3$, showing the pointlessness of Ibn Sīnā's rejection of fourth-figure syllogisms.

Note that even if the rules were cartesian, this would be no use for proving soundness, since we have to prove $\rhd$ rather than $\vdash$.

History 8.3.13 Ibn Sīnā also lists, following Aristotle, ways in which some assertoric moods can be derived from others by contraposition:

If $\phi, \psi \vdash \chi$ then $\neg\chi, \psi \vdash \neg\phi$ (and likewise $\phi, \neg\chi \vdash \neg\psi$).

(See Appendix A where these derivations are listed.) In Ibn Sīnā's view, contraposition is a device in propositional logic. So the proper place to see how it fits into a formal system is in connection with reductio ad absurdum in propositional logic; accordingly we deal with it in Chapter BELOW. Ibn Sīnā's account of reductio ad absurdum is one of two places where Ibn Sīnā attempts to combine rules of two different logics into a single formal system (and for the other place see History 10.4.1 below), very likely for the first time in the history of logic; this is another reason for giving separate attention to the issue.

8.4 Exercises

8.1. Show that no productive assertoric premise-sequence contains a sentence of the form (o) backwards.

Solution. Suppose to the contrary that $(o)(D, C)$ occurs backwards in the productive premise-sequence $T[B, A]$. Then by Lemma 8.1.7, the initial segment $U[B, D]$ of $T[B, A]$ whose final sentence is $(o)(D, C)$ is productive. But $U[B, D]$ contains a negative sentence and D is undistributed in $U[B, D]$, contradicting Theorem 7.2.6(b)(ii).

Part III

Two-dimensional

Chapter 9

Two-dimensional logic

9.1 The sentence forms

Definition 9.1.1 We write $\mathcal{L}_{2d}$ for *two-dimensional logic*, or for brevity *2D logic*. The logic $\mathcal{L}_{2d}$ is a subject-predicate logic (cf. Definition 3.1.1). As Ibn Sīnā presents it, the boundaries of the logic are a little hazy. But we need a precise definition and so we will make one. 2D logic has $4 \times 5 = 20$ sentence forms, all got from the schema

$$(f\text{-}g) \tag{9.1.1}$$

by putting one of a, e, i, o for f, and one of d, ℓ, m, t, z for g. We call (f) (i.e. (a) or (e) or (i) or (o) as appropriate) the *assertoric form* of the sentence, and (g) its *avicennan form*. By *core 2D logic* we mean the restriction of 2D logic where sentences with the avicennan form (z) are not used.

Definition 9.1.2 (a) By a *two-dimensional signature*, or for short a *2D signature*, we mean a two-sorted signature with first sort $object$ and second sort $time$, in which all the relation symbols are binary with first sort $object$ and second sort $time$; we write A, B etc. for these relation symbols. We sometimes write 2D signatures Σ as Σ_{2d} to distinguish them from the signatures used in assertoric logic.

(b) If Σ is a two-dimensional signature and δ a constant of the sort $time$, then we write $\Sigma(\delta)$ for the signature got from Σ by adding δ. Likewise we write $\Sigma(E)$ for the signature got from Σ by adding the binary relation symbol E; this is a reserved symbol used with a standard meaning to be explained at the end of this section.

(c) The domain of the 2D sentence forms is the class of pairs (B, A) of relation symbols in any 2D signature, not including the reserved relation symbol E.

(d) If Σ is a 2D signature, we write $L(\Sigma)$ for the corresponding two-sorted first-order language; we assume that neither $=$ nor $\perp$ is a symbol of $L(\Sigma)$. In $L(\Sigma)$, for variables and individual constants of the sort $object$ we will use lower-case latin letters: variables x, y, z and constants a, b, c. For the sort $time$ we will use lower-case greek letters: variables σ, τ and constants α, β.

(e) If Σ is a 2D signature, then by a Σ-*structure* we mean a two-sorted structure M with a pair of nonempty domains $\text{dom}_o(M)$ and $\text{dom}_t(M)$, such that in M each relation symbol A of Σ has an interpretation A^M which is a subset of $\text{dom}_o(M)\times$ $\text{dom}_t(M)$. The same applies to $\Sigma(\delta)$, with the difference that a $\Sigma(\delta)$-structure also carries an assignment of δ to an element δ^M of its domain of sort $time$.

Definition 9.1.3 The sentences of 2D logic are as in Figure 9.1 below; the sentences in core 2D logic are those above the horizontal line. The sentences of core 2D logic are sentences of $L(\Sigma(E))$ for some 2D signature Σ; the sentences of 2D logic but not in the core are sentences of $L(\Sigma(\delta))$.

We will spend the next few sections analysing this list of forms. Notice how each form is got by taking an assertoric sentence form and adding temporal features. This fact allows us to define a projection from 2D logic to assertoric logic:

Definition 9.1.4 Let $(f\text{-}g)(B, A)$ be a 2D sentence. Then the *assertoric projection* of this sentence is the sentence $(f)(B, A)$.

It also allows us to carry over some assertoric terminology to the 2D case, as follows.

Definition 9.1.5 (a) A 2D sentence $(f\text{-}g)(B, A)$ is classified as *universal*, *existential*, *affirmative* or *negative* according to its assertoric form (f). So for example the sentence $(o\text{-}d)(B, A)$ has assertoric form (o) and hence is negative existential.

(b) The clauses $\exists x\exists\tau Bx\tau$ and $\exists x Bx\delta$ in the 2D sentences of assertoric form (a) are called the *existential augments*, and the clauses $\forall x\forall\tau\neg Bx\tau$ and $\forall x\neg Bx\delta$ in the 2D sentences of assertoric form (o) are called the *universal augments*. (See Chatti [19] on these augments.)

form	sentence
$(a\text{-}d)(B, A)$	$(\forall x(\exists\tau Bx\tau \to \forall\tau(Ex\tau \to Ax\tau)) \wedge \exists x\exists\tau Bx\tau)$
$(a\text{-}\ell)(B, A)$	$(\forall x(\exists\tau Bx\tau \to \forall\tau(Bx\tau \to Ax\tau)) \wedge \exists x\exists\tau Bx\tau)$
$(a\text{-}m)(B, A)$	$(\forall x(\exists\tau Bx\tau \to \exists\tau(Bx\tau \wedge Ax\tau)) \wedge \exists x\exists\tau Bx\tau)$
$(a\text{-}t)(B, A)$	$(\forall x(\exists\tau Bx\tau \to \exists\tau(Ex\tau \wedge Ax\tau)) \wedge \exists x\exists\tau Bx\tau)$
$(e\text{-}d)(B, A)$	$\forall x(\exists\tau Bx\tau \to \forall\tau(Ex\tau \to \neg Ax\tau))$
$(e\text{-}\ell)(B, A)$	$\forall x(\exists\tau Bx\tau \to \forall\tau(Bx\tau \to \neg Ax\tau))$
$(e\text{-}m)(B, A)$	$\forall x(\exists\tau Bx\tau \to \exists\tau(Bx\tau \wedge \neg Ax\tau))$
$(e\text{-}t)(B, A)$	$\forall x(\exists\tau Bx\tau \to \exists\tau(Ex\tau \wedge \neg Ax\tau))$
$(i\text{-}d)(B, A)$	$\exists x(\exists\tau Bx\tau \wedge \forall\tau(Ex\tau \to Ax\tau))$
$(i\text{-}\ell)(B, A)$	$\exists x(\exists\tau Bx\tau \wedge \forall\tau(Bx\tau \to Ax\tau))$
$(i\text{-}m)(B, A)$	$\exists x(\exists\tau Bx\tau \wedge \exists\tau(Bx\tau \wedge Ax\tau))$
$(i\text{-}t)(B, A)$	$\exists x(\exists\tau Bx\tau \wedge \exists\tau(Ex\tau \wedge Ax\tau))$
$(o\text{-}d)(B, A)$	$(\exists x(\exists\tau Bx\tau \wedge \forall\tau(Ex\tau \to \neg Ax\tau)) \vee \forall x\forall\tau\neg Bx\tau)$
$(o\text{-}\ell)(B, A)$	$(\exists x(\exists\tau Bx\tau \wedge \forall\tau(Bx\tau \to \neg Ax\tau)) \vee \forall x\forall\tau\neg Bx\tau)$
$(o\text{-}m)(B, A)$	$(\exists x(\exists\tau Bx\tau \wedge \exists\tau(Bx\tau \wedge \neg Ax\tau)) \vee \forall x\forall\tau\neg Bx\tau)$
$(o\text{-}t)(B, A)$	$(\exists x(\exists\tau Bx\tau \wedge \exists\tau(Ex\tau \wedge \neg Ax\tau)) \vee \forall x\forall\tau\neg Bx\tau)$
$(a\text{-}z)(B, A)$	$(\forall x(Bx\delta \to Ax\delta) \wedge \exists xBx\delta)$
$(e\text{-}z)(B, A)$	$\forall x(Bx\delta \to \neg Ax\delta)$
$(i\text{-}z)(B, A)$	$\exists x(Bx\delta \wedge Ax\delta)$
$(o\text{-}z)(B, A)$	$(\exists x(Bx\delta \wedge \neg Ax\delta) \vee \forall x\neg Bx\delta)$

Figure 9.1: Two-dimensional sentences

History 9.1.6 The name 'two-dimensional' is not from Ibn Sīnā. It is adapted from Oscar Mitchell [77], a student and then colleague of C. S. Peirce at Rutgers in the early 1880s. See Dipert [22] for biographical information on him. Mitchell independently noticed a rather cruder family of sentences that involve temporal quantification, for example 'All the Browns were ill during some part of the year'. Ibn Sīnā's two-dimensional sentences are compared with Mitchell's in [43].

In the course of the next few chapters we will make a systematic study of the logical relationships between pairs of 2D sentences. But it will be helpful to identify the contradictory negations at once.

Lemma 9.1.7 *Write $a' = o, e' = i, i' = e, o' = a, d' = t, \ell' = m, m' = \ell, t' = d, z' = z$. Then every 2D sentence $(f\text{-}g)(B, A)$ has a contradictory negation $(f'\text{-}g')(B, A)$.*

The *proof* is by inspection. □

History 9.1.8 Street claims at [96] p. 47 and [97] pp. 138, 156 that Ibn Sīnā gave an incorrect contradictory negation for $(a\text{-}\ell)$, and that Rāzī and Ṭūsī took him to task for this. None of this claim is supported by the texts that Street cites. One can verify from Street's own translation at [97] p. 138 that Ibn Sīnā has the contradictory negation correctly; see also *Mašriqiyyūn* 81.1–10 where Ibn Sīnā discusses the contradictory negation of (m). The passage of Ṭūsī's commentary that Street cites at [96] p. 47f says only that the reader of Ibn Sīnā should avoid a certain mistake, not that Ibn Sīnā has made this mistake. The passage of Rāzī's *Lubāb* that Street cites at [97] p. 156 again gives the contradictory negation of the $(a\text{-}\ell)$ form correctly, and contains no suggestion that Ibn Sīnā said anything different.

But this may not be the whole story. If my very weak Persian doesn't let me down, Ṭūsī on page 147 of his *Asās al-Iqtibās* [100] says that according to Ibn Sīnā, 'if there are absolute conventionals that differ as negative and affirmative then they are mutually contradictory'. (I think by absolute conventionals he means (ℓ) sentences.) The appropriate comment on this seems to be that our texts of Ibn Sīnā don't confirm Ṭūsī's statement. Since Ṭūsī is reported to have written *Asās* in 1244/5 and the *Išārāt* commentary in 1246 (Landolt [69] p. 13), it's conceivable that he realised his error before finishing the latter work.

The two-dimensional sentences are already very much present in Ibn Sīnā's earliest surviving major work in logic, the *Muḵtaṣar* [58]. But here they are jumbled up with alethic modal sentences so as to create the *ḵabṭ* that Rāzī complained of (cf. Section 1.1 above). For example

(9.1.2) Know that 'impossible' means permanently absent, either absolutely or under a condition so that the absence persists for as long as that condition holds. (*Muḵtaṣar* 32a12f)

(9.1.3) ... it has been stated that A is true of everything fitting the description B for as long as it fits the description B, and [that] C is permanently a B. So [the thing] will be with necessity an A; and this necessity is proved of it not by the major premise [alone] but from the fact that the argument proves it. (*Muḵtaṣar* 53b16f)

For most pieces of text along these lines, we can suggest one or more plausible ways of separating out the alethic and the temporal. But when we try to reconcile the various texts into one coherent whole, we rapidly hit a combinatorial explosion. No wonder Rāzī complained.

Qiyās and *Mašriqiyyūn* were written just a few years apart. In both of them Ibn Sīnā takes a major step towards sorting out the confusion—he introduces the two-dimensional sentences *before* he starts to discuss alethic modalities. He signals in this way that the two-dimensional logic is meant to stand on its own two feet. In fact the list of sentence forms in Figure 9.1 above was derived mainly from these two passages, *Qiyās* i.3, 21.13–24.1 and *Mašriqiyyūn* 68.1–70.13, which occupy similar positions in the expositions of *Qiyās* and *Mašriqiyyūn*. In *Qiyās* Ibn Sīnā gives example sentences in Arabic with some explanatory comments, while in *Mašriqiyyūn* he sets out the sentence forms more abstractly. The two listings correlate closely, so we can match up the abstract forms with the examples.

The book [45] will go through this in more detail, but for the present here are some examples of the (a) forms. With the natural language examples we are expected to ask ourselves what we would reckon we are being told if we met them in a piece of scientific discourse; this question of interpretation of discourse is a constant theme in Ibn Sīnā.

(9.1.4)

$(a\text{-}d)$	Every whiteness is a colour. (*Qiyās* 21.16f)
$(a\text{-}\ell)$	Every white thing has a colour dispersed to the eye. (*Qiyās* 22.9)
$(a\text{-}m)$	Everyone who travels from Rayy [in Tehran] to Baghdad reaches Kermanshah. (*Qiyās* 22.12)
$(a\text{-}t)$	Everything that breathes in breathes out. (*Qiyās* 23.5)

The $(a\text{-}d)$ sentence would normally be read as stating a timeless truth, unlike say 'Both my aunts are staying with me' which has a similar syntax but would be understood as referring just to the present. The sentence in the $(a\text{-}\ell)$ case is one of Ibn Sīnā's favourite examples. The reference to 'dispersed to the eye' belongs to a physics of colour that we no longer accept, but the point is that the physical properties of whiteness belong to a white object only for as long as the thing stays white. In the case of the $(a\text{-}m)$ sentence, Kermanshah is a town near the border between Iran and Iraq on the main road from Tehran to Baghdad, and the point is that a traveller from Tehran to Baghdad reaches Kermanshah at some time while travelling from Tehran to Baghdad. The $(a\text{-}t)$ example cleverly forces us to interpret 'breathes out' as 'breathes out sometimes', because a thing can't breathe in and out simultaneously.

Note that we might meet the second sentence in a physics text (or in Ibn Sīnā's arrangement of the sciences, a psychology text), the third in a geography text and the fourth in a biology text. So Ibn Sīnā conveys that the two-dimensional sentences are ones that we should expect to meet in rational theoretical discourse.

In both *Qiyās* i.4, 31.15–32.1 and *Mašriqiyyūn* 68.6, Ibn Sīnā describes the (d) sentences as being 'necessary' (*ḍarūrī*) even though the word 'necessary' doesn't occur in them. We should take this as a technical term. But then when we come to a later place where Ibn Sīnā talks about sentences that are 'necessary', we have to decide whether he means (d) sentences or whether he is using 'necessary' in some more concrete or metaphysical sense. So the *k̲abṭ* persists. What is new, thanks to *Qiyās* and *Mašriqiyyūn*, is that we can build up two-dimensional logic in its own right as a distinguishable component of the *k̲abṭ*. This is the single most important point for making sense of Ibn Sīnā's formal logic; we will spell the point out in Section 12.1 below.

History 9.1.9 In fact the 'd' in (d) stands for *ḍarūrī*. In *Mašriqiyyūn* Ibn Sīnā makes an attempt to set up systematic names for the two-dimensional sentence forms. Most of these names didn't survive to *Išārāt* [61], so he must have decided they were not a success. But they do provide us with convenient letters to name the avicennan forms. Thus (d) is for *ḍarūrī* 'necessary', (ℓ) is for *lāzim* 'adherent', (m) is for *muwāfiq* 'compatible', and (t) is for the *ṭ* in *muṭlaq* *ᶜāmm* 'broad absolute'. The (z) is for *zamānī* which here means 'at a specific time, above all the present time' (*Mašriqiyyūn* 72.7f).

History 9.1.10 Ibn Sīnā's examples don't always answer the questions we would want answered today, but he chose them carefully to make the points he wanted to make. The example 'Every whiteness (*bayāḍ*) is a colour' in (9.1.4) above is a case in point. In *Maqūlāt* [53] Ibn Sīnā has said repeatedly that distinctions of (Aristotelian) category are irrelevant to formal logic. (See Gutas' commentary [34] pp. 300–303.) These distinctions include the distinction between substance and accident. And of course Ibn Sīnā is right about this; we can check through hundreds of examples in his logical writings, and there is not a single case where the validity of an inference depends on whether a term applies to substances or to accidents. To a modern logician it might seem strange that anybody ever thought otherwise; but Ibn Sīnā seems to have felt that his predecessor Al-Fārābī had missed the point. The present example sentence, put in a prominent place in *Qiyās*, makes the point again. For Ibn Sīnā a whiteness is an accident (more specifically a quality), not a substance—see for example *Maqūlāt* 20.12–15, 35.15ff, 117.1. The example is both a (d) sentence and a conceptual necessity, so it knocks out the suggestion—which I have seen made—that Ibn Sīnā's remarks about the category-free nature of logic don't apply to the logic of necessity.

In his explanations of the (d) forms, Ibn Sīnā often uses a strange phrase: 'for as long as its essence (*d̲āt*) is satisfied'. From his examples it's clear that he means 'for as long as it exists'. Presumably the essence is the individual essence of the individual, and to say that this essence is satisfied is equivalent to saying that the individual exists. Presumably also one of his reasons for using this strange phrase was to call attention to a notion that he wanted to emphasise. In any case the appropriate modern formalisation of the phrase is not in doubt, using $Ex\tau$ to mean 'x exists at time τ'.

But I should note one widespread misunderstanding. A number of published works refer to Ibn Sīnā's (d) sentences as 'substantial', apparently mistranslating Ibn Sīnā's word *ḏāt* 'essence' as 'substance'. It's hard to see how this came about, particularly in view of History 9.1.10 above. Al-Fārābī does say that *jawhar* (the normal Arabic word for 'substance') is sometimes used to mean essence (*Ḥurūf* [28] 63.9), and Ibn Sīnā confirms this at *Ḥudūd* Definition 15 ([62] p. 23) and at *Qiyās* 22.3. But if Ibn Sīnā ever goes the other way and uses *ḏāt* to mean substance—and Goichon [33] records no cases where he does—it would need an extremely strong argument to show that Ibn Sīnā has this in mind when he uses the word *ḏāt* in connection with the relation E. Goichon [33] pp. 134, 136 describes the translation of *ḏāt* by 'substantia' as an unfortunate and confusing error, and I can only agree.

Turning to the formal properties of E, we meet here our first example of meaning postulates (cf. Definition 2.2.1), as follows.

Definition 9.1.11 Let Σ_{2d} be a 2D signature. Then the Σ_{2d}-*theory of* E, for short the *theory of* E, in symbols Th(E), is the following set of sentences:

$$\begin{array}{l} \forall x \forall \tau (Ax\tau \to Ex\tau) \text{ (for each } A \text{ in } \Sigma_{2d}). \\ \forall x \exists \tau Ex\tau. \end{array} \tag{9.1.5}$$

From this theory we can deduce the equivalences

$$\begin{array}{rcl} Ax\tau & \equiv & (Ex\tau \wedge Ax\tau) \\ \neg Ax\tau & \equiv & (Ex\tau \to \neg Ax\tau) \end{array} \tag{9.1.6}$$

where A is any symbol in Σ_{2d}.

History 9.1.12 According to *ᶜIbāra* [54] 79.11–80.12, affirmative statements are true of a thing only while that thing exists. This accounts for the first kind of sentence in the theory of E. The second expresses that we consider only objects that do exist at some time. Ibn Sīnā says in several places that he will quantify only over things that are at some time actual; see for example *Qiyās* 20.14–21.12, 183.5–11, 209.3–6.

An immediate application:

Lemma 9.1.13 *Assuming the theory of E, we have the logical progression*

$$(f\text{-}d)(B,A) \;\vdash\; (f\text{-}\ell)(B,A) \;\vdash\; (f\text{-}m)(B,A) \;\vdash\; (f\text{-}t)(B,A)$$

for each assertoric form (f). (By serendipity the avicennan forms are in alphabetical order here.)

Proof. We prove this when (f) is (a). First assume $(a\text{-}d)(B,A)$, and let b be any object such that $Bb\alpha$ for some α. Then for every β such that $Eb\beta$ we also have $Ab\beta$. To prove $(a\text{-}\ell)(B,A)$, suppose $Bb\beta$. By the theory of E, if $Bb\beta$ then $Eb\beta$, and hence $Ab\beta$ as required.

Next assume $(a\text{-}\ell)(B,A)$. Let b be any object such that $Bb\alpha$ holds for some α. Then for each β, if $Bb\beta$ then $Ab\beta$. In particular $Ab\alpha$, proving $(a\text{-}m)(B,A)$.

Finally assume $(a\text{-}m)(B,A)$. Let b be any object such that $Bb\alpha$ holds for some α. Then there is some β such that $Bb\beta$ and $Ab\beta$. By the theory of E again, $Bb\beta$ implies $Eb\beta$. So there is β such that $Eb\beta$ and $Ab\beta$, proving $(a\text{-}t)(B,A)$.

The cases of (e), (i) and (o) are similar and are left to the reader. □

9.2 The modal paraphrase

Ibn Sīnā didn't have first order logic, so of course the sentences in Figure 9.1 are ours and not his. But there is another way of formalising Ibn Sīnā's 2D Arabic sentences within today's logic. It's a notational variant of the formalisation in Section 9.1, so there is no substantive issue about which of the two is correct. But the two do suggest different approaches, and some problems of metatheory may be easier to solve for one than for the other.

Definition 9.2.1 Let Σ be a monadic relational signature. Then we define a *Kripke Σ-frame* to be a quadruple $(\mathcal{D},\mathcal{W},\mathcal{S},\mathcal{E})$ as follows:

(a) $\mathcal{D}$ and $\mathcal{W}$ are nonempty sets;

(b) for each $w\in\mathcal{W}$, $\mathcal{S}_d$ is a Σ-structure with domain $\mathcal{D}$, and $\mathcal{E}_d$ is a subset of $\mathcal{D}$.

(Strictly we should call this object a Kripke frame for monadic first-order logic with constant universe, universal accessibility and an actuality predicate. But this is the only kind of Kripke frame that will concern us.) The elements w of $\mathcal{W}$ are called *worlds*.

Definition 9.2.2 Suppose Σ_{2d} is a two-dimensional signature. Then we write $\mathcal{K}(\Sigma_{2d})$ for the monadic relational signature got from Σ_{2d} by regarding each of the binary relation symbols of Σ_{2d} as a monadic relation symbol. Likewise if Σ_{1d} is a monadic relational signature, then by $\mathcal{J}(\Sigma_{1d})$ we mean the two-dimensional signature got by regarding each of the monadic relation symbols of Σ_{1d} as a binary relation symbol with first sort $object$ and second sort $time$.

Clearly the maps $\mathcal{K}$ and $\mathcal{J}$ on signatures are mutual inverses.

Definition 9.2.3 Suppose Σ_{2d} is a two-dimensional signature; we write Σ_{1d} for $\mathcal{K}(\Sigma_{2d})$. Let M be a $\Sigma_{2d}(E)$-structure. We construct a Kripke Σ_{1d}-frame $\mathcal{K}(M)$ as follows. $\mathcal{K}(M)$ is $(\mathcal{D}, \mathcal{W}, \mathcal{S}, \mathcal{E})$ where

(a) $\mathcal{D}$ and $\mathcal{W}$ are respectively the domain of sort $object$ and the domain of sort $time$ in M.

(b) For each $w \in \mathcal{W}$, $\mathcal{S}_w$ is the Σ_{1d}-structure whose domain is $\mathcal{D}$, with each relation symbol A interpreted as

(9.2.1) $$A^{\mathcal{S}_w} = \{a \in \mathcal{D} : (a, w) \in A^M\}.$$

(c) For each $w \in \mathcal{W}$,

(9.2.2) $$\mathcal{E}_w = \{a \in \mathcal{D} : (a, w) \in E^M\}.$$

Definition 9.2.4 Suppose Σ_{1d} is a monadic relational signature and $(\mathcal{D}, \mathcal{W}, \mathcal{S}, \mathcal{E})$ is a Kripke Σ_{1d}-frame. Write Σ_{2d} for $\mathcal{J}(\Sigma_{1d})$. We construct a $\Sigma_{2d}(E)$-structure $\mathcal{J}(\mathcal{D}, \mathcal{W}, \mathcal{S}, \mathcal{E})$, or for short M, as follows.

(a) The domains of sort $object$ and $time$ in M are respectively $\mathcal{D}$ and $\mathcal{W}$.

(b) For each relation symbol A of Σ_{2d},

(9.2.3) $$A^M = \{(a, w) \in \mathcal{D} \times \mathcal{W} : a \in A^{\mathcal{S}_w}\}.$$

(c)

(9.2.4) $$E^M = \{(a, w) \in \mathcal{D} \times \mathcal{W} : a \in \mathcal{E}_w\}.$$

Lemma 9.2.5 *The map $\mathcal{K}$ defined in Definition 9.2.3 and the map $\mathcal{J}$ defined in Definition 9.2.4 are mutual inverses, setting up a correspondence between $\Sigma_{2d}(E)$-structures and Kripke Σ_{1d}-frames.* □

Kripke frames are associated with languages that have modal operators.

Definition 9.2.6 Let Σ be a monadic relational signature. In Definition 2.1.14 we correlated Σ-structures with sentences of the first-order language $L(\Sigma)$. We now correlate Kripke Σ-frames with sentences of a modal predicate language $L_{modal}(\Sigma)$.

(a) The terms and atomic formulas of $L_{modal}(\Sigma)$ are the same as those of $L(\Sigma)$, and the set of formulas of $L_{modal}(\Sigma)$ is closed under the usual first-order operations. But also for each formula ϕ of $L_{modal}(\Sigma)$ there are formulas $\Box\phi$ and $\Diamond\phi$, neither of which binds any variable occurrences in ϕ.

(b) $\Box$ and $\Diamond$ are called the *modal operators*. We read $\Box\phi$ as 'Necessarily ϕ' and $\Diamond\phi$ as 'Possibly ϕ'.

(b) The *scope* of an occurrence of a modal operator in a formula is the subformula that begins with this occurrence of the modal operator. A formula of $L_{modal}(\Sigma)$ is said to be *completely modalised* if every atomic formula in it lies within the scope of an occurrence of a modal operator.

Sentences of $L_{modal}(\Sigma)$ are evaluated separately at each world of a Kripke frame; in other words, if w is a world then we have a relation $\models_w$ that expresses 'truth at w'. Details are given in standard references, for example Hughes and Cresswell [49] p. 243, or Chapter 2 of Blackburn, De Rijke and Venema [11]. I summarise what we will need.

Definition 9.2.7 Let Σ be a monadic relational signature, $K = (\mathcal{D}, \mathcal{W}, \mathcal{S}, \mathcal{E})$ a Kripke Σ-frame, w a world of $\mathcal{K}$, $\phi(\bar{x})$ a formula of $L_{modal}(\Sigma)$ and $\bar{a}$ a tuple of elements of $\mathcal{D}$. Then a relation between these ingredients is defined:

$$K \models_w \phi[\bar{a}], \tag{9.2.5}$$

in words, '$\bar{a}$ satisfies $\phi(\bar{x})$ at w in $\mathcal{K}$'. The definition is by induction on the complexity of ϕ, using the standard first-order clauses, together with two modal clauses:

$$\begin{array}{lcl} K \models_w \Box\phi[\bar{a}] & \Leftrightarrow & \text{for every world } u, K \models_u \phi[\bar{a}]; \\ K \models_w \Diamond\phi[\bar{a}] & \Leftrightarrow & \text{for some world } u, K \models_u \phi[\bar{a}]. \end{array} \tag{9.2.6}$$

Lemma 9.2.8 *If ϕ is a completely modalised formula, then the truth of $K \models_w \phi[\bar{a}]$ doesn't depend on w.*

Proof. Left to the reader. □

Definition 9.2.9 If ϕ is a completely modalised formula, then we write $K \models \phi[\bar{a}]$ to mean that $K \models_w \phi[\bar{a}]$ for some world w of K (or equivalently, for some world w of K, since the set of worlds is not empty).

Theorem 9.2.10 *Let Σ_{2d} be a 2D signature and $\phi(\bar{x})$ a formula of $L_{modal}(\Sigma_{2d})$. Then*

(a) There is a formula $\phi^{2d}(\bar{x}, \delta)$ of $L(\Sigma_{2d})$ such that for every Σ_{2d}-structure M, every tuple $\bar{a}$ of objects of M and every time β of M,

(9.2.7) $$\mathcal{K}(M) \models_\beta \phi[\bar{a}] \;\Leftrightarrow\; M \models \phi^{2d}[\bar{a}, \beta].$$

(b) If $\phi(\bar{x})$ is completely modalised then ϕ^{2d} can be chosen in the form $\phi^{2d}(\bar{x})$, i.e. with no free occurrence of σ.

Proof. (a) is proved by induction on the complexity of ϕ, following the standard definition of satisfaction for modal predicate logic with constant universe and universal accessibility relation. Then if ϕ is completely modalised, it follows from Lemma 9.2.8 that we can put $\forall\sigma$ at the beginning of ϕ^{2d}. □

Theorem 9.2.11 *Each two-dimensional sentence is of the form ϕ^{2d} for some formula ϕ of $\mathcal{L}_{modal}$, with the exception that for the sentences of the form (z) we write a time constant δ in place of the free time variable σ. The mapping is as in Figure 9.2 below.* □

It's important to stress once again that the map $\phi \mapsto \phi^{2d}$ is purely a change of notation. For example there is no justification for reading some particular kind of necessity into $\Box$ and then using Theorem 9.2.11 to transfer this notion of necessity into the two-dimensional sentences. But of course changes of notation can be valuable. In this case a change from ϕ to ϕ^{2d} may suggest useful modal methods. The converse change from ϕ^{2d} to ϕ gives extra structure in terms of quantifiers and variables, and we will make extensive use of this in Chapters BELOW.

ϕ	ϕ^{2d}
$(\forall x(\Diamond Bx \to \Box(Ex \to Ax)) \wedge \exists x Bx)$	$(a\text{-}d)(B,A)$
$\forall x \Box(Bx \to Ax)$	$(a\text{-}\ell)(B,A)$
$\forall x(\Diamond Bx \to \Diamond(Bx \wedge Ax))$	$(a\text{-}m)(B,A)$
$\forall x(\Diamond Bx \to \Diamond(Ex \wedge Ax))$	$(a\text{-}t)(B,A)$
$\forall x(\Diamond Bx \to \Box(Ex \to \neg Ax))$	$(e\text{-}d)(B,A)$
$\forall x \Box(Bx \to \neg Ax))$	$(e\text{-}\ell)(B,A)$
$\forall x(\Diamond Bx \to \Diamond(Bx \wedge \neg Ax))$	$(e\text{-}m)(B,A)$
$\forall x(\Diamond Bx \to \Diamond(Ex \wedge \neg Ax))$	$(e\text{-}t)(B,A)$
$\exists x(\Diamond Bx \wedge \Box(Ex \to Ax))$	$(i\text{-}d)(B,A)$
$\exists x(\Diamond Bx \wedge \Box(Bx \to Ax))$	$(i\text{-}\ell)(B,A)$
$\exists x \Diamond(Bx \wedge Ax)$	$(i\text{-}m)(B,A)$
$\exists x(\Diamond Bx \wedge \Diamond(Ex \wedge Ax))$	$(i\text{-}t)(B,A)$
$\exists x(\Diamond Bx \wedge \Box(Ex \to Ax))$	$(o\text{-}d)(B,A)$
$\exists x(\Diamond Bx \wedge \Box(Bx \to \neg Ax))$	$(o\text{-}\ell)(B,A)$
$\exists x \Diamond(Bx \wedge \neg Ax)$	$(o\text{-}m)(B,A)$
$\exists x(\Diamond Bx \wedge \Diamond(Ex \wedge Ax))$	$(o\text{-}t)(B,A)$
$(\forall x(Bx \to Ax) \wedge \exists x Bx)$	$(a\text{-}z)(B,A)$
$\forall x(Bx \to \neg Ax)$	$(e\text{-}z)(B,A)$
$\exists x(Bx \wedge Ax)$	$(i\text{-}z)(B,A)$
$(\exists x(Bx \wedge \neg Ax) \vee \forall x \neg Bx)$	$(o\text{-}z)(B,A)$

Figure 9.2: Modalised two-dimensional sentences

9.3 At most two sentences

Theorem 9.3.1 *Assuming Th(E) as meaning postulates, the following entailments hold between pairs of core two-dimensional sentences with a given subject relation symbol and a given predicate relation symbol; they and the ones they imply by transitivity and reflexivity of entailment are the only such entailments.*

$$\begin{array}{ccccccc}
(a\text{-}d) & \Rightarrow & (a\text{-}\ell) & \Rightarrow & (a\text{-}m) & \Rightarrow & (a\text{-}t) \\
\Downarrow & & \Downarrow & & \Downarrow & & \Downarrow \\
(i\text{-}d) & \Rightarrow & (i\text{-}\ell) & \Rightarrow & (i\text{-}m) & \Rightarrow & (i\text{-}t) \\
 & & & & & & \\
(e\text{-}d) & \Rightarrow & (e\text{-}\ell) & \Rightarrow & (e\text{-}m) & \Rightarrow & (e\text{-}t) \\
\Downarrow & & \Downarrow & & \Downarrow & & \Downarrow \\
(o\text{-}d) & \Rightarrow & (o\text{-}\ell) & \Rightarrow & (o\text{-}m) & \Rightarrow & (o\text{-}t)
\end{array} \tag{9.3.1}$$

Proof. The lower half of (9.3.1) follows from the upper by taking contradictory negations as in Lemma 9.1.7, so we concentrate on the upper half.

The vertical entailments all have the pattern

$$(\forall x(\exists\tau Bx\tau \to \phi(x)) \wedge \exists x\exists\tau Bx\tau) \;\vdash\; \exists x(\exists\tau Bx\tau \wedge \phi(x)). \quad (9.3.2)$$

This is clearly valid. The horizontal entailments are included in Lemma 9.1.13 above.

To prove the failures of entailment, let $L(\Sigma)$ be a language containing the sentences in question. First take a Σ-structure M_1 in which there are just two objects a, b and one time α, and the following hold: $Ea\alpha$, $Eb\alpha$, $Ba\alpha$, $Bb\alpha$, $Aa\alpha$, $\neg Ab\alpha$. Then

$$M_1 \;\models\; (i\text{-}d)(B,A) \wedge \neg(a\text{-}t)(B,A). \quad (9.3.3)$$

This prevents there being any other entailments between the sentences in the top half of (9.3.1).

Next take a Σ-structure M_2 with just the one object a and the one time α, such that the following hold: $Ea\alpha$, $Ba\alpha$, $Aa\alpha$. Then

$$M_2 \;\models\; (a\text{-}d)(B,A) \wedge \neg(o\text{-}t)(B,A). \quad (9.3.4)$$

Hence there is no entailment from any sentence in the top rows to any sentence in the bottom rows.

Finally take a Σ-structure M_3 with just the one object a and the one time α, such that the following hold: $Ea\alpha$, $\neg Ba\alpha$, $\neg Aa\alpha$. Then

$$M_2 \;\models\; (e\text{-}d)(B,A) \wedge \neg(i\text{-}t)(B,A). \quad (9.3.5)$$

Hence there is no entailment from any sentence in the bottom rows to any sentence in the top rows. □

Theorem 9.3.2 *Among two-dimensional sentences with subject symbol B and predicate symbol A, the only entailments between two distinct sentences of avicennan form (z), or one of avicennan form (z) and one in the core, are the following;*

$$\begin{array}{l} (a\text{-}z)(B,A) \;\Rightarrow\; (i\text{-}z)(B,A) \;\Rightarrow\; (i\text{-}m)(B,A); \\ (e\text{-}\ell)(B,A) \;\Rightarrow\; (e\text{-}z)(B,A) \;\Rightarrow\; (o\text{-}z)(B,A) \end{array} \quad (9.3.6)$$

together with any other entailments that follow from these by Theorem 9.3.1.

Proof. The second row follows from the first by taking contradictory negations, so we concentrate on the first row. The first entailment in the first row holds for the same reason as in the assertoric sentences. For the second, suppose there is an object a which is both a B and an A at time δ; then there is a time at which a is both a B and an A, confirming $(i\text{-}m)(B, A)$.

For the failures of entailment, we note that there are no other entailments between (z) sentences, for the same reason as in the assertoric sentences. Also the arguments with M_2 and M_3 in the proof of Theorem 9.3.1 show that there are no entailments from affirmative to negative, or from negative to affirmative. It remains to show that $(a\text{-}z)(B, A)$ entails no affirmative core sentence with tag (B, A) apart from $(i\text{-}m)(B, A)$ and its consequences, and that no affirmative core sentence with tag (B, A) entails any (z) sentence with the same tag.

For the first of these, it suffices to find a model of $(a\text{-}z)(B, A)$ that is not a model of either $(i\text{-}\ell)(B, A)$ or $(a\text{-}t)(B, A)$. For the second it suffices to find a model of $(a\text{-}d)(B, A)$ that is not a model of $(i\text{-}z)(B, A)$. The reader deserves an exercise here, so let this be it. □

These two theorems allow us to infer facts about covers, cf. Definition 4.1.1 above.

Corollary 9.3.3 *(a)* $\{(a\text{-}d), (e\text{-}d)\}$ *is a cover for core two-dimensional logic but not for all of two-dimensional logic.*

(b) $\{(a\text{-}d), (a\text{-}z), (e\text{-}d)\}$ *is a cover for the whole of two-dimensional logic.*

□

Theorem 9.3.2 also raises the possibility that there might be premise-pairs in two-dimensional logic that have more than one optimal conclusion (cf. Definition 3.3.2(b)). And indeed there are—see Exercise 9.N below. But this never happens in core 2D logic; see Exercise 11.N.

We turn to conversions within 2D logic, i.e. conversions of 2D sentences to 2D sentences.

Theorem 9.3.4 *We assume Th(E) as meaning postulates.*

(a) The 2D sentences that convert symmetrically (cf. Definition 3.3.9) are those of the forms

(9.3.7) $$(e\text{-}d),\ (e\text{-}\ell),\ (e\text{-}z),\ (i\text{-}m),\ (i\text{-}t),\ (i\text{-}z).$$

(b) *The following sentences are convertible but not symmetrically, and they have the following converses:*

- $(a\text{-}d)(B,A)$, $(a\text{-}\ell)(B,A)$, $(a\text{-}m)(B,A)$, $(i\text{-}d)(B,A)$ *and* $(i\text{-}\ell)(B,A)$ *all have converse* $(i\text{-}m)(A,B)$.
- $(a\text{-}t)(B,A)$ *has converse* $(i\text{-}t)(A,B)$.
- $(a\text{-}z)(B,A)$ *has converse* $(i\text{-}z)(A,B)$.

(c) *The remaining 2D sentences, namely those of the forms* (o), $(e\text{-}m)$ *and* $(e\text{-}t)$, *are not convertible.*

Proof. One can check directly that a sentence of any of the forms listed in (9.3.7) converts symmetrically. Also the sentences listed in (b) entail the sentences given as their converses, by symmetric conversion of the forms in (9.3.7) together with Theorems 9.3.1 and 9.3.2, and hence are convertible. It remains to show (for (a) and (b)) that none of the sentences listed in (b) entails a stronger converse, and (for (c)) that none of the sentences listed in (c) is convertible at all.

We consider first the sentences listed in (b). The five listed as having converse $(i\text{-}m)(A,B)$ are all of them either $(a\text{-}d)(B,A)$ or weakenings of $(a\text{-}d)(B,A)$, so it will suffice to show that $(a\text{-}d)(B,A)$ entails only $(i\text{-}m)(A,B)$ and $(i\text{-}t)(A,B)$ among 2D sentences with tag (A,B). For this it suffices to show that $(a\text{-}d)(B,A)$ doesn't entail either $(a\text{-}t)(A,B)$ or $(i\text{-}\ell)(A,B)$.

We show these as follows. Let Σ be the 2D signature consisting of the relation symbols A and B. Let M_1 be the $\Sigma(E)$-structure with object domain $\{1,2\}$ and time domain $\{\alpha,\beta\}$, in which 1 satisfies A at time α but not time β, and never satisfies B; and 2 satisfies A at both times but B only at α. E holds everywhere. Then M_1 is a model of $(a\text{-}d)(B,A)$ but not of either $(a\text{-}t)(A,B)$ or $(i\text{-}\ell)(A,B)$. Since E holds everywhere, M_1 is a model of Th(E).

For $(a\text{-}t)(B,A)$ we must show that it doesn't entail either $(a\text{-}t)(A,B)$ or $(i\text{-}m)(A,B)$. For this, take the same Σ as above, and let M_2 be the $\Sigma(E)$-structure with object domain $\{1,2\}$ and time domain $\{\alpha,\beta\}$, in which both 1 and 2 satisfy A at α and not at β, and 1 satisfies B at β but not at α, while 2 never satisfies B; E holds everywhere. Then M_2 is a model of $(a\text{-}t)(B,A)$ and Th(E), but not of either $(a\text{-}t)(A,B)$ or $(i\text{-}m)(A,B)$.

For $(a\text{-}z)(B,A)$ we must show it doesn't entail $(a\text{-}z)(A,B)$. For this, take M_3 to be a $\Sigma(E,\delta)$-structure with two objects 1 and 2, and one time δ, where

1 satisfies both A and B, but 2 satisfies only A, and E holds everywhere. Then M_3 is a model of $(a\text{-}z)(B, A)$ and Th(E), but not of $(a\text{-}z)(A, B)$.

Finally we consider the sentences listed in (c), and show that none of them is convertible. It suffices to show this for the sentences of maximal strength in this group, namely those of the forms $(o\text{-}d)(B, A)$, $(o\text{-}z)(B, A)$ and $(e\text{-}m)(B, A)$. In each case it suffices to show that the sentence doesn't entail either of the weakest possible converses, which are $(o\text{-}t)(A, B)$ and $(o\text{-}z)(A, B)$.

Let Σ be, as before, the 2D signature whose symbols are A and B. Let M_4 be the $\Sigma(E, \delta)$-structure with two objects $\{1, 2\}$ and one time $\{\delta\}$, satisfying $A1\delta$, $\neg A2\delta$, $B1\delta$ and $B2\delta$, with E satisfied everywhere. Then M_4 is a model of $(o\text{-}t)(A, B)$ and $(o\text{-}z)(B, A)$ but not of either $(o\text{-}t)(A, B)$ or $(o\text{-}z)(A, B)$. Since E holds everywhere, M_4 is a model of Th(E).

Let M_5 be the $\Sigma(E, \delta)$-structure with two objects $\{1, 2\}$ and two times α, δ, satisfying $\neg A1\alpha$, $A1\delta$, $\neg A2\alpha$, $\neg A2\delta$, $B1\alpha$, $B1\delta$, $B2\alpha$, $\neg B2\delta$, with E holding everywhere. Then M_5 is a model of $(e\text{-}m)(A, B)$ but not of either $(o\text{-}t)(A, B)$ or $(o\text{-}z)(A, B)$. Since E holds everywhere, M_5 is a model of Th(E). □

Ectheses are not entailments between pairs of sentences. But in Ibn Sīnā's proof theory for assertoric logic they appeared alongside the conversions, so we have some excuse for slipping in a mention of them here. This will have to be preliminary. Ecthetic rules are in general not cartesian—because of the universal augments on (o) sentences—and the justification that we offered in Chapter 8 for the use of non-cartesian rules involved some substantial preliminaries that we haven't yet carried over to the two-dimensional case. (See the use of distribution at Lemma 8.3.10.) So for the present we give partial proofs that ignore the universal augments. At BELOW we will give a proper justification. Meanwhile a handwaving justification is available along the lines 'Take separately the case where the subject relation is empty'.

In 2D logic the number of different ways of defining a new term from two given terms is much greater than in assertoric logic, giving greater scope for ingenious applications of ecthesis. I haven't surveyed all the possibilities; the ones listed in the next theorem are just the ones we need for verifying Ibn Sīnā's applications in proof theory.

Theorem 9.3.5 *Let B and A be relation symbols. In the five cases listed below, the definition of D from B and A is in the middle column; this definition together* *with the sentence in the left column and Th(E) justifies introducing the sentences in the right column.*

	premise	$Dx\tau$	*conclusions*
(1)	$(o\text{-}t)(B, A)$	$(\exists\sigma Bx\sigma \wedge \neg Ax\tau)$	$(a\text{-}t)(D, B), (e\text{-}\ell)(A, D)$
(2)	$(o\text{-}d)(B, A)$	$(Bx\tau \wedge \forall\sigma\neg Ax\sigma)$	$(a\text{-}\ell)(D, B), (e\text{-}d)(A, D)$
(3)	$(i\text{-}d)(B, A)$	$(Bx\sigma \wedge \forall\tau(Ex\tau \rightarrow Ax\tau))$	$(a\text{-}\ell)(D, B), (a\text{-}d)(D, A)$

Proof, ignoring universal augments.

(1) Assume $(o\text{-}t)(B, A)$. Then (ignoring the augment) there are an object a such that $\exists\sigma Ba\sigma$ and a time β such that $\neg Aa\beta$. Then $Da\beta$, so

(9.3.8) $$\exists x\exists\tau Dx\tau.$$

Let b be any object such that $\exists\tau Db\tau$. Then $\exists\sigma Bb\sigma$. Together with (9.3.8), this shows that

(9.3.9) $$(a\text{-}t)(D, B).$$

Also if c is an object and γ a time such that $Ac\gamma$, then $\neg Da\gamma$. This proves

(9.3.10) $$(e\text{-}\ell)(A, D).$$

(2) Assume $(o\text{-}d)$. Then there is a such that $\exists\tau Ba\tau$ and $\forall\tau\neg Aa\sigma$ (quoting (9.1.6 for the second sentence), so $\exists x\exists\tau Dx\tau$. Also if b and γ are an object and a time such that $Db\gamma$, then $Bb\gamma$, so $(a\text{-}\ell)(D, B)$. If $\exists\sigma Ac\sigma$ then for all times γ, $\neg Dc\gamma$, so $(e\text{-}d)(A, D)$.

(3) In this case there is no universal augment, so we can give a proper proof. Assume $(i\text{-}d)(B, A)$, and suppose a is such that $Ba\alpha$ for some time α and $Aa\beta$ for all times β at which a exists. Then $Da\alpha$, and for every b such that $Db\gamma$ for some γ, $Ab\beta$ for all times β at which b exists. Thus $(a\text{-}d)(D, A)$. But also if $Db\gamma$ for some time γ, then $Bb\gamma$; hence also $(a\text{-}\ell)(D, B)$. □

9.4 Assertoric-like fragments of 2D logic

We noted that Ibn Sīnā gives a detailed review of Aristotle's account of assertoric syllogisms in four of his surviving logical works (*Muḵtaṣar*, *Najāt*, *Qiyās*, *Dānešnāmeh*). Each of these four accounts appears inside a study of syllogisms from 'absolute' premises. (Cf. History 5.1.4 above.)

In each of these four passages, Ibn Sīnā identifies a problem about the justification of second figure syllogisms. There are several different kinds of absolute sentence, he says, and it's not true that in all these kinds the universal negative sentences convert. This shows, he continues, that the justification of *Cesare*, *Camestres* and *Festino* doesn't always work, and in fact these moods are productive only if we 'take' (*'aka̱dnā*, *Najāt* 59.3) the relevant sentences as convertible. In other words, we need to confine these moods to the appropriate kinds of absolute sentence.

In his discussion of second figure for absolute sentences in *Qiyās* (ii.4, 113.5–114.1), Ibn Sīnā goes on to specify some kinds of absolute sentence where negative universal sentences do convert. At *Qiyās* 114.1 he suggests there are two such kinds, but in fact his text briefly mentions four. One (*Qiyās* 113.9) is the 'standard' (*maš hūr*) case; I take him to mean assertoric sentences. The second (*Qiyās* 113.10) is sentences with wide scope for the time quantifier, as in

$$\exists\tau\forall x(Bx\tau \to \neg Ax\tau) \tag{9.4.1}$$

which is logically equivalent to

$$\exists\tau\forall x(Ax\tau \to \neg Bx\tau). \tag{9.4.2}$$

The third (*Qiyās* 113.11) is the case of sentences with avicennan form $(e\text{-}z)$, and the fourth (113.15) is those with avicennan form $(e\text{-}\ell)$. Besides these four kinds of absolute sentence, Ibn Sīnā also notes that it is 'uncontroversial' (*lā munāza*^{c}, *Qiyās* 131.8f) that negative universal 'with necessity' statements convert; presumably this includes the 'necessary' 2D sentences of the form $(e\text{-}d)$. These last three kinds of sentence are the sentences with assertoric form (e) that we noted in Theorem 9.3.4(a) as converting symmetrically.

So, leaving aside the alethic modals for the present, Ibn Sīnā has pointed out five kinds of universal negative statement that convert. If you check out the four non-assertoric cases you will probably notice a further fact: in each case the convertible universal negative sentence belongs to a family of sentences that behaves very much like assertoric logic. This fact needs to be made more precise. It's not a fact that Ibn Sīnā himself explicitly comments on, but it will be useful for us to pursue it in those three cases that lie within 2D logic.

We begin with the $(e\text{-}\ell)$ forms. The formulas given for them in Figure 9.1 were needlessly complicated; the subject part can be eliminated because it is already expressed in the predicate. The same is true for three other (ℓ) or (m) forms, as follows.

Lemma 9.4.1 *The forms (a-ℓ), (e-ℓ), (i-m) and (o-m) are logically equivalent to the following simpler forms:*

$$\begin{array}{ll} (a\text{-}\ell)(B,A): & (\forall x\forall\tau(Bx\tau \to Ax\tau) \wedge \exists x\exists\tau Bx\tau) \\ (e\text{-}\ell)(B,A): & \forall x\forall\tau(Bx\tau \to \neg Ax\tau) \\ (i\text{-}m)(B,A): & \exists x\exists\tau(Bx\tau \wedge Ax\tau) \\ (o\text{-}m)(B,A): & (\exists x\exists\tau(Bx\tau \wedge \neg Ax\tau) \vee \forall x\forall\tau\neg Bx\tau) \end{array}$$

□

Definition 9.4.2 (a) We call the part of 2D logic consisting of the sentences of the forms $(a\text{-}\ell)$, $(e\text{-}\ell)$, $(i\text{-}m)$ and $(o\text{-}m)$ the *ulem fragment* (for 'Universal ℓ and Existential m').

(b) More generally a *fragment* of 2D logic is a family of 2D sentences that can be considered as a logic in its own right. (For example a fragment would be expected to be closed under both contradictory and simple negation.) So the collection of 2D sentences of the forms (d) and (t) forms the *dt fragment* of 2D logic, and the collection of 2D sentences of the forms (d), (t) and (z) forms the *dtz fragment* of 2D logic.

In the sense of Definition 4.3.1 above, the $ulem$ sentences can be paraphrased into assertoric logic. The paraphrase treats an ordered pair consising of an object and a time as a single object.

Definition 9.4.3 Let Σ_{2d} be a two-dimensional signature, and let Σ_{1d} be the monadic relational signature containing exactly the same relation symbols as Σ_{2d}, but regarded as monadic. For each assertoric sentence ϕ of $L(\Sigma_{2d})$ we define a $ulem$ sentence ϕ^{ulem} of $L(\Sigma_{2d})$ by replacing quantifiers $\exists x$ by $\exists x\exists\tau$, and $\forall x$ by $\forall x\forall\tau$, and replacing each atomic formula Ax by $Ax\tau$.

Theorem 9.4.4 *(a) Up to logical equivalence, the mapping $\phi \mapsto \phi^{ulem}$ is a bijection between the assertoric sentences and the $ulem$ sentences.*

(b) Let T be a set of assertoric sentences of $L(\Sigma_{1d})$, and write T^{ulem} for the set $\{\phi^{ulem} : \phi \in T\}$ of ulem sentences of $L(\Sigma_{2d})$. Then T^{ulem} is consistent if and only if T is consistent.

(c) The mapping $\phi \mapsto \phi^{ulem}$ respects conversion and contradictory negation.

Proof. (a) and (c) can be checked directly.

(b) The direction $\Rightarrow$ comes from the interpretation of assertoric terms as talking about ordered pairs of object and time. Let the Σ_{2d}-structure M be a model of T^{ulem}. Form the Σ_{1d}-structure $M^{\downarrow}$ as follows. The domain of $M^{\downarrow}$ is the cartesian product of the object domain of M and the time domain of M, i.e. the set of ordered pairs ⟨object, time⟩. For each relation symbol A we put

(9.4.3) $$\langle a, \alpha\rangle \in A^{M^{\downarrow}} \;\;\Leftrightarrow\;\; (a, \alpha) \in A^{M}.$$

One can then see by inspection that if $M \models \phi^{ulem}$ then $M^{\downarrow} \models \phi$.

A different idea is needed for the direction $\Leftarrow$, since the elements of a Σ_{1d}-structure need not be ordered pairs. Suppose the Σ_{1d}-structure N is a model of T. We form a Σ_{2d}-structure $N^{\uparrow}$ as follows. The object domain of $N^{\uparrow}$ is the domain of N. Take an arbitrary object α; the time domain of $N^{\uparrow}$ is $\{\alpha\}$. For each relation symbol A, put

(9.4.4) $$(a, \alpha) \in A^{N^{\uparrow}} \;\;\Leftrightarrow\;\; a \in A^{N}.$$

Since α is the unique time in $N^{\uparrow}$, 'at all times' and 'at some time' both express 'at time α'. From this it's easy to see that if ϕ is an assertoric sentence in $L(\Sigma_{1d})$ and $N \models \phi$ then $N^{\uparrow} \models \phi^{ulem}$. □

History 9.4.5 The use of ordered pairs for packing two universal quantifiers into a single quantifier goes back at least to Alexander of Aphrodisias, and it forms an essential part of the background to C. S. Peirce's discovery of quantificational logic; see [43] on the history. By Ibn Sīnā's time it was very likely standard equipment; certainly he uses it regularly as a tool of logical analysis. For Ibn Sīnā's use of ordered pairs, see *Qiyās* 59.17, 256.14, 476.10, *Burhān* 146.14, *Jadal* 146.2, *Išārāt* 72.9, and further discussion in [47].

We turn next to the sentences of the form (z), which we can refer to collectively as the z fragment of 2D logic. We can paraphrase this fragment into assertoric logic by incorporating 'at time z', or more simply 'now', into the terms.

Definition 9.4.6 Let Σ_{2d} be a 2D signature and δ a constant of the sort $time$. Let Σ_{1d} be the monadic relational signature containing exactly the same relation symbols as Σ_{2d}, but regarded as monadic. For each assertoric sentence ϕ of $L(\Sigma_{1d})$ we define a sentence ϕ^{z} of $L(\Sigma_{2d}(\delta))$ by replacing each atomic formula Ax by $Ax\delta$.

Theorem 9.4.7 *(a) The mapping $\phi \mapsto \phi^z$ is a bijection between the assertoric sentences and the sentences of the z fragment of 2D logic.*

(b) Let T be a set of assertoric sentences of $L(\Sigma_{1d})$, and write T^z for the set $\{\phi^z : \phi \in T\}$ of sentences of the z fragment. Then T^z is consistent if and only if T is consistent.

(c) The mapping $\phi \mapsto \phi^z$ respects conversion and contradictory negation.

The *proof* is like that of Lemma 9.4.4 but simpler, and is left to the reader. □

The third case that we consider is the $atnd$ part of the dt fragment. The dt fragment is a part of 2D logic that has a major importance in its own right; so we will devote the next chapter to it, and meet the $atnd$ part along the way.

9.5 Exercises

9.1. With reference to Exercise 5.1 above:

(a) For which avicennan forms (g) are $(a\text{-}g)(B, A)$ and $(e\text{-}g)(A, B)$ not contraries?

(b) For which avicennan forms (g) are $(i\text{-}g)(B, A)$ and $(e\text{-}g)(A, B)$ not subcontraries?

Solution: For (a): (m) and (t). For (b): (d) and (ℓ).

9.2. Prove the entailments in Lemma 9.1.13 in the case where f is o.

Solution. Each of the sentences $(o\text{-}g)(B, A)$ is true if nothing is ever a B; so we can assume in proving the entailments that at least one thing is sometimes a B.

$(o\text{-}d)(B, A) \vdash (o\text{-}\ell)(B, A)$: Suppose some object a is sometimes a B but never an A at any time when it exists. By the theory of E, a exists at all times when it is a B. Therefore a is never an A at any time when it is a B.

$(o\text{-}\ell)(B, A) \vdash (o\text{-}m)(B, A)$: Suppose some object a is sometimes a B but never an A at any time when it is a B. Then a is at some time a B but not an A.

$(o\text{-}m)(B, A) \vdash (o\text{-}t)(B, A)$: Suppose some object a is at some time a B but not an A. At this time it exists, by the theory of E. So a is at some time a B, and at some time during its existence it is not an A.

9.3. Show that the premise-pair

$$(i\text{-}z)(B, C),\ (a\text{-}d)(B, A)$$

is productive, and find two distinct optimal conclusions for it. [This example was found by translating an example of Buridan into 2D logic by the Avicenna-Johnston semantics, cf. Section 12.3 below and Exercises 12.n–m.]

Solution. We read δ as 'now'. By the first premise there is an object a which is a B now and a C now. By the second premise, since a is sometimes a B, it is an A at all times when it exists. Therefore there is something that is a C now and an A at all times when it exists; and these times include now, by the theory of E. We can deduce $(i\text{-}z)(C, A)$ and $(i\text{-}d)(C, A)$, neither of which entails the other. The only 2D strengthening of $(i\text{-}z)$ is $(a\text{-}z)$, which is not obtainable by these premises. Likewise the only 2D strengthening of $(i\text{-}d)$ is $(a\text{-}d)$, also unobtainable.

9.4. The theory of E expresses a condition on $\Sigma_{2d}(E)$-structures M. State the corresponding condition on Kripke frames $\mathcal{K}(M)$.

Solution. A Kripke frame $(\mathcal{D}, \mathcal{W}, \mathcal{S}, \mathcal{E})$ is of the form $\mathcal{K}(M)$ for a $\Sigma_{2d}(E)$-structure M satisfying Th(E) if and only if (i) for each relation symbol A of $K(\Sigma_{2d})$ and each world $w \in \mathcal{W}$, $A^{\mathcal{S}_m} \subseteq \mathcal{E}_w$, and (ii) $\mathcal{D} \subseteq \bigcup_{w\in\mathcal{W}} \mathcal{E}_w$.

9.5. Let Σ be a monadic relational signature and ϕ, ψ formulas of the modal language $\mathcal{L}_{modal}(\Sigma)$. We say that ϕ and ψ are *logically equivalent* if for every Kripke Σ-frame K, every world w of K and every tuple $\bar{a}$ of elements, $K \models_w \phi[\bar{a}]$ if and only if $K \models_w \psi[\bar{a}]$. Show:

(a) Let ϕ be a formula of $\mathcal{L}_{modal}(\Sigma)$.

(b) For any two formulas ϕ, ψ of $\mathcal{L}_{modal}(\Sigma)$, $\forall x\Box\phi$ is logically equivalent to $\Box\forall x\phi$, and $\exists x\Box\phi$ is logically equivalent to $\Box\exists x\phi$.

(c) If ϕ and ψ are formulas of $\mathcal{L}_{modal}(\Sigma)$, ϕ is completely modalised and $\star$ is one of $\rightarrow$, $\wedge$, $\vee$, then $\Box(\phi \star \psi)$ is logically equivalent to $(\phi \star \Box\psi)$, and $\Diamond(\phi \star \psi)$ is logically equivalent to $(\phi \star \Diamond\psi$.

9.6. Prove or refute: For all sentences ϕ and ψ and formula $\theta(x)$:

(a) $\exists x \Box \theta(x)$ is logically equivalent to $\Box \exists x \theta(x)$.

(b) $(\Box \phi \wedge \Box \psi)$ is logically equivalent to $\Box(\phi \wedge \psi)$.

(c) $(\Box \phi \vee \Box \psi)$ is logically equivalent to $\Box(\phi \vee \psi)$.

Solution. (a) False. Take a Kripke frame with two objects a and b, and two worlds α and β, such that Ba is true at α and not at β, while Bb is true at β and not at α.

(b) True. $(\Box \phi \wedge \Box \psi)$ says that ϕ is true at all worlds and ψ is true at all worlds; which is equivalent to saying that at all worlds $(\phi \wedge \psi)$ is true, i.e. $\Box(\phi \wedge \psi)$.

(c) False. Let ϕ be $\exists x A x$ and ψ be $\exists x B x$. Take a Kripke frame with a single object a and two worlds α and β, such that at world α, Ba holds and Aa fails, and at world β, Aa holds and Ba fails. Then $\Box(\phi \vee \psi)$ holds at both worlds, while neither $\Box \phi$ nor $\Box \psi$ holds at either world.

9.7.

(a) Show that $\Box\Box\phi$ and $\Diamond\Box\phi$ are both logically equivalent to $\Box\phi$.

(b) Since the two-dimensional language and the modal language are notational variants of each other, (a) must be equivalent to some fact about the two-dimensional language. Determine what this fact is. Try (as hard as you like) to justify the claim that this fact tells us anything important to know about two-dimensional logic.

(c) At *Qiyās* ii.1, 86.6 Ibn Sīnā writes 'Every human necessarily can be a writer'. One can derive from (a) above that $\Box\Diamond\phi$ is logically equivalent to $\Diamond\phi$. Why would it be wrong to infer that Ibn Sīnā's sentence says no more than 'Every human being can be a writer'?

9.8.

(a) Find a model of $(a\text{-}z)(B, A)$ that is not a model of either $(i\text{-}\ell)(B, A)$ or $(a\text{-}t)(B, A)$.

(b) Find a model of $(a\text{-}d)(B, A)$ that is not a model of $(i\text{-}z)(B, A)$.

(Cf. Theorem 9.3.2 above.)

Solution. We take a 2D signature Σ whose relation symbols are just B and A, and we construct $\Sigma(E)$-structures to meet the requirements.

(a) The $\Sigma(E)$-structure M_1 has object domain $\{a, b\}$ and time domain $\{\alpha, \delta\}$. It satisfies: $Ba\alpha$, $\neg Aa\alpha$, $Ba\delta$, $Aa\delta$, $Bb\alpha$, $\neg Ab\alpha$, $\neg Bb\delta$, $Ab\delta$, E holds everywhere.

(b) The $\Sigma(E)$-structure M_2 has object domain $\{a\}$ and time domain $\{\alpha, \delta\}$. It satisfies: $Ba\alpha$, $Aa\alpha$, $Ea\alpha$, $\neg Ba\delta$, $\neg Aa\delta$, $\neg Ea\delta$.

9.9. I had this on conversions. It says that (z) sentences convert iff their assertoric projections do.

SOMETHING ON DE RE
SOMETHING ON CONVERSIONS

9.10. PROVE ECTHESIS ignoring universal augments (2) $(o\text{-}t)(B, A)$ $(Bx\tau \wedge \exists\sigma(Ex\sigma \wedge \neg Ax\sigma))$ $(a\text{-}\ell)(D, B), (e\text{-}t)(D, A)$

Solution. Assume $(o\text{-}t)$. Then there is a such that $\exists\tau Ba\tau$ and $\exists\tau(Ea\tau \wedge \neg Aa\tau)$, so $\exists x\exists\tau Dx\tau$. As in (2) we derive $(a\text{-}\ell)(D, B)$. If $\exists\sigma Dc\sigma$ then $\exists\sigma(Ec\sigma \wedge \neg Ac\sigma)$, so $(e\text{-}t)(D, A)$.

9.11. Let Σ be a 2D signature and M a $\Sigma(E, \delta)$-structure. Show that there is a $\Sigma(E, \delta)$-structure N such that

(a) for every relation symbol A in Σ, $N \models \exists x\exists\tau Ax\tau$, and

(b) for every CORE 2D sentence in $L(\Sigma(E))$, $M \models \phi$ if and only if $N \models \phi$.

9.12. Show that in there are models of $\{(a\text{-}d)(B, A), (e\text{-}z)(B, A)\}$ but that in any such model, no object satisfies B at time δ.

Solution. We first deal with truth at δ. Let $\mathbb{S}$ be the set of relation symbols in Σ such that $M \models \forall x\neg Ax\delta$. For each $A \in \mathbb{S}$ introduce a new element a_A, and let N be the $\Sigma(E, \delta)$-structure got by adding these new elements, putting $Aa_A\delta$ and $Ea\delta$ and nothing else. We check the requirements.

Sentences of the form $(a\text{-}z)(B, A)$ or $(i\text{-}z)$: No new element satisfying B at δ is added. So OK.

Sentences $(e\text{-}z)(B, A)$: If there already was an element satisfying B at δ then we add no new such element, so no change. If there wasn't, then we add an element satisfying B but not A at δ, which is OK. Similarly with $(o\text{-}z)(B, A)$.

What does this move do to core sentences? Suppose for example that there is an element satisfying B but none at δ, and every element satisfying B anywhere satisfies A somewhere.

Problem will be when we have $(a\text{-}d)(B, A)$ but no element satisfies B at δ. Suppose in fact $(e\text{-}z)(B, A)$. So these two sentences together entail that nothing satisfies B at δ, so the EXERCISE IS WRONG. MAKES NEW EXERCISE?

Chapter 10

The dt fragment

10.1 The dt reduction

We turn now to the convertible sentences of the form $(e\text{-}d)$. These can be reduced to convertible assertoric sentences by taking the time quantifiers into the predicates. This has the effect of moving the 'always' inside the scope of the negation:

(10.1.1) No sometime-B is a sometime-A.

So the paraphrase here should incorporate 'sometime' into the terms. This paraphrase will work on negative (d) sentences as above, and on affirmative (t) sentences.

Definition 10.1.1 (a) We define the *atnd fragment* of 2D logic, $\mathcal{L}_{atnd}$, to consist of the sentences that are either Affirmative with avicennan form (t), or Negative with avicennan form (d). This fragment is a subset of the *dt* fragment.

(b) For each 2D signature Σ_{2d} we define a corresponding monadic relational signature Σ^+_{1d}; the relation symbols of Σ^+_{1d} are the symbols A^+ for each relation symbol A of Σ_{2d}. For each assertoric sentence ϕ of $L(\Sigma^+_{1d}$ we define a 2D sentence ϕ^{atnd} of 2D logic, as follows:

(10.1.2)

ϕ	ϕ^{atnd}
$(a)(B^+, A^+)$	$(a\text{-}t)(B, A)$
$(e)(B^+, A^+)$	$(e\text{-}d)(B, A)$
$(i)(B^+, A^+)$	$(i\text{-}t)(B, A)$
$(o)(B^+, A^+)$	$(o\text{-}d)(B, A)$

We show that these maps constitute a paraphrase of $\mathcal{L}_{atnd}$ into assertoric logic, assuming the theory of E in $\mathcal{L}_{atnd}$.

Theorem 10.1.2 *(a) The mapping $\phi \mapsto \phi^{atnd}$ is a bijection between the class of assertoric sentences and the class of 2D sentences in the atnd fragment.*

(b) Let T be an assertoric theory and let T' be the theory $\{\phi^{atnd} : \phi \in T\}$. Then $T' \cup Th(E)$ is consistent if and only if T is consistent.

(c) The mapping $\phi \mapsto \phi^{atnd}$ respects contradictory negation and conversion.

Proof. (a) and (c) are left to the reader.

The easy direction of (b) is ⇒, and it consists of using the interpretation of terms of assertoric logic so as to define an operation taking models N of $T' \cup \text{Th}(E)$ to models $N^{\downarrow}$ of T. The domain of $N^{\downarrow}$ will be the object domain of N. For each relation symbol A of Σ_{2d} we put

$$(A^{+})^{N^{\downarrow}} = \{a \in \text{dom}(N^{\downarrow}) : N \models \exists\tau(Ea\tau \wedge Aa\tau)\}. \tag{10.1.3}$$

We show for example that if $N \models (o\text{-}d)(B, A)$ and is a model of Th(E) then $N^{\downarrow} \models (o)(B^{+}, A^{+})$. Since $(o\text{-}d)$ sentences are disjunctions there are two cases to consider. The first is that for some object a,

$$N \models (\exists\tau Ba\tau \wedge \forall\tau(Ea\tau \rightarrow \neg Aa\tau)). \tag{10.1.4}$$

Then $N \models \exists\tau(Ea\tau \wedge Ba\tau)$ by the theory of D, and so $a \in (B^{+})^{N^{\downarrow}}$ by (10.1.3); also by (10.1.3), $a \notin (A^{+})^{N^{\downarrow}}$. It follows that

$$N^{\downarrow} \models (B^{+}a \wedge \neg A^{+}a). \tag{10.1.5}$$

The second case is that B^{N} is empty. In this case $(B^{+})^{N^{\downarrow}}$ is empty too, by (10.1.3) again. The other sentence forms are left to the reader.

In the direction ⇐ we need to choose an interpretation for E, and we do it in the simplest possible way. We take the object domain of $M^{\uparrow}$ to be the domain of M, and the time domain of $M^{\uparrow}$ to be the set $\{0\}$. For every relation symbol A^{+} of Σ^{+}_{1d} we define

$$A^{M^{\uparrow}} = (A^{+})^{M} \times \{0\} \tag{10.1.6}$$

and likewise we put

$$E^{M^{\uparrow}} = \text{dom}(M) \times \{0\}. \tag{10.1.7}$$

We show that if $M \models (e)(B^+, A^+)$ then $M^\uparrow \models (e\text{-}d)(B, A)$, as follows.

$$\begin{array}{lll} & M \models (e)(B^+, A^+) & \\ \Rightarrow & (B^+)^M \cap (A^+)^M = \emptyset & \\ \Rightarrow & B^{M^\uparrow} \cap A^{M^\uparrow} = \emptyset & \text{by (10.1.6)} \\ \Rightarrow & M^\uparrow \models \forall x(\exists \tau Bx\tau \to \forall \tau \neg Ax\tau) & \text{since only one time} \\ \Rightarrow & M^\uparrow \models \forall x(\exists \tau Bx\tau \to \forall \tau (Ex\tau \to \neg Ax\tau)) & \\ \Rightarrow & M^\uparrow \models (e\text{-}d)(B, A). & \end{array}$$

The other cases are equally straightforward. Also(10.1.7) makes $M^\uparrow$ a model of Th(E). □

Theorem 10.1.2 shows that the $atnd$ fragment is, in terms of inference and conversions, just an isomorphic copy of assertoric logic. In particular those parts of Avicenna's assertoric proof theory that depend only on inference and conversions will transfer straightforwardly to this fragment. For safety we will check in due course that the ecthetic proofs transfer too.

But surely we can do better than this. If 'sometimes an A' can be made into a term, then so can 'always an A', and then we get a translation of the whole of the dt fragment into assertoric logic. Does this work? Not without some extra tweaking. Nevertheless this avenue seems to give us by far the easiest access to the logical properties of the dt fragment. The rest of this Chapter will explain how.

Definition 10.1.3 Let Σ_{2d} be a two-dimensional signature, and let $\Sigma^{\pm}_{1d}$ be the monadic relational signature whose relation symbols are got by taking, for each relation symbol A of Σ_{2d}, a pair of new symbols A^+ and A^-. The $+$ and $-$ are called the *polarities* of these new relation symbols. For each assertoric sentence ϕ of $L(\Sigma^{\pm}_{1d})$ whose subject symbol has polarity $+$, we define a 2D sentence ϕ^{dt} of 2D logic, as follows:

(10.1.8)

ϕ	ϕ^{dt}
$(a)(B^+, A^+)$	$(a\text{-}t)(B, A)$
$(a)(B^+, A^-)$	$(a\text{-}d)(B, A)$
$(e)(B^+, A^+)$	$(e\text{-}d)(B, A)$
$(e)(B^+, A^-)$	$(e\text{-}t)(B, A)$
$(i)(B^+, A^+)$	$(i\text{-}t)(B, A)$
$(i)(B^+, A^-)$	$(i\text{-}d)(B, A)$
$(o)(B^+, A^+)$	$(o\text{-}d)(B, A)$
$(o)(B^+, A^-)$	$(o\text{-}t)(B, A)$

When the predicate symbol of ϕ has polarity $+$, ϕ^{dt} as defined above is identical with ϕ^{atnd} as defined in (10.1.2).

It would be foolish to hope that all the logical relationships between sentences in the dt fragment could be deduced from the logical properties of their assertoric counterparts. For example $(a\text{-}d)(B, A)$ entails $(a\text{-}t)(B, A)$ (modulo the theory of E), but nothing has been said to connect the A^- in $(a)(B^+, A^-)$ with the A^+ in $(a)(B^+, A^+)$. We should at least feed in the information that

(10.1.9) for every relation symbol A in $\Sigma^{\pm}_{1d}$, $\forall x(A^-x \to A^+x)$.

Agreeably, the new sentences introduced in (10.1.9) are the unaugmented assertoric sentences $(a, uas)(A^-, A^+)$ (Definition 3.1.5), so it shouldn't be too hard to make them go to work alongside the assertoric sentences.

Definition 10.1.4 We define $\mathcal{L}_{as\pm}$ to be the logic whose relation symbols are the monadic relation symbols of the form A^+ or A^- (i.e. the relation symbols in the signatures $\Sigma^{\pm}_{1d}$ just defined), and whose sentences are the assertoric sentences with subject symbol of polarity $+$, together with the unaugmented assertoric sentences of the form $\forall x(A^-x \to A^+x)$.

We show that the mapping $\phi \mapsto \phi^{dt}$ as defined in Definition 10.1.3 gives a paraphrase of the dt fragment in the logic $\mathcal{L}_{as\pm}$.

Theorem 10.1.5 *(a) The mapping $\phi \mapsto \phi^{dt}$ is a bijection between the set of assertoric sentences of $\Sigma^{\pm}_{1d}$ with subject symbols of polarity $+$, and the set of 2D sentences in the dt fragment in $\Sigma_{2d}(E)$.*

(b) Suppose T is a set of assertoric sentences of $\Sigma^{\pm}_{1d}$ with subject symbol of polarity $+$, and T' is the theory $\{\phi^{dt} : \phi \in T$. Then $T' \cup Th(E)$ is consistent if and only if $T \cup Th(\pm)$ is consistent.

(c) The mapping $\phi \mapsto \phi^{dt}$ respects contradictory negation and conversion.

Proof. (a) and (c) are left to the reader.

For (b), again the direction $\Rightarrow$ is the easy and conceptual direction. Suppose N is a model of $T' \cup \text{Th}(E)$. We use (10.1.3) as before, together with:

(10.1.10) $$(A^-)^{N\downarrow} = \{a \in \text{dom}(N^{\downarrow}) : N \models \forall\tau(Ea\tau \to Aa\tau)\}.$$

Given the theory of E, $\forall\tau(Ea\tau \to Aa\tau)$ implies $\exists\tau Ea\tau)$, so $N^{\downarrow}$ is a model of Th($\pm$).

$\Leftarrow$ Suppose M is a model of $T \cup \mathrm{Th}(\pm)$. We define the $\Sigma_{2d}(E)$-structure $M^{\uparrow}$ as follows. We introduce two new objects, say 0 and 1. We take the object domain of $M^{\uparrow}$ to be the domain of M, and the time domain of $M^{\uparrow}$ to be the set $\{0, 1\}$. We define

(10.1.11) $$E^{M^{\uparrow}} = \mathrm{dom}(E) \times \{0, 1\}$$

and for every relation symbol A of Σ_{2d} we put

(10.1.12) $$A^{M^{\uparrow}} = ((A^{+})^{M} \times \{0\}) \cup ((A^{-})^{M} \times \{1\}).$$

Since M is a model of $\mathrm{Th}(\pm)$, $(A^{-})^{M} \subseteq (A^{+})^{M}$. From this we calculate that for any element a of M,

(10.1.13) $$\begin{array}{lcl} M^{\uparrow} \models \exists\tau Ax\tau & \Leftrightarrow & M \models A^{+}a, \\ M^{\uparrow} \models \forall\tau Ax\tau & \Leftrightarrow & M \models A^{-}a, . \end{array}$$

Also by the definition of $E^{M^{\uparrow}}$, $M^{\uparrow}$ satisfies the theory of E. □

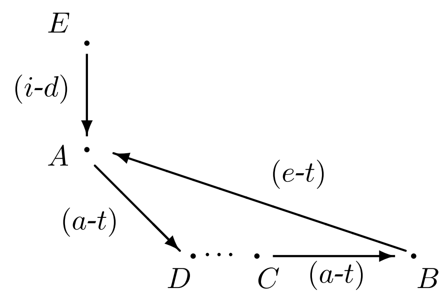
10.2 Classification of optimal MITs

Theorem 10.1.5 makes it possible to calculate a listing of the optimal inconsistent sets of dt sentences by using the corresponding results in the assertoric logic $\mathcal{L}_{as\pm}$. We will carry through this calculation in Section 10.6 below. Meanwhile we quote the results of Section 10.6 and analyse what they tell us about the dt fragment.

As in Section 5.3 above, we describe theories in terms of their subject-predicate digraphs. The optimal MITs in the dt fragment fall into four disjoint families. The families $(i)II\uparrow$, $(i)II\downarrow$ and (o) correspond to the families $(i)\uparrow$, $(i)\downarrow$ and (o) of Section 5.3, with the difference that for most families, the members of the family are now distinguished not just by the lengths of the tracks, but also by the avicennan forms. There is a new family $(i)I$ whose digraphs are not circular; in fact they consist of a circle with a short tail added at top left.

Type $(i)I$.

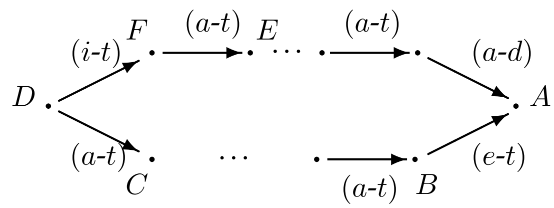
(10.2.1)



An MIT has type $(i)I(n)$ if it has a subject-predicate digraph as shown, where the upper track from D to A has length 0 and the lower track from D to A has length n. The parameter n can have any value $\geqslant 2$.

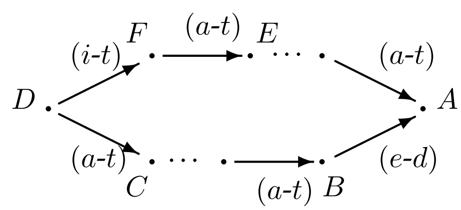
Type $(i)II\!\uparrow$

(10.2.2)



or

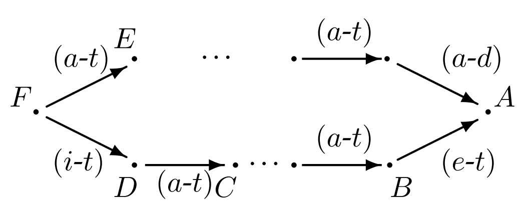
(10.2.3)



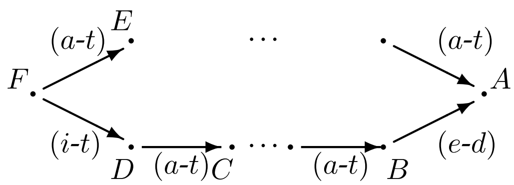
An MIT has type $(i)\!\uparrow(m,n)$ if it has a subject-predicate digraph as shown, where the upper track from D to A has length m and the lower track from D to A has length n. These two parameters can each have any value $\geqslant 1$.

Type $(i)II\!\downarrow$

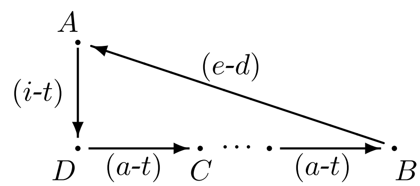
(10.2.4)



(10.2.5)



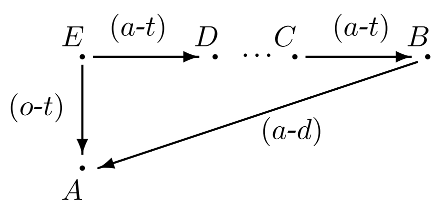
(10.2.6)



An MIT has type $(i)\!\downarrow(m,n)$ if it has a subject-predicate digraph as shown, where the upper track from D to A has length m and the lower track from D to A has length n. Here m can be any number $\geqslant 0$ and n can be any number $\geqslant 1$, subject to the requirement that $m+n \geqslant 2$. MITs of this type are always optimal inconsistent.

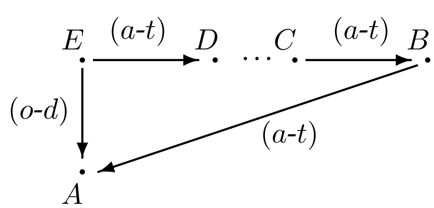
Type (o)

(10.2.7)



and

(10.2.8)



An MIT has type $(o)(m)$ if it has a subject-predicate digraph as shown, where the uppper (i.e. righthand) track from E to A has length m. Here m can be any number $\geqslant 1$. MITs of this type are always optimal inconsistent.

Theorem 10.2.1 *A set of* dt *sentences is optimal inconsistent if and only if it has a subject-predicate digraph of one of the types* $(i)I$, $(i)II\uparrow$, $(i)II\downarrow$ *or* (o) *as above.*

Proof. This is Theorem 10.6.7 of Section 10.6 below. □

Corollary 10.2.2 *Let* T *be an optimal inconsistent set of* dt *sentences. Then:*

(a) T *contains exactly one negative sentence.*

(b) T *contains exactly one existential sentence.*

(c) T *contains exactly one sentence of avicennan form* (d).

We remark that (a) and (c) remain true for the whole of 2D logic, but (b) fails already for the dtz fragment; see Example BELOW.

Corollary 10.2.3 *A graph-circular set* T *of* dt *sentences is optimal inconsistent if and only if it satisfies the following two conditions:*

(a) The assertoric projection (see Definition 9.1.4) of T *is an optimal inconsistent assertoric theory.*

(b) T *has exactly one sentence of avicennan form* d*, and this sentence is the final sentence in one track of the digraph of* T.

Corollary 10.2.4 (Orthogonality Principle for the dt **fragment)** *Let* $T[B, A]$ *be a non-retrograde premise-sequence in the* dt *fragment. Then the conditions for* $T[B, A]$ *to be productive, with conclusion of a given* dt *form, are the conjunction of*

(a) a condition that the assertoric projections of the premises form a productive assertoric premise-pair, and

(b) a condition depending only on the figure of $T[B, A]$ *and the avicennan forms of the sentences.*

The Orthogonality Principle fails for retrograde premise-sequences because with these the figure is not sufficient to identify the final sentence in the antilogism (see Exercise 3.6). But the good news is that the Principle holds not only for the dt fragment but for the whole of core 2D logic. However, we will see in Section 12.3 that it fails massively for the dtz fragment; this is probably the most visible difference between the logic of Ibn Sīnā and the logic of Buridan.

10.3 Productive two-premise moods

We can now lift the results of Section 5.4 (which are classical) to the dt fragment (which first appeared with Ibn Sīnā). Recall from Section 4.1 that the Genetic Principle requires Ibn Sīnā to look for conclusions in the dt fragment. The results in Section 5.4 only tell us about conclusions in the dt fragment anyway. In fact there are a few cases where stronger conclusions can be found elsewhere in core 2D logic; we note these at the end of the section.

Definition 10.3.1 We name a two-premise dt mood by naming its assertoric projection and then adding '(h, j, k)' where h, j and k are letters for the avicennan moods of the minor premise, the major premise and the conclusion respectively.

Before listing the productive moods, we need to remember that Ibn Sīnā's practice is to list *all* productive moods, even those whose premises are unnecessarily strong for the conclusions that they yield. So we need to run through the possible avicennan modalities on the premises, and check from Section 10.2 which of them are productive. For each of the first three figures the results are independent of the mood, by the Orthogonality Principle, Corollary 10.2.4 above. The results are as follows.

First Figure:

(10.3.1)

premise-pair	productive	strongest conclusion
(d, d)	Yes	d
(d, t)	Yes	t
(t, d)	Yes	d
(t, t)	Yes	t

Second Figure:

(10.3.2)

premise-pair	productive	strongest conclusion
(d, d)	Yes	d
(d, t)	Yes	d
(t, d)	Yes	d
(t, t)	No	

Third Figure:

(10.3.3)

premise-pair	productive	strongest conclusion
(d, d)	Yes	d
(d, t)	Yes	t
(t, d)	Yes	d
(t, t)	Yes	t

The digraphs of the antilogisms are the same as in Section 5.4 but with the avicennan moods added. Since we already have these digraphs listed, I now revert to the ordering in figures.

First Figure

mood	minor	major	conclusion	notes
1.1 *Barbara(t,d,d))*	$(a\text{-}t)(C, B)$	$(a\text{-}d)(B, A)$	$(a\text{-}d)(C, A)$	
1.1 *Barbara(t,t,t)*	$(a\text{-}t)(C, B)$	$(a\text{-}t)(B, A)$	$(a\text{-}t)(C, A)$	*atnd*
1.2 *Celarent(t,d,d)*	$(a\text{-}t)(C, B)$	$(e\text{-}d)(B, A)$	$(e\text{-}d)(C, A)$	*atnd*
1.2 *Celarent(t,t,t)*	$(a\text{-}t)(C, B)$	$(e\text{-}t)(B, A)$	$(e\text{-}t)(C, A)$	
1.3 *Darii(t,d,d)*	$(i\text{-}t)(C, B)$	$(a\text{-}d)(B, A)$	$(i\text{-}d)(C, A)$	
1.3 *Darii(t,t,t)*	$(i\text{-}t)(C, B)$	$(a\text{-}t)(B, A)$	$(i\text{-}t)(C, A)$	*atnd*
1.4 *Ferio(t,d,d)*	$(i\text{-}t)(C, B)$	$(e\text{-}d)(B, A)$	$(o\text{-}d)(C, A)$	*atnd*
1.4 *Ferio(t,t,t)*	$(i\text{-}t)(C, B)$	$(e\text{-}t)(B, A)$	$(o\text{-}t)(C, A)$	

Besides these optimal moods there are the corresponding non-optimal moods with (d, d, d) and (t, d, t), making a total of 16 productive moods.

Second Figure

mood	minor	major	conclusion	notes
2.1 *Cesare(d,t,d)*	$(a\text{-}d)(C, A)$	$(e\text{-}t)(B, A)$	$(i\text{-}d)(C, B)$	
2.1 *Cesare(t,d,d)*	$(a\text{-}t)(C, A)$	$(e\text{-}d)(B, A)$	$(i\text{-}d)(C, B)$	*atnd*
2.2 *Camestres(d,t,d)*	$(e\text{-}d)(C, A)$	$(a\text{-}t)(B, A)$	$(e\text{-}d)(C, B)$	*atnd*
2.2 *Camestres(t,d,d)*	$(e\text{-}t)(C, A)$	$(a\text{-}d)(B, A)$	$(e\text{-}d)(C, B)$	
2.3 *Festino(d,t,d)*	$(i\text{-}d)(C, A)$	$(e\text{-}t)(B, A)$	$(o\text{-}d)(C, B)$	
2.3 *Festino(t,d,d)*	$(i\text{-}t)(C, A)$	$(e\text{-}d)(B, A)$	$(o\text{-}d)(C, B)$	*atnd*
2.4 *Baroco(d,t,d)*	$(o\text{-}d)(C, A)$	$(a\text{-}t)(B, A)$	$(o\text{-}d)(C, B)$	*atnd*
2.4 *Baroco(t,d,d)*	$(o\text{-}t)(C, A)$	$(a\text{-}d)(B, A)$	$(o\text{-}d)(C, B)$	

Besides these optimal moods there are the corresponding non-optimal moods with (d, d, d), making a total of 12 productive moods.

History 10.3.2 These charts describe the logical facts, not what Ibn Sīnā says about the logical facts. Nevertheless when Ibn Sīnā claims that there is no productive second figure mood with two 'broad absolute' premises, it's bizarre not to think that he is referring to the absence of any mood with avicennan forms $(t, t, -)$ in the list above. Less clear is what he means when he suggests that this marks a major difference between his logic and that of his predecessors. He refers to views of Theophrastus and Themistius discussing broad absolute sentences; unfortunately the relevant texts haven't survived. Did he find, in these authors or in Aristotle or Alexander, an example claimed to be a valid second figure syllogism, but with premises that he read as being broad absolute? Or was his point just that these earlier authors, when handling sentences that in his view should be read as broad absolute, failed to point out that such sentences never generate a second figure syllogism? ADD REFS.

Third Figure

mood	minor	major	conclusion	notes
3.1 *Darapti(t,d,d)*	$(a\text{-}t)(B, C)$	$(a\text{-}d)(B, A)$	$(i\text{-}d)(C, A)$	
3.1 *Darapti(t,t,t)*	$(a\text{-}t)(B, C)$	$(a\text{-}t)(B, A)$	$(i\text{-}t)(C, A)$	*atnd*
3.2 *Felapton(t,d,d)*	$(a\text{-}t)(B, C)$	$(e\text{-}d)(B, A)$	$(o\text{-}d)(C, A)$	*atnd*
3.2 *Felapton(t,t,t)*	$(a\text{-}t)(B, C)$	$(e\text{-}t)(B, A)$	$(o\text{-}t)(C, A)$	
3.3 *Datisi(t,d,d)*	$(i\text{-}t)(C, B)$	$(a\text{-}d)(C, A)$	$(i\text{-}d)(B, A)$	
3.3 *Datisi(t,t,t)*	$(i\text{-}t)(C, B)$	$(a\text{-}t)(C, A)$	$(i\text{-}t)(B, A)$	*atnd*
3.4 *Disamis(t,d,d)*	$(a\text{-}t)(C, B)$	$(i\text{-}d)(C, A)$	$(i\text{-}d)(B, A)$	
3.4 *Disamis(t,t,t)*	$(a\text{-}t)(C, B)$	$(i\text{-}t)(C, A)$	$(i\text{-}t)(B, A)$	*atnd*
3.5 *Bocardo(t,d,d)*	$(a\text{-}t)(C, A)$	$(o\text{-}d)(C, B)$	$(o\text{-}d)(B, A)$	*atnd*
3.5 *Bocardo(t,t,t)*	$(a\text{-}t)(C, A)$	$(o\text{-}t)(C, B)$	$(o\text{-}t)(B, A)$	
3.6 *Ferison(t,d,d)*	$(i\text{-}t)(C, B)$	$(e\text{-}d)(C, A)$	$(o\text{-}d)(B, A)$	*atnd*
3.6 *Ferison(t,t,t)*	$(i\text{-}t)(C, B)$	$(e\text{-}t)(C, A)$	$(o\text{-}t)(B, A)$	

Besides these avicennan-optimal moods there are the corresponding non-optimal moods with (d, d, d) and (d, t, t), making a total of 24 productive moods.

Fourth Figure

mood	minor	major	conclusion	notes
4.1 *Bamalip*(t, t, t)	$(a\text{-}t)(B, C)$	$(a\text{-}t)(A, B)$	$(i\text{-}t)(C, A)$	*atnd*
4.2 *Dimatis*(t, t, t)	$(a\text{-}t)(C, B)$	$(i\text{-}t)(A, C)$	$(i\text{-}t)(B, A)$	*atnd*
4.3 *Calemes*(d, t, d)	$(e\text{-}d)(B, A)$	$(a\text{-}t)(C, B)$	$(e\text{-}d)(A, C)$	*atnd*
4.4 *Fesapo*(t, d, d)	$(a\text{-}t)(A, B)$	$(e\text{-}d)(C, A)$	$(o\text{-}d)(B, C)$	*atnd*
4.5 *Fresison*(t, d, d)	$(i\text{-}t)(A, C)$	$(e\text{-}d)(B, A)$	$(o\text{-}d)(C, B)$	*atnd*

In Fourth Figure the non-optimal moods need to be listed separately for each assertoric mood here, because the Orthogonality Principle fails here. *Bamalip* allows any of the four assignments of premises, but only a (t) conclusion; *Dimatis* likewise. *Calemes* allows two premise assignments: (d, t) and (d, d), and gives a (d) conclusion both ways. *Fesapo* allows the two premise assignments (t, d) and (d, d), and gives a d conclusion for both; likewise *Fresison*. So the total number of productive moods is 14.

Example 10.3.3 We noted at the beginning of this section that in some cases there are conclusions that are optimal in the dt fragment but not optimal in core 2D logic. These cases can be read off from the charts in Section 11.1 below. There are just two such cases:

mood	minor	major	conclusion
Datisi(d,t,m)	$(i\text{-}d)(B, C)$	$(a\text{-}t)(B, A)$	$(i\text{-}m)(C, A)$
Dimatis(t,d,m)	$(a\text{-}d)(B, C)$	$(i\text{-}t)(A, B)$	$(i\text{-}m)(C, A)$

The premises of *Datisi(d,t,m)* say that some sometime-B is always a C, and every sometime-B is sometimes an A. Clearly it follows that some sometime-A is always a C, i.e. $(i\text{-}d)(A, C)$. But this would be a conclusion in *Disamis* with the premises swapped around; it is not a conclusion in *Datisi* because the terms are in the wrong order in the conclusion. The best conclusion we can get in *Datisi* is to weaken $(i\text{-}d)(A, C)$ to $(i\text{-}m)(A, C)$ and then convert to $(i\text{-}m)(C, A)$ by $(i\text{-}m)$-conversion as in Theorem 9.3.4. The *Dimatis* case is similar.

10.4 Avicennan proof theory for the dt fragment

How does the proof theory of Section 8.3 above adapt to the dt fragment?

Our question here is not what adaptations Ibn Sīnā made; it is what adaptions are in fact needed and what are possible. Our answers will sometimes use logical facts that Ibn Sīnā discovered and pointed out, but the question how he viewed the situation is a separate matter and one of the central topics of [45]. I show there that it provides an essential clue for understanding the *ḳabṭ* that we met already in Section 1.1 above. (Section 12.1 below will give a brief preview.)

(1) Avicennan weakenings

The dt moods include a number of non-optimal moods where the avicennan form of a premise can be weakened without weakening the conclusion. In the assertoric case, Section 8.3 handled weakenings by use of conversions, following Aristotle. But in 2D logic this approach is no longer feasible. For example to weaken $(a\text{-}d)(B, A)$ to $(a\text{-}t)(B, A)$ by conversion we would need to be able to convert a sentence with assertoric form (a) to another sentence with assertoric form (a), which is clearly impossible. I will assume for the present that avicennan weakenings can be regarded as perfect inferences, so that a premise can have its avicennan form weakened at the point where the premise is used. (This is not a claim about Ibn Sīnā's practice; that needs to be checked separately.)

(2) First-figure moods

We will also assume for the present that the first figure dt moods can be taken as perfect, just as the assertoric ones were. Again this will need checking against Ibn Sīnā's account. He might disagree about whether the dt first-figure moods are cognitively self-evident; this is not really a question of logic. For our present purposes it suffices to fix a set of axioms in the style of Section 8.3.

The two rules added in Definition 8.3.4 to form the rule-book $\mathbb{R}3^{=}$ are simply first-figure moods with the premises reversed, so for the present we can assume also that these rules carry over to the dt situation.

Having settled (1) and (2), for the rest of this section we can restrict attention to the avicennan-optimal moods in Figures Two to Four, as listed in the charts of the previous section.

(3) Moods in the $atnd$ fragment

For these moods the proofs given in Section 8.3 carry over exactly, modulo a tacit use of the theories of E and $\pm$. (In fact for the avicennan-optimal moods in the $atnd$ fragment the theory of $\pm$ is never needed; EXERCISE?) All the productive fourth-figure moods in the dt fragment fall under this case.

(4) Conversions that lift

We can read off from Theorem 9.3.4 that in the dt fragment, symmetric (e)-conversion works only for the avicennan form (d), and symmetric (i)-conversion works only for the avicennan form (t); $(a\text{-}d)$ and $(a\text{-}t)$ both convert only to $(i\text{-}t)$ in this fragment.

The following moods lie in the dt fragment and outside the $atnd$ fragment, but the conversions used for them in Section 8.3 are still valid for the given avicennan forms:

(10.4.1) *Darapti(t,d,d), Felapton(t,t,t), Datisi(t,d,d), Ferison(t,t,t)*

In Section 8.3 these were proved by conversions. For the first two of them, the (a) minor premise was converted to (i), and in the third and fourth the (i) minor premise was converted to (i). In the dt versions the minor premise has the form either $(a\text{-}t)$ or $(i\text{-}t)$, and these forms do convert as required.

Thus far, the proof methods of Section 8.3 lift to the dt fragment unproblematically. There remain the following six moods:

(10.4.2) *Cesare(d,t,d), Camestres(t,d,d), Festino(d,t,d), Baroco(t,d,d), Disamis(t,d,d), Bocardo(t,t,t).*

All six are in Figures Two and Three, hence of interest to Ibn Sīnā. The methods used in Section 8.3 for *Cesare(d,t,d)*, *Camestres(t,d,d)* and *Festino(d,t,d)* would require symmetrical $(e\text{-}t)$, which is not available. The methods used for *Disamis(t,d,d)* would require symmetrical $(i\text{-}d)$-conversion, which is not available.

(5) Ectheses

In Section 8.3 we used ecthesis to justify two moods: *Baroco* and *Bocardo*. The proof that we gave for *Baroco* seems unlikely to lift to *Baroco(t,d,d)*, since taking the obvious route would require us to convert $(e\text{-}t)((C \setminus B), B)$. But in any case Ibn Sīnā's ecthesis always introduces a new term by definition, and we haven't checked how this procedure adapts to the dt fragment. So it would be safest to reconsider all the valid cases of *Baroco* and *Bocardo* in the dt fragment. There are four cases:

(i) *Baroco(d,t,d)* requires expanding $(o\text{-}d)(C, B)$.

(ii) *Baroco(t,d,d)* requires expanding $(o\text{-}t)(C, B)$.

(iii) *Bocardo(t,d,d)* requires expanding $(o\text{-}d)(B, A)$.

(iv) *Bocardo(t,t,t)* requires expanding $(o\text{-}t)(B, A)$.

In the case of *Baroco(t,d,d)* we can use the ecthesis (1) of Theorem 9.3.5, which gives us the move

(10.4.3) $$(o\text{-}t)(C,B) \vdash (a\text{-}t)(D,C), (e\text{-}\ell)(B,D)$$

This rescue takes us outside the dt fragment and into logical territory that we will explore in later chapters. For the present we quote a First Figure mood from Section 11.2 below:

(10.4.4) $$(a\text{-}d)(C,B), (e\text{-}\ell)(B,A) \ \rhd (e\text{-}d)(C,A).$$

This does the required job as follows:

1.	$<(o\text{-}t)(C,B), (a\text{-}d)(A,B)> [C,A]$	posit
2.	$<(a\text{-}t)(D,C), (e\text{-}\ell)(B,D), (a\text{-}d)(A,B)> [C,A]$	by ecthesis, Theorem 9.3.5 (1)
3.	$<(a\text{-}t)(D,C), (e\text{-}d)(A,D)> [C,A]$	by 5 in Def 8.3.4 and (10.4.3)
4.	$<(i\text{-}t)(C,D), (e\text{-}d)(A,D)> [C,A]$	by $(a\text{-}t)$-conversion
5.	$<(i\text{-}t)(C,D), (e\text{-}d)(D,A)> [C,A]$	by $(e\text{-}d)$-conversion
6.	$<(o\text{-}d)(C,D)> [C,A]$	by *Ferio(t,d,d)*

NOTES: For *Baroco(d,t,d)* we need:

1.	$<(o\text{-}d)(C,B), (a\text{-}t)(A,B)> [C,A]$	posit
2.	$<(a\text{-}t)(D,C), (e\text{-}d)(B,D), (a\text{-}t)(A,B)> [C,A]$	by ecthesis, Theorem 9.3.5 (1)
3.	$<(a\text{-}t)(D,C), (e\text{-}d)(A,D)> [C,A]$	by 5 in Def 8.3.4 and (10.4.3)
4.	$<(i\text{-}t)(C,D), (e\text{-}d)(A,D)> [C,A]$	by $(a\text{-}t)$-conversion
5.	$<(i\text{-}t)(C,D), (e\text{-}d)(D,A)> [C,A]$	by $(e\text{-}d)$-conversion
6.	$<(o\text{-}d)(C,D)> [C,A]$	by *Ferio(t,d,d)*

For *Bocardo(t,d,d)* we need:

1.	$<(a\text{-}t)(B,C), (o\text{-}d)(B,A)> [C,A]$	posit
2.	$<(a\text{-}t)(B,C), (a\text{-}\ell)(D,B), (e\text{-}d)(A,D)> [C,A]$	Theorem 9.3.5 (2)
3.	$<(a\text{-}t)(D,C), (e\text{-}d)(D,A)> [C,A]$	by *Barbara* and a SWITCH
4.	$<(i\text{-}t)(C,D), (e\text{-}d)(D,A)> [C,A]$	by $(a\text{-}t)$-conversion
5.	$<(o\text{-}d)(C,A)> [C,A]$	by *Ferio*

For *Bocardo(t,t,t)* we need:

1.	$<(a\text{-}t)(B,C), (o\text{-}t)(B,A)> [C,A]$	posit
2.	$<(a\text{-}t)(B,C), (a\text{-}t)(D,B), (e\text{-}\ell)(D,A)> [C,A]$	Theorem 9.3.5 (1)
3.	$<(a\text{-}t)(D,C), (e)(D,A)> [C,A]$	by SWITCHED *Barbara*
4.	$<(i\text{-}t)(C,D), (e)(D,A)> [C,A]$	by $(a\text{-}t)$-conversion
5.	$<(o\text{-}t)(C,A)> [C,A]$	by *Ferio*

Since we now have several options for choosing the defined term in ecthesis, it makes sense to check whether any of the dt moods that we haven't yet justified could be justified by a suitable choice of ecthesis. As it happens, Ibn Sīnā had exactly this thought; he showed that it takes care of *Disamis(t,d,d)*.

1.	$<(a\text{-}t)(B,C),(i\text{-}d)(B,A)> [C,A]$	posit
2.	$<(a\text{-}t)(B,C),(a\text{-}t)(D,B),(a\text{-}d)(D,A)> [C,A]$	by ecthesis, Fact 9.3.5 (3)
3.	$<(a\text{-}t)(D,C),(a\text{-}d)(D,A)> [C,A]$	by reversed *Barbara*
4.	$<(i\text{-}t)(C,D),(a\text{-}d)(D,A)> [C,A]$	by $(i\text{-}t)$-conversion
5.	$<(i\text{-}d)(C,A)> [C,A]$	by *Darii*

Ibn Sīnā gives the proof without spelling out the avicennan forms. As in the cases discussed just above, he specifies D with the inadequate description 'some B that is an A', but he adds at once that this should be adjusted so as to prove a conclusion with the same modality as the second premise. In fact the definition

(10.4.5) $$Dx\tau \;\equiv\; (Bx\tau \wedge \forall\sigma(Ex\sigma \rightarrow Ax\sigma))$$

works here by giving the ecthesis

(10.4.6) $$(i\text{-}d)(B,A) \;\vdash\; (a\text{-}t)(D,B),(a\text{-}d)(D,A)$$

which is a weakening of Fact 9.3.5 (5).

(6) *Cesare(d,t,d), Camestres(t,d,d), Festino(d,t,d)*

In Section 8.3 these were proved by converting the (e) premise. This premise now has the form $(e\text{-}t)$, which doesn't convert. Converting the (a) premise to (i) in *Cesare* and *Camestres* is a hopeless move, because it can give only an existential conclusion. If Ibn Sīnā wants to find direct proofs of these conclusions from these premises, he will have to go outside Aristotelian methods. The first two have no existential premises, so at least for these there is no obvious route by ecthesis.

Probably my own preference here is clear. I would go for reduction to assertoric proofs by incorporating the temporal modalities into the predicates. This would be a major turnaround in methods of proof: it would incorporate two non-syllogistic re-terming steps (Section 4.3) into the most basic level of formal proofs. To illustrate the case of *Festino(d,t,d)* in the style of Section 8.3:

1.	$<(i\text{-}d)(C,B),(e\text{-}t)(A,B)>[C,A]$	posit
2.	$<(i)(C^+,B^-),(e)(A^+,B^-)>[C^+,A^+]$	re-term by (10.1.8)
3.	$<(i)(C^+,B^-),(e)(B^-,A^+)>[C^+,A^+]$	by e-conversion
4.	$<(o)(C^+,A^+)>[C^+,A^+]$	by *Ferio*
5.	$<(o\text{-}d)(C,A)>[C,A]$	re-term by (10.1.8)

Closely similar moves work for the other two cases.

History 10.4.1 Of course it would be an outrageous liberty to assume that Ibn Sīnā would have used this method for the first three moods under consideration. Fortunately we don't need to assume it, because we have it from the horse's mouth. Ibn Sīnā tells us in his own words that at this point we should 'count the permanence or non-permanence as a part of the predicate' (*Qiyās* [55] 130.11f), which is precisely the re-terming move used above. Other details of this passage in *Qiyās* are confused and clearly need some kind of editing; but the fact that he tells us to incorporate a re-terming, at precisely the point in the formal development where Aristotelian methods cease to work and this device does the job required, is a clear indication that he has made the breakthrough. This is one of the lightbulb moments, not only in Ibn Sīnā's logic but in the history of pre-Fregean logic. See [45] for further discussion.

10.5 Productivity conditions and compound proof theory for dt

Besides the Philoponus conditions, the extra condition for productivity is that if there is a negative sentence ϕ, then either the predicate symbol of ϕ is the major extreme of the premise-sequence, or the predicate symbol of ϕ is the predicate symbol of another sentence ψ in the premise-sequence and at least one of ϕ and ψ has the form (d), or the premise-sequence is retrograde and ϕ has the form (d).

So for premise-pairs: if there is a negative premise, then either it is the major premise in first or third figure [which can be shortened to 'we are in first or third figure' by the assertoric conditions], or we are in second figure and at least one premise is (d), or we are in fourth figure and the negative premise has the form (d).

So ignoring fourth figure the condition is just that if we are in second figure then at least one premise is (d).

Turning to the rule of following:

If the major premise is forwards, the mood of the conclusion follows it.

If the major premise is backwards but the premise-sequence is not retrograde, then the conclusion is (d).

If the premise-sequence is retrograde, then the conclusion is (d) if and only if there is a negative premise.

This gives for premise-pairs:

In first and third figures the mood of the conclusion follows that of the major premise.

In second figure the conclusion is (d).

In fourth figure the conclusion is (d) unless the premises are all affirmative, in which case it's (t).

10.6 Listing the dt MITs

We reduced the dt fragment to a form of assertoric logic that we called $\mathcal{L}_{as\pm}$, containing sentences of the following nine forms:

(10.6.1)

1	$(a)^-$	$(a)(B^+, A^-)$
2	$(a)^+$	$(a)(B^+, A^+)$
3	$(e)^-$	$(e)(B^+, A^-)$
4	$(e)^+$	$(e)(B^+, A^+)$
5	$(i)^-$	$(i)(B^+, A^-)$
6	$(i)^+$	$(i)(B^+, A^+)$
7	$(o)^-$	$(o)(B^+, A^-)$
8	$(o)^+$	$(o)(B^+, A^+)$
9	$(a, uas)^+$	$(a, uas)(B^-, B^+)$

The nuisance (o) form is here, but for metamathematical purposes we can eliminate it by the device of Lemma 6.1.6 as follows.

Lemma 10.6.1 *Let T be a theory consisting of sentences of the forms in (10.6.1). Let T' be the theory constructed from T by replacing each sentence of the form $(o)^-$ by the sentence of the form $(o, uas)^-$ with the same tag, and each sentence of the form $(o)^+$ by the sentence of the form $(o, uas)^+$ with the same tag. Then T is consistent if and only if T' is consistent.*

Proof. The proof is exactly as that of Lemma 6.1.6. To see that the previous proof applies, we need to check that in T there is no sentence of the form $(a, uas)^-$ with the same subject symbol as a sentence of the form $(o)^-$ or $(o)^+$. But this is immediate, since a sentence of the form $(a, uas)^+$ has a subject symbol of negative polarity B^-, whereas the sentences of the form $(o)^-$ or $(o)^+$ have subject symbols B^+ of positive polarity. □

As a result of this lemma, we can find the MITs by finding them first for a logic with the nine sentence forms

(10.6.2) $(a)^-,\ (a)^+,\ (e)^-,\ (e)^+,\ (i)^-,\ (i)^+,\ (o, uas)^-,\ (o, uas)^+,\ (a, uas)^+,$

and then translating back. We will write $\mathcal{L}_{dt1}$ for the logic above, and $\mathcal{L}_{dt2}$ for the new logic. Warning: Strengthenings and weakenings are not always the same in the two logics $\mathcal{L}_{dt1}$ and $\mathcal{L}_{dt2}$. So to describe the optimally inconsistent theories in $\mathcal{L}_{dt1}$, we need first to find the MITs in $\mathcal{L}_{dt2}$ and then translate back into $\mathcal{L}_{dt1}$ before considering the possible weakenings.

The Propositional Constraints impose some restrictions on the tracks of a refutation in $\mathcal{L}_{dt2}$.

Lemma 10.6.2 *Let ρ be a refutation in $\mathcal{L}_{dt2}$. Then in both tracks of ρ:*

(a) A node carrying a relation symbol of polarity $-$ is either terminal or followed immediately by a node with theory label of the form (a, uas).

(b) A node carrying a relation symbol of polarity $+$ is either terminal or followed immediately by a node with theory label of the form (a) or (e).

Proof. If a node ν is not terminal and carries a relation symbol C, then by the Propositional Constraint, C is the subject symbol of the theory label of ν^+. □

Lemma 10.6.3 *Let ρ be a terse refutation in $\mathcal{L}_{dt2}$. If two affirmative nodes of ρ in the same track carry the same relation symbol, then they are both in the tail of ρ.*

Proof. For exactly the same reasons as BEFORE, no node is left constrained away from any other node. So the proof is covered already by that of Lemma le:6.5.2y. □

Lemma 6.4.10 no longer gives us a support in this case, because we now have sentences of the form (a, uas). Instead we can say:

Lemma 10.6.4 *Let ρ be a refutation in $\mathcal{L}_{dt2}$. Suppose μ and ν are affirmative nodes in the upper and lower tracks of ρ respectively, and the same relation symbol B^+ is at both μ and ν. Then there exists a support for μ and ν.*

Proof. The same proof as for Lemma 6.4.10 shows that at least one of the successor nodes μ^+ and ν^+ has an affirmative theory label with subject symbol B^+. Since the sentences (a, uas) all have subject symbol of polarity $-$, the theory label in question must have the form (a), so that it provides a support for μ and ν. □

We infer:

Lemma 10.6.5 *Let ρ be a terse refutation in $\mathcal{L}_{dt2}$, and μ, ν a pair of AFFIRMATIVE nodes of ρ in different tracks and carrying the same relation symbol B^+ of polarity $+$. Then μ and ν are initial nodes with theory label of the form (a), and this same label is also the theory label of the second one of one of the two tracks.*

The *proof* is AS BEFORE.

What happens when in the lemma above, the relation symbol at μ and ν has polarity $-$, say B^-? Since both nodes are affirmative, the lower is not terminal, and hence by ABOVE, ν^+ has theory label $(a, uas)(B^-, B^+)$.

We claim that μ must be terminal. For otherwise also μ^+ has theory label $(a, uas)(B^-, B^+)$ and hence μ^+ and ν^+ are affirmative nodes carrying the same relation symbol B^+. So by the previous lemma, both μ^+ and ν^+ are initial, which is a contradiction. Claim proved.

We can separate out the nominator types exactly as before:

Theorem 10.6.6 *Let ρ be a refutation in $\mathcal{L}$ without (o). Then exactly one of the following holds.*

Case (**i**) *Exactly one of the sentences in the theory of ρ has the form (i), and this sentence is the nominator sentence.*

Case (**o**) *Exactly one of the sentences in the theory of ρ has the form (o, uas), and this sentence is the nominator sentence and the theory label of the negative node.*

Case (**a**) *The theory of ρ contains no sentence of either of the forms (i) and (o, uas), and the nominator sentence has the form (a).*

We consider each case in turn. We write χ for the nominator sentence.

Case (**i**). As before, the initial nodes are the nominator nodes, and their Skolem labels are Skolem pieces of χ. As before, we rule out the possibility that both initial nodes have the secondary Skolem piece of χ; and as before, in the case where χ has the form $(i)(B^+, A^+)$ we also rule out the possibility that both initial nodes have the primary Skolem piece of χ. However, there remains the case that χ has the form $(i)(B^+, A^-)$ and both initial nodes have the primary Skolem piece of χ as their Skolem labels, so that both carry the relation symbol A^-. None of our arguments rule out this possibility. But ABOVE shows that in this case the refutation has a tail of length 1, and that the upper initial node is also terminal.

Case (**o**). As before, we rule out the possibility that both initial nodes have the secondary Skolem piece of χ. It's also impossible for both initial nodes to have the primary Skolem piece of χ as their Skolem labels, since this piece is negative and can only be the Skolem label of the negative node.

Case (**a**). Here as with (i) there is a new possibility, namely that χ has the form $(a)(B^+, A^-)$ and that the second nodes of both tracks have χ as their theory label. Since χ has only one 2-part Skolem piece, the refutation will have a tail of length at least 2; but the CLAIM shows that the upper second node in this case is terminal, so that the tail has length exactly 2.

We have enough information now to carry these refutations back to subject-predicate digraphs. Then from the digraphs we can eliminate those MITs that are not optimal. For example in the case (i) where χ has the form $(i)(B^+, A^-)$ and the refutation has a tail, the PROCEDURE produces a digraph of the form (10.2.1) above, except that any of the sentences labelled $(a\text{-}t)$ could in fact be $(a\text{-}d)$, and likewise with $(e\text{-}t)$ and $(e\text{-}d)$. But none of these theories with a d in place of a t are optimal, since (10.2.1) is inconsistent as it stands, and passing from $(a\text{-}d)$ to $(a\text{-}t)$ (or likewise with $(e\text{-}d)$) is a weakening. The same argument applies to the other subject-predicate digraphs in cases (i) and (o).

Case (a) never yields an optimally inconsistent theory. This is because the first sentence in the upper track is always of the form $(a\text{-}g)$, which can be weakened to $(i\text{-}g)$ producing an inconsistent theory of type $(i)II\uparrow$.

Theorem 10.6.7 *A set of dt sentences is optimal inconsistent if and only if it has a subject-predicate digraph of one of the types $(i)I$, $(i)II\uparrow$, $(i)II\downarrow$ or (o) as in Section 10.2 above.*

10.7 Exercises

10.1. In the proof of (b)$\Leftarrow$ of Theorem 10.1.2 we had to invoke the theory of E to get from $\exists\tau Ba\tau$ to $\exists\tau(Ea\tau \wedge Ba\tau)$. Might it be easier to define

$$(A^+)^{N^\downarrow} = \{a \in \text{dom}(N^\downarrow) : N \models \exists\tau Aa\tau\}, \tag{10.7.1}$$

making the appeal to the theory of E unnecessary? Show that the definition (10.7.1) works for the proof of the theorem just as well as (10.1.3), but it requires an appeal to the theory of E at different points in the proof.

Solution: If N is a model of Th(E) then the two definitions (10.1.3) and (10.7.1) define the same sets $(A^+)^{N^\downarrow}$. But for example in the proof that $N^\downarrow \models (o)(B^+, A^+)$ using (10.7.1), we would need to argue from $N \models \forall\tau(Ea\tau \rightarrow \neg Aa\tau)$ to $N \models \forall\tau\neg Aa\tau$, and this inference depends on Th(E).

10.2. WAS A THEOREM. USE? Every set of sentences in the dt fragment in $L(\Sigma_{2rel}(E))$ can be written as T^{DT} for a set T of assertoric sentences in $L(\Sigma^{dt}_{1rel})$ with subject symbols of the form R^+. Moreover

(a) T^{DT} is two-dimensional inconsistent if and only if T is $\pm$-inconsistent.

(b) T^{DT} is minimal two-dimensional inconsistent if and only if T is minimal $\pm$-inconsistent.

(c) T^{DT} is optimally minimal two-dimensional inconsistent if and only if T is optimally minimal $\pm$-inconsistent.

10.3. List the 14 productive fourth-figure moods in the dt fragment, together with their conclusions.

Solution. ALREADY DONE IN SECTION 10.3.

mood	minor	major	conclusion	notes
1.*Bamalip*(t,t,t)	$(a\text{-}t)(B,C)$	$(a\text{-}t)(A,B)$	$(i\text{-}t)(C,A)$	
2.*Bamalip*(t,d,t)	$(a\text{-}t)(B,C)$	$(a\text{-}t)(A,B)$	$(i\text{-}t)(C,A)$	
3.*Bamalip*(d,t,t)	$(a\text{-}t)(B,C)$	$(a\text{-}t)(A,B)$	$(i\text{-}t)(C,A)$	
4.*Bamalip*(d,d,t)	$(a\text{-}t)(B,C)$	$(a\text{-}t)(A,B)$	$(i\text{-}t)(C,A)$	
5.*Dimatis*(t,t,t)	$(a\text{-}t)(C,B)$	$(i\text{-}t)(A,C)$	$(i\text{-}t)(B,A)$	
6.*Dimatis*(t,d,t)	$(a\text{-}t)(C,B)$	$(i\text{-}t)(A,C)$	$(i\text{-}t)(B,A)$	
7.*Dimatis*(d,t,t)	$(a\text{-}t)(C,B)$	$(i\text{-}t)(A,C)$	$(i\text{-}t)(B,A)$	
8.*Dimatis*(d,d,t)	$(a\text{-}t)(C,B)$	$(i\text{-}t)(A,C)$	$(i\text{-}t)(B,A)$	
9.*Calemes*(d,t,d)	$(e\text{-}t)(B,A)$	$(a\text{-}d)(C,B)$	$(e\text{-}d)(A,C)$	
10.*Calemes*(d,d,d)	$(e\text{-}t)(B,A)$	$(a\text{-}d)(C,B)$	$(e\text{-}d)(A,C)$	
11.*Fesapo*(t,d,d)	$(a\text{-}d)(A,B)$	$(e\text{-}t)(C,A)$	$(o\text{-}d)(B,C)$.	
12.*Fesapo*(d,d,d)	$(a\text{-}d)(A,B)$	$(e\text{-}t)(C,A)$	$(o\text{-}d)(B,C)$.	
13.*Fresison*(t,d,d)	$(i\text{-}d)(A,C)$	$(e\text{-}t)(B,A)$	$(o\text{-}d)(C,B)$	
14.*Fresison*(d,d,d)	$(i\text{-}d)(A,C)$	$(e\text{-}t)(B,A)$	$(o\text{-}d)(C,B)$	

Chapter 11

The graphable moods

11.1 Listing the graphable MITs

Our treatment of 2D logic in the preceding two chapters has a serious flaw: it's piecemeal, taking different pieces of the logic one at a time without finding the patterns that govern this logic overall. The same is true of the approach taken by Ibn Sīnā himself; in concentrating on the dt fragment we have been following his lead. But in his time the general methods of today's logic were not available, so he had an excuse that we don't have.

Part of the problem is that by setting himself the task of systematically listing the two-premise inferences, Ibn Sīnā goes into an area that most modern logicians would shrink from. How would you set about listing all the first-order sequents with two premises? But besides that, there still is a lot that we just don't know about 2D logic.

In any case we certainly need methods that are more general than the reductions to assertoric logic that got us through the previous two chapters. We have already mentioned syllogisms that involve both (d) and (ℓ) sentences. These syllogisms are beyond the reach of those reductions, because the type of interpretation used to reduce sentences in the dt fragment is incompatible with the type used for the $ulem$ sentences. The assertoric objects for the second reduction are ordered pairs of object and time, whereas for the first reduction the times are incorporated into the terms rather than the objects.

It might be possible to extend the dt reduction to the dtz fragment, taking in Ibn Sīnā's 'true now' (z) sentences. At the time of writing this has not yet been tried. But it may be worthwhile, both because Ibn Sīnā himself has at least a dozen pages of material in this territory that nobody has yet made sense of, and because (as we will see in the next Chapter) the dtz fragment brings us very close to the logic of the 14th century Scholastic Jean Buridan.

However, we do still have the option of using the Skolem methods of Chapter 6 above. These are highly general and they certainly lift to the whole of 2D logic, as we will see BELOW. They tell us that every minimal inconsistent set of 2D sentences has a two-track terse refutation of the same kind of shape that we saw in Section 6.4. This doesn't immediately guarantee the Rule of Quality for 2D logic, because in general there are negative 2D sentences with affirmative Skolem pieces, and these would need to be eliminated from terse refutations. But with some work we can still recover the full Rule of Quality:

Theorem 11.1.1 *Every minimal inconsistent set of 2D sentences contains exactly one negative sentence.*

This will be Theorem BELOW.

The Rule of Quantity fails already for optimal inconsistent sets in the dtz fragment:

Example 11.1.2 The following theory is optimal inconsistent, but has two existential sentences:

$$(i\text{-}z)(D, C), (a\text{-}d)(C, B), (a\text{-}z)(B, A), (o\text{-}z)(C, A).$$

I leave it to the reader to check these properties.

We can recover this much of the Rule of Quantity:

Theorem 11.1.3 *Let T be a minimal inconsistent 2D theory. Then:*

(a) T contains at most two existential sentences.

(b) If T lies in core 2D logic then T contains at most one existential sentence.

(c) If T has more than one existential sentence then T includes both a (d) sentence and a (z) sentence, and the two existential sentences have the assertoric forms (i) and (o).

See BELOW.

A more substantial issue is the procedure for converting refutations to subject-predicate digraphs. We could always do this recasting in the assertoric case, because in this case a refutation never has two affirmative nodes outside the tail and carrying the same relation symbol REF. In core 2D logic this is no longer true:

Example 11.1.4 The horrible example.

Definition 11.1.5 Let Γ be a finite abstract digraph with a distinguished node called the *contradictory node*. We say that Γ is *normal* if it has the following properties:

(a) Γ is loopless and has at least two arrows.

(b) Each node of Γ, except possibly the contradictory node, has at most one immediate predecessor; the contradictory node has at most two.

(c) Γ has a unique circular subgraph, and it contains the contradictory node.

(d) Γ has at most one terminal node, and if it has one then the node is the contradictory node.

Theorem 11.1.6 *Let Γ be a normal finite abstract digraph. Then Γ has one of the following three forms:*

(i) Γ is a closed path of length at least 2.

(ii) Γ consists of a closed path of length at least two, and a linear subgraph of arrow-length at least 1, whose only node in common with the closed path is the contradictory node κ.

(iii) Γ is the union of two distinct linear subgraphs with the same initial node and the same terminal node, and if ν is a node that is in both subgraphs and is neither initial nor κ, then the predecessors of ν are also in both subgraphs.

Proof. Write κ for the contradictory node of Γ. Let $\mathbb{T}$ be the set of all subgraphs of Γ that are tracks whose terminal node is κ, and that are maximal in the sense that they are not final segments of other tracks with this property.

Claim One. $\mathbb{T}$ has either one or two elements, and every node of Γ is in at least one track in $\mathbb{T}$.

Proof of Claim. We can find an element of $\mathbb{T}$ by starting at κ, taking an immediate predecessor μ_1 of κ, then an immediate predecessor μ_2 of μ_1, and so on until either we come to an initial node of Γ, or we come to a node with no immediate predecessors not already listed. So $\mathbb{T}$ has cardinality at least 1. But by condition (b) this process is uniquely determined except for at most two possibilities at the first step. So $\mathbb{T}$ has cardinality either 1 or 2.

Suppose that some node ν_1 of Γ is not in any element of $\mathbb{T}$. Then by the construction of the previous paragraph, there is no track in Γ that starts at κ and finishes at ν_1. In particular ν_1 is not κ, so by condition (d) ν_1 is not terminal, and thus we can form an infinite sequence $\nu_1, \nu_2, \ldots$ where each ν_{i+1} is an immediate successor of ν_i. Since Γ is finite, there must be a repetition in this sequence, giving a circular subgraph of Γ that doesn't contain κ, and this contradicts condition (c). □ Claim One.

Claim Two. Let τ be a track in $\mathbb{T}$, and let μ be the initial node of τ. Then just one of two cases holds: (1) μ is an initial node of Γ, or (2) κ is an immediate predecessor of μ in Γ.

Proof of Claim. Suppose μ is not an initial node of Γ, so that in Γ μ has an immediate predecessor ν. Since τ is maximal, ν is a node already in τ. If ν is not κ then we have a circular subgraph of Γ that doesn't contain κ, contradicting (c). The two cases are disjoint, since if μ has an immediate predecessor then μ is not initial. □ Claim Two.

In case (1) we say that τ is *line-like*; in case (2) we say that τ is *circle-like*.

Claim Three. If $\mathbb{T}$ contains just one track, then this track is circle-like. If $\mathbb{T}$ contains two tracks, then they are not both circle-like; if they are both line-like then they have the same initial node.

Proof of Claim. If $\mathbb{T}$ consists of a single line-like track then Γ has no circular subgraph, contradicting condition (c). If $\mathbb{T}$ consists of two circle-line tracks, then there are two distinct circular subgraphs of Γ, contradicting (c) again. If τ_1 and τ_2 are two line-like tracks in $\mathbb{T}$ with distinct initial nodes, then by the construction of Claim Two, they have no nodes in common except for κ. But then Γ has no circular subgraph, contradicting (c) again. □ Claim Three.

Claim Three allows just the three possibilities (i), (ii) and (iii). □

11.2 The optimal core 2D moods

We consider here only premise-pairs and their optimal conclusions.

Barring the ungraphable cases, the syllogistic moods for core two-dimensional logic can be read off from the results of the previous section. At present it seems that the ungraphable cases were unknown until 28 November 2014, and there is no indication that Ibn Sīnā even suspected their existence. There must be very few of them, but I hope to report more on this soon.

We get an optimal syllogism by taking any optimal MIT $\{\phi_1, \phi_2, \phi_3\}$, choosing the contradictory negation of one of its sentences for the conclusion, and taking the two remaining sentences as premises. All valid but not optimal syllogisms can then be found by strengthening the premises and/or weakening the conclusion. We reviewed the results for assertoric syllogisms in Chapter 6.

Recall that we follow Ibn Sīnā in putting the minor premise first, then the major premise, then the conclusion. The major premise is the one which contains the predicate term of the conclusion.

Since every core two-dimensional syllogism projects down to an assertoric syllogism, we can name the core two-dimensional syllogisms as follows: Take the name of the assertoric syllogism that is the projectum, and add in parentheses at the end the avicennan forms of the minor premise, major premise and conclusion respectively.

For example consider *Barbara*. As we know from Section 5.3, an optimal assertoric syllogism in mood *Barbara* is got by taking an assertoric theory whose digraph is in the family **C** with two arrows in the upper track, taking the sentences corresponding to these arrows as premise and the contradictory negation of the sentence on the downwards vertical arrow as conclusion. We find the optimal core two-dimensional moods in *Barbara* by applying this same recipe to the family **2DC**. Definition de:10.2.1 tells us the possible assignments of avicennan forms; of course we must switch between d and t, and between ℓ and m, when we take the contradictory negation of the sentence on the vertical arrow.

For example one of the assignments of avicennan forms in Definition de:10.2.1 is as follows: the vertical arrow is labelled d, and the arrows in the other track are all labelled t. Listing minor premise, major premise and conclusion, and remembering to switch for the conclusion, this gives the optimal mood

$$Barbara(t, t, t). \tag{11.2.1}$$

You can read off the remaining cases from Definition de:10.2.1; there are five possible assignments in all.

It turns out that exactly the same assignments of avicennan forms apply to all first figure syllogisms, regardless of their aristotelian forms. We read them off as follows:

	minor	*major*	*conc*
(i)	t	t	t
(ii)	d	ℓ	d
(iii)	t	d	d
(iv)	ℓ	ℓ	ℓ
(v)	m	ℓ	m

Figure 11.1: Optimal avicennan moods in first figure

The same applies to second, third and fourth figures. For the fourth figure we are in the special case of family **2DB** with empty upper track, as in Definition de:10.2.2. Here are the results:

	minor	*major*	*conc*
(i)	t	ℓ	t
(ii)	t	d	d
(iii)	d	t	d
(iv)	ℓ	ℓ	ℓ
(v)	m	ℓ	m

Figure 11.2: Optimal avicennan moods in second figure

	minor	*major*	*conc*
(i)	d	t	m
(ii)	t	t	t
(iii)	t	d	d
(iv)	m	ℓ	m
(v)	ℓ	m	m

Figure 11.3: Optimal avicennan moods in third figure

MUST REDO THIS ONE PROPERLY.

	$minor$	$major$	$conc$
(i)	$(i\text{-}t)$	d	d
(ii)	t	$(i\text{-}t)$	t
(iii)	d	t	$(e\text{-}d)$

Figure 11.4: Optimal avicennan moods in fourth figure

Theorem 11.2.1 *Each of the twelve optimal assertoric moods in first to third figures yields five core two-dimensional moods, giving 60 in all. The three optimal assertoric fourth figure moods yield a further 9 core two-dimensional moods.*

We begin by using the results of Section se:24 to catalogue the optimal two-premise syllogisms in core 2D logic. These are the cases that Ibn Sīnā discusses in most detail.

The first three figures refer to tailless two-track digraphs that have one track with two arrows and one track with one arrow. For example in first figure the minor premise is the label on the root arrow of the two-arrow track, the major premise is the label on the other arrow of this track, and the conclusion is the contradictory negation of the label on the arrow of the other track. For the relevant results in Section se:24 it makes no difference whether the existential sentence is in the one-arrow track or the two-arrow, or whether it is (i) or (o). This is thanks to Orthogonality. So we can give a chart for all first figure syllogisms without distinguishing between moods.

As a first step we chart the possibilities for an optimal inconsistent theory:

	2-arrow track 1st arrow	2-arrow track 2nd arrow	1-arrow track
(i)	d	ℓ	t
(ii)	ℓ	ℓ	m
(iii)	m	ℓ	ℓ
(iv)	t	d	t
(v)	t	t	d

The ordering of the rows in the chart above is lexicographic, taking strong before weak. The rows themselves come from Corollary co:amanita6 above. There are three 'strong' rows according to where we put the d, namely (i), (iv) and (v). There are two 'weak' rows according to whether the m is at the start of the 2-arrow track or the 1-arrow track, namely (iii) and (ii).

Converting the chart to syllogisms means replacing the sentence in the 1-arrow track by its contradictory negation, which switches its avicennan form (permuting d and t, and permuting m and ℓ), as follows.

	$minor$	$major$	$conc$
(i)	d	ℓ	d
(ii)	ℓ	ℓ	ℓ
(iii)	m	ℓ	m
(iv)	t	d	d
(v)	t	t	t

Converting this table to second and third figures is a routine task. For second figure:

	$minor$	$major$	$conc$
(i)	d	t	d
(ii)	ℓ	ℓ	ℓ
(iii)	m	ℓ	m
(iv)	t	d	d
(v)	t	ℓ	t

And for third figure:

	$minor$	$major$	$conc$
(i)	d	t	m
(v)	ℓ	m	m
(iv)	m	ℓ	m
(iii)	t	d	d
(ii)	t	t	t

We may as well add the facts for fourth figure, though Ibn Sīnā has already declared his lack of interest in this case. I give the results as for *Dimatis*, i.e. with the conclusion being the contradictory negation of the arrow pointing to the root.

	$minor$	$major$	$conc$
(i)	m	ℓ	m
(ii)	t	d	m
(iii)	t	t	t

Ibn Sīnā lists the conclusion-optimal syllogisms; i.e. for each set of premises that yields a syllogistic conclusion, he lists the strongest conclusion that can be got. We can reconstruct his list by allowing all strengthenings of the premises in the tables above, except where the strengthening allows a stronger conclusion. To save the reader the trouble I give the results below.

First figure, conclusion-optimal:

	$minor$	$major$	$conc$
(i)	d	d	d
(ii)	d	ℓ	d
(iii)	d	m	t
(iv)	d	t	t
(v)	ℓ	d	d
(vi)	ℓ	ℓ	ℓ
(vii)	ℓ	m	t
$(viii)$	ℓ	t	t
(ix)	m	d	d
(x)	m	ℓ	m
(xi)	m	m	t
(xii)	m	t	t
$(xiii)$	t	d	d
(xiv)	t	ℓ	t
(xv)	t	m	t
(xvi)	t	t	t

Second figure, conclusion-optimal:

	$minor$	$major$	$conc$
(i)	d	d	d
(ii)	d	ℓ	d
(iii)	d	m	d
(iv)	d	t	d
(v)	ℓ	d	d
(vi)	ℓ	ℓ	ℓ
(vii)	m	d	d
$(viii)$	m	ℓ	m
(ix)	t	d	d
(x)	t	ℓ	t

Third figure, conclusion-optimal:

	minor	*major*	*conc*
(i)	d	d	d
(ii)	d	ℓ	m
(iii)	d	m	m
(iv)	d	t	m
(v)	ℓ	d	d
(vi)	ℓ	ℓ	m
(vii)	ℓ	m	m
$(viii)$	ℓ	t	t
(ix)	m	d	d
(x)	m	ℓ	m
(xi)	m	m	t
(xii)	m	t	t
$(xiii)$	t	d	d
(xiv)	t	ℓ	t
(xv)	t	m	t
(xvi)	t	t	t

Ibn Sīnā's lists are complete only for the dt fragment. Thus in first figure he lists:

$$ddd, dtt, tdd, ttt.$$

And in second figure:

$$ddd, dtd, tdd.$$

And in third figure:

$$ddd, tdd, ttt.$$

In fact Ibn Sīnā adds to this list dtt. From the table we see that this is not conclusion-optimal; we could have got dtm. So dtt is conclusion-optimal only relative to the dt fragment. This is consistent with what we find elsewhere, that Ibn Sīnā's listings are all relative to some particular fragment of logic. Thus in listing first-order figures he ignores premise-sets that are productive in other figures but not in first figure.

We turn to theories with tailed digraphs. Ibn Sīnā was aware of these, but he treated them as inconsistent sets and not as syllogisms. It suits us fine to do the same here.

For the tailed two-track digraph there are just two possibilities. We use the numbering of arrows from the diagrams in Section se:22.

	$T1$	$U1$	$L1$
(i)	d	ℓ	t
(ii)	d	t	ℓ

For the tailed fourth-figure digraph there is only one possibility:

	$T1$	$L1$	$L2$
(i)	d	t	t

11.3 Implications for compound proofs

Is Ibn Sīnā right to claim his assertoric proof theory lifts to recombinant in general?

11.4 Janus sentences

Ibn Sīnā also introduced to 2D logic some sentences that we haven't considered yet; I will call them *Janus* sentences because they face two ways, affirmative and negative. In Ibn Sīnā's time it wasn't yet the habit to pin down a logic to an exact set of sentences. (It seems to have been Faḵr al-Dīn in the late twelfth century who first recommended this practice.) So for convenience I excluded Janus sentences from 2D logic in earlier sections of this paper. With these sentences included, I will speak of *extended 2D logic*. When the Janus sentences are excluded I speak of *unextended 2D logic*.

In fact the exact choice of Janus sentences in Ibn Sīnā's own system is not entirely clear. I will go for inclusivity; but one can still raise questions about where he intended to take the existential or universal augments.

Let ψ be a universal 2D sentence $\forall x\phi(x)$. Then we write $\psi(x)^+$ for the formula $\phi(x)$ got by removing the initial object quantifier, and $\psi(x)^-$ for the internal negation of $\psi(x)^+$, i.e. the formula got by changing a to e or vice versa. Then we can consider four new sentences as follows; their labels are on the left.

(11.4.1)
$$\begin{array}{ll} J_\wedge\psi: & \forall x(\phi(x)^+ \wedge \phi(x)^-) \\ J^\wedge\psi: & (\forall x\phi(x)^+ \wedge \forall x\phi(x)^-) \\ J_\vee\psi: & \forall x(\phi(x)^+ \vee \phi(x)^-) \\ J^\vee\psi: & (\forall x\phi(x)^+ \vee \forall x\phi(x)^-) \end{array}$$

(Upstairs the propositional operator has wide scope; downstairs it has narrow scope.) The first and second sentences in (11.4.1) are logically equivalent.

We call the two sentences $\forall x\phi(x)^+$ and $\forall x\phi(x)^-$ the two *faces* of each of the sentences above; one is affirmative and the other is negative.

Dually we can do the same thing to existential sentences ψ of the form $\exists x\phi(x)$, again getting four Janus sentences:

(11.4.2)
$$\begin{array}{ll} J_\wedge\psi: & \exists x(\phi(x)^+ \wedge \phi(x)^-) \\ J^\wedge\psi: & (\exists x\phi(x)^+ \wedge \exists x\phi(x)^-) \\ J_\vee\psi: & \exists x(\phi(x)^+ \vee \phi(x)^-) \\ J^\vee\psi: & (\exists x(\phi(x)^+ \vee \exists x\phi(x)^-). \end{array}$$

We get the same eight sentences if we take ϕ to be negative as we do with ϕ affirmative. Each of these sentences has two faces, $\exists x\phi(x)^+$ and $\exists x\phi(x)^-$, one affirmative and one negative.

We describe a Janus sentence as *disjunctive* if it has $\vee$ in the middle, and *conjunctive* if it has $\wedge$ in the middle. We describe the Janus sentences in (11.4.1) as *universal*, and those in (11.4.2) as *existential*. We describe the Janus sentences of the form $J^\wedge\phi$ and $J^\vee$ as *propositional*, to indicate that they are propositional compounds of unextended 2D sentences.

Not all the Janus sentences are useful. For example:

Lemma 11.4.1 *A Janus sentence of any of the following forms is inconsistent:*

(11.4.3)
$$\begin{array}{lll} J^\wedge(a\text{-}d), & J^\wedge(a\text{-}\ell), & J^\wedge(a\text{-}z), \\ J_\wedge(a\text{-}d), & J_\wedge(a\text{-}\ell), & J_\wedge(a\text{-}z), \\ J_\wedge(i\text{-}d), & J_\wedge(i\text{-}\ell), & J_\wedge(i\text{-}z). \end{array}$$

□

Another general fact:

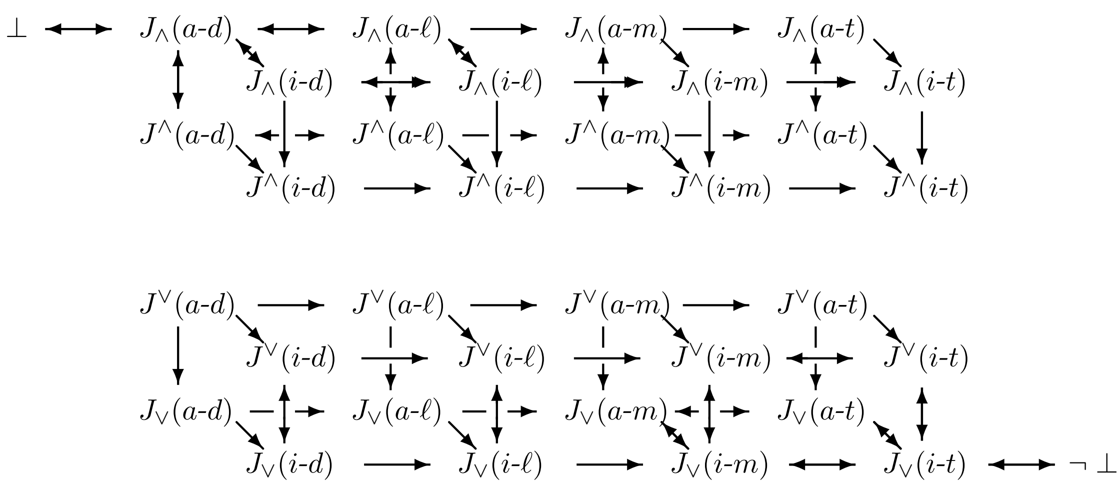
Lemma 11.4.2 *Let ϕ be any unextended 2D sentence. Then*

$$J_\wedge\phi \;\vdash\; J^\wedge\phi \;\vdash\; \phi \;\vdash\; J^\vee\phi \;\vdash\; J_\vee\phi.$$

□

The next lemma gives more detail for the extended core 2D sentences:

Lemma 11.4.3 *The following implications hold. Each sentence in the upper block entails the corresponding sentence in the lower block.*



Lemma 11.4.4 *Writing* $^{-}$ *for contradictory negation,*

$$\overline{J^{\wedge}\phi} = J^{\vee}\overline{\phi}$$

and

$$\overline{J_{\wedge}\phi} = J_{\vee}\overline{\phi}.$$

□

Ibn Sīnā introduces $J_{\wedge}(a\text{-}t)$ and $J_{\wedge}(i\text{-}t)$ as *wujūdī*, evidently as temporal analogues of Aristotle's 'contingent'. In *Easterners* he also mentions the pair $J_{\wedge}(a\text{-}m)$ and $J_{\wedge}(i\text{-}m)$.

Lemma 11.4.5 *Extended 2D logic is decidable.*

Proof. The forms of the Janus sentences show that each of them can be written using just two variables, so that we can quote Mortimer's theorem again. □

Adding propositional Janus sentences to a set of 2D sentences introduces nothing essentially new. Conjunctive propositional Janus sentences can be regarded as pairs of sentences. We can handle disjunctive Janus sentences with the following elementary fact.

Fact 11.4.6 *Let T be a first-order theory and for each $i \in I$ let ϕ_i be a sentence $(\phi_i^1 \vee \phi_i^2)$. Then the theory*

$$T \cup \{\phi_i : i \in I\}$$

is inconsistent if and only if for every function $f : I \to \{1, 2\}$ the theory

$$T \cup \{\phi_i^{f(i)} : i \in I\}$$

is inconsistent. (The function f in this context is called a transversal*, and we write T_f for the theory exhibited above.)*

From this point, we could aim to develop a logic of extended 2D sentences. The patterns would be more complicated than for unextended 2D logic. For example here is a minimal inconsistent theory:

(11.4.4) $J^{\vee}(i\text{-}m)(C, B), (e\text{-}\ell)(C, A), (a\text{-}\ell)(B, A), (a\text{-}\ell)(C, D), (a\text{-}\ell)(B, D).$

Simiar ideas rapidly lead to arbitrarily complicated examples. Going down this road we would double the length of the paper. But happily Ibn Sīnā considers only some very special cases, as in the following section.

11.5 Exercises

1.

Chapter 12

Cashing in

12.1 Ibn Sīnā's *ḵabṭ*

To understand Ibn Sīnā's alethic modal logic is above all to make sense of his *ḵabṭ*. As we noted in Section 1.1, this is his practice of confusing temporal modalities with alethic ones—and Rāzī was surely right to suppose that at least sometimes Ibn Sīnā uses the alethic modality 'necessary' to express 'inevitable'.

Any explanation of the *ḵabṭ* has to rest on a close study of Ibn Sīnā's text. For that reason it belongs in [45] rather than the present book. But the reader is entitled to be told the outcome, so let me sketch it here. I concentrate on *Qiyās*, which is Ibn Sīnā's fullest surviving mature treatment.

Permanence is a kind of necessity, and Ibn Sīnā himself labels his 'permanence' (d) sentences as 'necessary' (*ḍarūrī*). So there is a body of material in *Qiyās* that can be understood as talking simply about the two-dimensional logic, with 'necessary' understood as (d). This material includes most of books i–iv of *Qiyās*, but excluding the material (chiefly in book iv) that explicitly discusses possibility and contingency. Ibn Sīnā himself clearly demarcates the material on possibility and contingency. So we are left with what I will call the *permanence material* of the development of formal logic in *Qiyās* i–iv.

Resolution of the *ḵabṭ* problem falls into two parts. The first, which is now easy, is to explain the permanence material. The second is to show how the permanence material fits into *Qiyās* books i–iv as a whole, and how it relates to the material in *Qiyās* that is not about permanence. This second part will probably always give rise to controversy, but I think the core points now become clear.

The permanence material forms a textbook of two-dimensional logic. It begins with a statement of the sentence forms of two-dimensional logic, and goes on to consider their contradictory negations, conversions and ecTheses, concentrating on the *dt* fragment but not neglecting other parts of the logic. After this, the textbook lists all the moods of the *dt* fragment in First, Second and Third Figures, and spells out a complete proof calculus for these. A later part of *Qiyās* extends the proof theory to take account of syllogisms in the *dt* fragment with more than two premises.

This material is a textbook of a particular logic in very much the same sense as, say, *Prior Analytics* i is a textbook of syllogisms, or Hilbert and Ackermann [37] is a textbook of first and second order logic. It establishes Ibn Sīnā as one of the major creative minds in the history of logic.

The textbook contains very few errors. The only one of any significance that I know of is an incorrect statement of the contradictory negations of the Janus sentences that Ibn Sīnā describes as *wujūdī*. This mistake was pointed out already by Rāzī (as we saw in Section 1.1 above), and in the modern literature it has been noted by Chatti REF. The faulty material is easily detached and has few consequences elsewhere.

On the plus side, the textbook contains some very original and sophisticated material. Three items that stand out are the use of paraphrase in the justification of *Cesare* and other Second Figure syllogisms; the discovery of minimal inconsistent sets that are not graph-circular (correcting a subtle error in Aristotle's account as it is normally read); and the proof search algorithm.

So the permanence material can take its place as one of the classics of pre-modern logic. But there remains the important task of accounting for the material involving possibility and contingency, and also of explaining why the textbook material was embedded in the *ḵabṭ*. There seems to be something not altogether professional about the way Ibn Sīnā has set out his stall.

The first observation to make—and it is not a new one—is that the moods that Ibn Sīnā accepts in necessary/possible are in exact correspondence with the ones that he accepts in $(d)/(t)$. Now the textbook derives these moods in the *dt* fragment from axioms; so if Ibn Sīnā wants to show that necessary/possible obeys the same logical rules, he need only show that it obeys the axioms. Looking at how he lays out this part of the workings, we see that the axioms are exactly where he concentrates his arguments. So I believe we can say: Ibn Sīnā aims to show that the pair necessary/possible obeys the same logical axioms as the pair $(d)/(t)$.

The next puzzle is why Ibn Sīnā's arguments to verify these axioms are so unconvincing. They are actually worse than is generally realised, because in them Ibn Sīnā uses methods of Aristotle which elsewhere he rightly dismisses as unsound. The answer to this puzzle lies in Ibn Sīnā's theory of the nature of logic as a science. The axioms of logic are not derived by formal argument from other truths of logic; if they were, they wouldn't be the axioms. So in this part of his work, Ibn Sīnā has to rely on other approaches, such as conceptual analysis, which for him involves tasting and testing concepts from any angle that may help to consolidate our intuitions. This exactly matches the style and presentation of the arguments that Ibn Sīnā deploys in this section of his work.

If this approach to Ibn Sīnā is right, then we should see Ibn Sīnā's treatment of necessity and possibility in *Qiyās* as an attempt to discover the logic of the notion of necessity as an abstract notion. I say only 'necessity' here, because Ibn Sīnā spells out in detail that possibility is a derivative concept, definable in terms of necessity. This incidentally helps to account for one conspicuous difference between his treatment of the dt fragment and his treatment of necessary/possible, which is that in the latter case he relies heavily on reductio ad absurdum. He seems to believe that reductio is relevant precisely because 'possible' is 'not necessary that not' (the De Morgan dual).

There are of course various kinds of necessity. To the extent that they are kinds of necessity at all, they will obey the same logic as the abstract notion. So all logical laws of necessity in the abstract will also be laws obeyed by (d) in the dt fragment. Ibn Sīnā hopes to show that the converse is true too: any logical laws of (d) will be laws of necessity too. Not everybody has been convinced; there is certainly room for alternative views. But I hope that we can agree about the nature of the disagreement. It is not a disagreement within formal logic; it's a disagreement about whether all forms of necessity obey the same formal laws.

The current literature contains two main claims that contradict the account in the previous two paragraphs. One of these is the claim that Ibn Sīnā's treatment of necessary/possible is aimed at commenting on Aristotle's account. The other is the claim that Ibn Sīnā aims to study not necessity in the abstract, but a particular form of metaphysical or ontological necessity. Ibn Sīnā anticipated that people would read both of these motivations into his logic, and so he took care to deny both of them. Gutas's book [34] is excellent on this issue; see particularly his pages 29–41 on Ibn

Sīnā's attitude to Aristotle's logic, and 300–303 on Ibn Sīnā's separation of logic and metaphysics. These are not issues of formal logic, so I need say little more about them here; they will both be discussed more fully in [45]. I will just comment that the first claim is true in the literal sense that Ibn Sīnā in *Muḵtaṣar*, *Najāt* and *Qiyās* presents his material in the form of a commentary on the *Organon*. This is a matter of style rather than motivation; but it's entirely possible that his motivation became clearer and narrower as his understanding matured.

A loose end is that the account above deals with Ibn Sīnā's treatment of the pair necessary/possible, but not with his treatment of the pair (t)/possible. This latter pair of concepts doesn't fit into the rest of Ibn Sīnā's scheme, because (t) and 'possible' are the De Morgan duals of different kinds of necessity. Ibn Sīnā—unlike Rāzī—has no interest in building up a logic that handles two kinds of necessity simultaneously. So Ibn Sīnā can deal with these cases only ad hoc. This accounts for the fact that these moods include the one case where Ibn Sīnā states in different places different conclusions for the same mood. It's also worth noting that Ḵūnajī's counterexamples to Ibn Sīnā's modal logic rest precisely on taking two different kinds of necessity at different points in one pair of premises. REF. This is an undeniable black hole in Ibn Sīnā's modal logic.

There still remains the problem of the *ḵabṭ*. Why say a thing confusingly if you can say it clearly?

My understanding of this issue has been hugely clarified by reading Spencer Johnston's PhD thesis [64]. For the last two years I have been saying that for Ibn Sīnā the two-dimensional logic is related to the alethic logic more or less as a Kripke semantics is related to a modern modal logic; but then I have apologised for being anachronistic. Johnston applies a Kripke semantics to the divided modal logic of Jean Buridan. It comes to light that Johnston's Kripke semantics is formally identical to parts of the dtz fragment of two-dimensional logic—except for the small and mainly irrelevant proviso that Johnston makes slightly different assumptions from Ibn Sīnā about the role of the E relation. The translation between Johnston and Ibn Sīnā will be set out in the next two sections below.

The question asks itself: If our understanding of Buridan's divided modals benefits from having a Kripke semantics, then doesn't it follow that Ibn Sīnā's understanding of alethic modal logic should benefit from its translation into two-dimensional logic, in exactly the same way? A full answer needs to pin down the benefits more precisely. I think most modern logicians would point to the fact that two-dimensional logic is extensional, i.e. that it has a semantics in elementary set theory, so that the laws of two-dimensional logic are objective and uncontroversial. So the 2D logic is like Kripke semantics in that it provides a fixed point or fulcrum for tackling the much more controversial concepts that tend to be involved in the study of alethic modal logic.

So in a nutshell, Ibn Sīnā hopes to clarify the logic of necessity by giving it an extensional semantics. It sticks out like a sore thumb that this is a modern programme, not one we expect from an author in the 11th century. But for this reason it should be no surprise to find that Ibn Sīnā is feeling his way and is unsteady in his way of presenting this material. Today we would say 'Here on the left is a modal sentence, and here on the right is its semantic interpretation'. Ibn Sīnā doesn't do this; instead he runs together the alethic and the 2D formulations, hoping that we can see them as carrying the same logical information. This is his *ḵabṭ*. It was a bad move, though it was made for insightful reasons.

12.2 E and E-free

When we come to develop a Skolem proof theory for 2D logic, any simplification of the sentence forms will be welcome. So it's good to know that within core 2D logic, the class of inconsistent sets of sentences is unaffected if we simply delete the relation symbol E. Deleting E means passing to a different logic, which we define as follows.

Definition 12.2.1 By *E-free two-dimensional logic* we mean the following logic $\mathcal{L}_{2d/E}$. The sentence forms of $\mathcal{L}_{2d/E}$ are named the same way as those of two-dimensional logic $\mathcal{L}_{2d}$, but with '$/E$' immediately after the avicennan form, as in $(a\text{-}t/E)$, $(e\text{-}d/E)$ etc. The corresponding sentences are as in Figure 12.1 below. If ϕ is a two-dimensional sentence, we write $\phi^{/E}$ for the corresponding sentence of E-free 2D logic. By *core E-free 2D logic* we mean the fragment of E-free 2D logic whose sentences are of the form $\phi^{/E}$ with ϕ in core 2D logic.

Theorem 12.2.2 *(a) The mapping $\phi \mapsto \phi^{/E}$ is a bijection between the sentences of 2D logic and the sentences of E-free 2D logic, and its restriction to core 2D logic is a bijection between the sentences of core 2D logic and those of core E-free 2D logic.*

(b) Let T be a set of sentences of core 2D logic, and write $T^{/E}$ for $\{\phi^{/E} : \phi \in T\}$. Then $T^{/E}$ is consistent if and only if $T \cup Th(E)$ is consistent.

name	sentence
$(a\text{-}d/E)(B,A)$	$(\forall x(\exists\tau Bx\tau \to \forall\tau Ax\tau) \land \exists x\exists\tau Bx\tau)$
$(a\text{-}\ell/E)(B,A)$	$(\forall x(\exists\tau Bx\tau \to \forall\tau(Bx\tau \to Ax\tau)) \land \exists x\exists\tau Bx\tau)$
$(a\text{-}m/E)(B,A)$	$(\forall x(\exists\tau Bx\tau \to \exists\tau(Bx\tau \land Ax\tau)) \land \exists x\exists\tau Bx\tau)$
$(a\text{-}t/E)(B,A)$	$(\forall x(\exists\tau Bx\tau \to \exists\tau Ax\tau) \land \exists x\exists\tau Bx\tau)$
$(e\text{-}d/E)(B,A)$	$\forall x(\exists\tau Bx\tau \to \forall\tau\neg Ax\tau)$
$(e\text{-}\ell/E)(B,A)$	$\forall x(\exists\tau Bx\tau \to \forall\tau(Bx\tau \to \neg Ax\tau))$
$(e\text{-}m/E)(B,A)$	$\forall x(\exists\tau Bx\tau \to \exists\tau(Bx\tau \land \neg Ax\tau))$
$(e\text{-}t/E)(B,A)$	$\forall x(\exists\tau Bx\tau \to \exists\tau\neg Ax\tau)$
$(i\text{-}d/E)(B,A)$	$\exists x(\exists\tau Bx\tau \land \forall\tau Ax\tau)$
$(i\text{-}\ell/E)(B,A)$	$\exists x(\exists\tau Bx\tau \land \forall\tau(Bx\tau \to Ax\tau))$
$(i\text{-}m/E)(B,A)$	$\exists x(\exists\tau Bx\tau \land \exists\tau(Bx\tau \land Ax\tau))$
$(i\text{-}t/E)(B,A)$	$\exists x(\exists\tau Bx\tau \land \exists\tau Ax\tau)$
$(o\text{-}d/E)(B,A)$	$(\exists x(\exists\tau Bx\tau \land \forall\tau\neg Ax\tau) \lor \forall x\forall\tau\neg Bx\tau)$
$(o\text{-}\ell/E)(B,A)$	$(\exists x(\exists\tau Bx\tau \land \forall\tau(Bx\tau \to \neg Ax\tau)) \lor \forall x\forall\tau\neg Bx\tau)$
$(o\text{-}m/E)(B,A)$	$(\exists x(\exists\tau Bx\tau \land \exists\tau(Bx\tau \land \neg Ax\tau)) \lor \forall x\forall\tau\neg Bx\tau)$
$(o\text{-}t/E)(B,A)$	$(\exists x(\exists\tau Bx\tau \land \exists\tau\neg Ax\tau) \lor \forall x\forall\tau\neg Bx\tau)$
$(a\text{-}z/E)(B,A)$	$(\forall x(Bx\delta \to Ax\delta) \land \exists xBx\delta)$
$(e\text{-}z/E)(B,A)$	$\forall x(Bx\delta \to \neg Ax\delta)$
$(i\text{-}z/E)(B,A)$	$\exists x(Bx\delta \land Ax\delta)$
$(o\text{-}z/E)(B,A)$	$(\exists x(Bx\delta \land \neg Ax\delta) \lor \forall x\neg Bx\delta)$

Figure 12.1: The E-free two-dimensional sentences

(c) The mapping $\phi \mapsto \phi^{/E}$ from 2D sentences to E-free 2D sentences respects contradictory negation and conversion.

Proof. (a) and (c) are clear.

(b) Let Σ be a 2D signature. We consider models of T as $\Sigma(E)$-structures and models of $T^{/E}$ as Σ-structures.

$\Rightarrow$: This is the conceptual direction; the idea is simply to make E true everywhere. Let the Σ-structure M be a model of T. Convert M to a $\Sigma(E)$-structure $M^{\downarrow}$ by putting

(12.2.1) $(a, \alpha) \in E^{M^{\downarrow}}$ if and only if a is an object of M and α a time of M.

Then $M^{\downarrow}$ is a model of Th(E). Moreover in $M^{\downarrow}$, every formula $\psi(x, \tau)$ is equivalent to both $(Ex\tau \wedge \psi(x,\tau))$ and $(Ex\tau \to \psi(x,\tau))$; it follows that $M^{\downarrow}$ is a model of T.

$\Leftarrow$: This direction is more severely technical and we assume some knowledge of model theory. Let the $\Sigma(E)$-structure N be a model of $T \cup \mathrm{Th}(E)$. For each object a of N, we write $E(a,N)$ for the set of times α such that $(a,\alpha) \in E^N$. By the theory of E, each set $E(a,N)$ is nonempty. Since $=$ is not in the language, we can add copies of elements without changing what sentences are true in N. So we can assume that for some cardinal κ, the time domain of N and all the sets $E(a,N)$ have cardinality κ.

We will analyse how the set of sentences true in N depends on what is true at single objects. We define a *splay on* N to be a nonempty family $(N_\omega : \omega \in \Omega)$ of $\Sigma(E)$-structures such that each N_ω has a single object a_ω, together with a map i_ω taking the single object of N_ω to an object of N, and an injective map j_ω from the time domain of N_ω to the time domain of N, in such a way that

(a) for each $\omega \in \Omega$, i_ω and j_ω together form an embedding of N_ω in N;

(b) the object domain of N is the union of the images of the maps i_ω ($\omega \in \Omega$);

(c) j_ω is bijective between $E(a_\omega, N_\omega)$ and $E(i_\omega a_\omega, N)$.

REMOVE REDUNDANT COMPLICATIONS.

Claim One. For each object a of N, let N_a be the substructure of N whose object domain is $\{a\}$ and whose time domain is the time domain of N. Then the family $(N_a : a \text{ an object of } N)$, with the identity maps i_a and j_a, forms a splay on N.

Proof of Claim. This is immediate from the definitions. $\square$ Claim One.

Let Φ be the set of formulas $\phi(x)$ of $L(\Sigma(E))$ such that

(i) $\phi(x)$ has just one free variable x, which is of sort $object$;

(ii) $\phi(x)$ has no quantifiers of sort $object$; and

(iii) all occurrences of time quantifiers $\exists\sigma$ in $\phi(x)$ are in a context $\exists\sigma(Ex\sigma \wedge \psi)$, and likewise all occurrences of quantifiers $\forall\sigma$ are in a context $\forall\sigma(Ex\sigma \rightarrow \psi)$.

Note that the subformulas of a formula in Φ also satisfy (ii) and (iii), and have no object variables apart from x, though they may have other time variables.

Claim Two. Suppose M_1 and M_2 are $\Sigma(E)$-structures and (i, j) is an embedding of M_1 in M_2 with i mapping the object domain of M_1 into that of M_2 and j mapping the time domain of M_1 into that of M_2. Suppose that for every object a of M_1, j maps $E(a, M_1)$ onto $E(ia, M_2)$. Let a be an object of M_1 and $\phi(x)$ a formula in Φ. Then

(12.2.2) $$M_1 \models \phi[a] \;\Leftrightarrow\; M_2 \models \phi[ia].$$

Proof of Claim. We prove the corresponding result for all subformulas of formulas in Φ, by induction on complexity. The only nontrivial cases are where the subformula begins with a quantifier; by (ii) this quantifier is of sort $time$, and (iii) requires it to be in a certain context. Thus we have

$$\begin{array}{lll} & M_1 \models \forall\sigma(Ea\sigma \rightarrow \psi(a, \sigma, \bar{\beta})) & \\ \Leftrightarrow & M_1 \models \psi(a, \alpha, \bar{\beta}) \text{ for all } \alpha \in E(a, M_1) & \\ \Leftrightarrow & M_2 \models \psi(ia, j\alpha, j\bar{\beta}) \text{ for all } \alpha \in E(a, M_1) & \text{by ind hyp} \\ \Leftrightarrow & M_2 \models \psi(ia, \alpha, j\bar{\beta}) \text{ for all } \alpha \in E(ia, M_2) & \text{by assumption on } j \\ \Leftrightarrow & M_2 \models \forall\sigma(E(ia, \sigma) \rightarrow \psi(ia, \sigma, j\bar{\beta})). & \end{array}$$

The argument for $\exists\sigma$ is closely similar. □ Claim Two.

Claim Three. Every core 2D sentence is a boolean combination of sentences that are equivalent, modulo Th(E), to one of the two forms $\forall x\phi(x)$ and $\exists x\phi(x)$ with $\phi \in \Phi$.

Proof of Claim. One can check this for each of the sentences in Figure 19.1, using (9.1.6). For example $\forall\tau(Bx\tau \rightarrow Ax\tau)$ is equivalent modulo

Th(E) to $\forall\tau((Ex\tau \wedge Bx\tau) \to Ax\tau)$, and hence to $\forall\tau(Ex\tau \to (Bx\tau \to Ax\tau))$.
□ Claim Three.

Claim Four. Consider a splay $(N'_\omega : \omega \in \Omega)$ on a $\Sigma(E)$-structure N'. Let T' be a set of sentences of $L(\Sigma(E))$, each of which has the form either $\exists x\phi(x)$ or $\forall x\phi(x)$ for some $\phi(x)$ in Φ. Then the following are equivalent:

(a) N' is a model of T'.

(b) Each sentence in T' of the form $\forall x\phi(x)$ is true in each N'_ω, and each sentence in T' of the form $\exists x\phi(x)$ is true in at least one N'_ω.

Proof of Claim. For each N'_ω, write a_ω for the unique object of N'_ω.

Assume (a). Then for each object a of N', if $N' \models \forall x\phi(x)$ then $N' \models \phi[a]$, so $N'_a \models \phi[a_\omega]$ by Claim Three. Hence $N'_\omega \models \forall x\phi(x)$ since a_ω is the only object. If $N' \models \exists x\phi(x)$ then for some object a, $N' \models \phi[a]$, so $N'_a \models \phi[a'_\omega$ as before, whence $N'_\omega \models \exists x\phi(x)$. This proves (b).

Assume (b). Suppose $N'_a \models \forall x\phi(x)$ for all objects a. Then for each ω, $N'_a \models \phi[a'_\omega$, so $N' \models \phi[a]$ by Claim Three. Then $N' \models \forall x\phi(x)$. The argument for $\exists x\phi(x)$ is similar. This proves (a). □ Claim Four.

Now define $T_\forall$ to be the set of all sentences of the form $\forall x\phi(x)$ with $\phi \in \Phi$ which are true in N, and likewise $T_\exists$ with $\exists x\phi(x)$. If we apply Claim Four to the splay of Claim One, then we infer that every sentence of $T_\forall$ is true in every N_a, and every sentence of $T_\exists$ is true in at least one N_a.

Now choose any set $\mathcal{T}$ of cardinality κ, and for each object a of N, choose a bijection $j_a : E(a, N) \to \mathcal{T}$. Build a $\Sigma(E)$-structure N' by taking the object domain of N' to be that of N, and the time domain to be $\mathcal{T}$, and interpreting the symbols of $\Sigma(E)$ in N' so that each (i_a, j_a) is an embedding of N_a into N'. Then $(N_a : a$ an object of $N)$, with the maps i_a, j_a, is a splay on N'. So by Claims Five and Four, N' is a model of $T_\forall \cup T_\exists$. Also by construction $E^{N'}$ is the set of all pairs (a, α) where a is an object of N' and α is a time of N'.

Claim Five. N' is a model of T.

Proof of Claim. By Claim Four it suffices to show that every sentence of $T_\forall$ is true in every N_a, and every sentence of $T_\exists$ is true in at least one N_a □ Claim Six. But we have already seen that this follows from applying Claim

Four to the splay of Claim One. □ Claim Five.

It has taken us a lot of work to start from a model N of T and finish up with a model N' of T. But the added value is that the relation E holds everywhere in N', and so N' is also a model of $T^{/E}$. Take $N^{\uparrow}$ to be the Σ-structure got from N' by ignoring E. Then $N^{\uparrow}$ is a model of $T^{/E}$ as required. □

Corollary 12.2.3 *Let Σ be a two-dimensional signature. Then for every $\Sigma(E)$-structure N there is a $\Sigma(E)$-structure N' which is a model of the same two-dimensional sentences as N, and is a model of $\forall x \forall \tau Ex\tau$.*

Proof. The proof of Theorem 12.2.2(b) contained a proof of this. □

The proof of Theorem 12.2.2 takes place in an atmosphere close to the 'guarded formulas' of Andréka, Németi and Van Benthem [6]. At present I don't know whether this fact is significant, or whether guarded methods would simplify the proof. Be that as it may, an unexpectedly difficult proof can be an indication that there is a near-counterexample in the wings. In the present case we get an immediate counterexample if we replace core 2D logic by the dtz fragment, as follows.

Example 12.2.4 The mood *Datisi(d,d,z)* is valid in E-free 2D logic, but not in 2D logic. We can see this easily by writing out the sentences. In E-free logic,

(12.2.3)
Some sometime-B is a C at all times.
Every sometime-B is an A at all times.
Therefore at time δ some C is an A.

This is clearly valid. But restore the E:

(12.2.4)
Some sometime-B is a C throughout its existence.
Every sometime-B is an A throughout its existence.
Therefore at time δ some C is an A.

This fails, because it could happen that nothing that is sometimes a B exists at time δ.

History 12.2.5 Ibn Sīnā was certainly aware of difficulties that arise in extending from core 2D logic to include the (z) sentences. For example at *Qiyās* iv.2, 193.9–17 he refers to problems which he says made Aristotle unwilling to consider syllogisms with premises that 'are true at a certain time'. He connects this with problems about using two-dimensional sentences where the time quantifier has sentence-wide scope; see for example *Qiyās* i.4, 30.16–31.5; iii.3, 155.1–8. In fact *Qiyās* contains a body of material in this area which has never been analysed—it would be hard to analyse it meaningfully without already having some understanding of two-dimensional logic.

12.3 The Avicenna-Johnston semantics

Our main target in this section is to demonstrate the mathematical fact that a semantics proposed by Spencer Johnston [64] for the divided modal syllogisms of Jean Buridan is a notational variant of a translation of Buridan's syllogisms into the dtz fragment of E-free two-dimensional logic. Implications of this fact are discussed in [48].

Johnston proposes a Kripke-style semantics for Buridan's divided modal syllogisms. He shows that all of Buridan's claims of validity or invalidity of moods are supported by this semantics. His Kripke structures have one difference from our Kripke frames of Section 9.2 above, namely that he includes an accessibility relation. But since this accessibility relation turns out to make all worlds accessible to all worlds, we can ignore it here. So I will use the notation of Section 9.2 rather than Johnston's notation.

Johnston identifies in Buridan the following eight forms of modal sentence, where A and B are distinct monadic relation symbols:

(12.3.1) $$A \overset{L}{a} B,\ A \overset{M}{a} B,\ A \overset{L}{e} B,\ A \overset{M}{e} B,\ A \overset{L}{i} B,\ A \overset{M}{i} B,\ A \overset{L}{o} B,\ A \overset{M}{o} B$$

We will call these *Buridan's divided modal sentences*. Johnson also identifies a further four forms of sentence:

(12.3.2) $$AaB,\ AeB,\ AiB,\ AoB$$

which we will call *Buridan's de inesse sentences*. We will describe the divided modal and the *de inesse* sentences together as *Buridan's sentences*. Johnston proposes necessary and sufficient conditions for a given sentence from either of these lists to be true at a given world in a Kripke Σ-frame, where Σ is a monadic relational signature containing the symbols A and B. His proposals are as follows.

Definition 12.3.1 Let Σ be a monadic relational signature and $K = (\mathcal{D}, \mathcal{W}, \mathcal{S}, \mathcal{E})$ a Kripke Σ-frame. For each relation symbol A in Σ and each world $\alpha \in \mathcal{W}$ we define:

$$\begin{aligned} &V^K(\alpha, A) = \mathcal{E}_\alpha \cap A^{\mathcal{S}_\alpha}. \\ &M^K(A) = \{d \in \mathcal{D} :\ \text{for some } \beta \in \mathcal{W}, d \in V^K(\beta, A)\}. \\ &L^K(A) = \{d \in \mathcal{D} :\ \text{for every } \beta \in \mathcal{W}, d \in V^K(\beta, A)\}. \end{aligned} \tag{12.3.3}$$

Then we define, for any world $\alpha \in \mathcal{W}$ and any two distinct relation symbols A and B in Σ,

$$\begin{array}{lcl}
K \models_\alpha A \overset{L}{a} B & \Leftrightarrow & M^K(A) \subseteq L^K(B) \text{ and } M^K(A) \neq \emptyset. \\
K \models_\alpha A \overset{M}{a} B & \Leftrightarrow & M^K(A) \subseteq M^K(B) \text{ and } M^K(A) \neq \emptyset. \\
K \models_\alpha A \overset{L}{e} B & \Leftrightarrow & M^K(A) \cap M^K(B) = \emptyset, \\
K \models_\alpha A \overset{M}{e} B & \Leftrightarrow & M^K(A) \cap L^K(B) = \emptyset. \\
K \models_\alpha A \overset{L}{i} B & \Leftrightarrow & M^K(A) \cap L^K(B) \neq \emptyset. \\
K \models_\alpha A \overset{M}{i} B & \Leftrightarrow & M^K(A) \cap M^K(B) \neq \emptyset. \\
K \models_\alpha A \overset{L}{o} B & \Leftrightarrow & M^K(A) \not\subseteq M^K(B) \text{ or } M^K(A) = \emptyset. \\
K \models_\alpha A \overset{M}{o} B & \Leftrightarrow & M^K(A) \not\subseteq M^K(B) \text{ or } M^K(A) = \emptyset. \\
K \models_\alpha AaB & \Leftrightarrow & V^K(\alpha, A) \subseteq V^K(\alpha, B) \text{ and } V^K(\alpha, A) \neq \emptyset. \\
K \models_\alpha AeB & \Leftrightarrow & V^K(\alpha, A) \cap V^K(\alpha, B) = \emptyset. \\
K \models_\alpha AiB & \Leftrightarrow & V^K(\alpha, A) \cap V^K(\alpha, B) \neq \emptyset. \\
K \models_\alpha AoB & \Leftrightarrow & V^K(\alpha, A) \not\subseteq V^K(\alpha, B) \text{ or } V^K(\alpha, A) = \emptyset.
\end{array}$$

In the first eight of these clauses, the α on the left is redundant and can be omitted.

By inspection of Definition 12.3.1, we see that Johnston assigns, to each of Buridan's sentences ϕ in signature Σ, a set-theoretic formula with variables for a Kripke frame K and a world α of K. We can adopt a language which has these set-theoretic formulas as its formulas; in this language any one of Johnston's formulas can be said to be satisfied or not satisfied at α in K. So the formula can be written as $\phi^{j1}(\xi)$, where ξ is a variable for α; thus

$$K \models \phi^{j1}[\alpha] \tag{12.3.4}$$

is a notational variant of what Johnston's set-theoretic formula says about K and α. His intention is that an entailment

$$\phi_1, \ldots, \phi_n \ \vdash\ \psi \tag{12.3.5}$$

between Buridan's sentences should hold if and only if

for every Kripke frame K and world α of K,

(12.3.6) $$K \models \phi_1^{j1}[\alpha],\ \ldots,\ K \models \phi_n^{j1}[\alpha] \ \Rightarrow\ K \models \psi^{j1}[\alpha].$$

Definition 12.3.2 The mapping $\phi \mapsto \phi^{j1}(\xi)$ defined through Definition 12.3.1 above will be called the *Johnston semantics*.

In Definitions 9.2.3 and 9.2.4 of Section 9.2 above, we defined an operation $\mathcal{K}$ taking two-dimensional structures to Kripke frames, and an operation $\mathcal{J}$ taking Kripke frames to two-dimensional structures. The operations are mutual inverses, so there is no information lost in either direction. Thus the Johnston semantics can be paraphrased to be expressed in terms of 2D structures $\mathcal{J}(K)$ rather than Kripke frames K.

Lemma 12.3.3 *Let K be a Kripke Σ-structure and N the $L(\Sigma(E))$-structure $\mathcal{J}(K)$. Then the notions of (12.3.3) above are definable from N as follows.*

$$\begin{array}{rcl} V^K(\alpha, A) & = & \{a : N \models (Ea\alpha \wedge Aa\alpha)\}. \\[1em] M^K(A) & = & \bigcup\{V^N(\beta, A) : \beta \text{ a time in } N\} \\ & = & \{a : N \models \exists\tau(Ea\tau \wedge Aa\tau)\}. \\[1em] L^K(A) & = & \bigcap\{V^N(\beta, A) : \beta \text{ a time in } N\} \\ & = & \{a : N \models \forall\tau(Ea\tau \wedge Aa\tau)\}. \end{array}$$

This induces translations of the Johnston formulas $\phi^{j1}(\xi)$ into formulas $\phi^{j2}(\xi)$ of $L(\Sigma(E))$ as in Figure 12.2 below, with the property that for every Kripke Σ-frame K and world α of K,

(12.3.7) $$K \models \phi^{j1}[\alpha] \ \Leftrightarrow\ \mathcal{J}(K) \models \phi^{j2}[\alpha].$$

The *proof* is straightforward by checking the definitions. □

Although each of Johnston's formulas is written with a world variable ξ, in the case of Buridan's modal sentences the variable is redundant and can be omitted. In a different sense, the E relation is redundant too, though its removal goes along very different lines from the results of Section 12.2 above. The key point is that we can incorporate E into the relations A etc., because they never occur separately; E occurs only in the context $(Ex\tau \wedge Ax\tau)$ for some A, and A occurs only in the context $(Ex\tau \wedge Ax\tau)$. So we can reinterpret each A as 'E and A', and drop the symbol E.

$$
\begin{array}{lcl}
(A \overset{L}{a} B)^{j2}(\xi) & \equiv & (\forall x(\exists\tau(Ex\tau \wedge Aa\tau) \to \forall\tau(Ex\tau \wedge Bx\tau)) \\
 & & \wedge\ \exists x\exists\tau(Ex\tau \wedge Ax\tau)). \\
(A \overset{M}{a} B)^{j2}(\xi) & \equiv & (\forall x(\exists\tau(Ex\tau \wedge Aa\tau) \to \exists\tau(Ex\tau \wedge Bx\tau)) \\
 & & \wedge\ \exists x\exists\tau(Ex\tau \wedge Ax\tau)). \\
(A \overset{L}{e} B)^{j2}(\xi) & \equiv & \forall x(\exists\tau(Ex\tau \wedge Aa\tau) \to \forall\tau\neg(Ex\tau \wedge Bx\tau)). \\
(A \overset{M}{e} B)^{j2}(\xi) & \equiv & \forall x(\exists\tau(Ex\tau \wedge Aa\tau) \to \exists\tau\neg(Ex\tau \wedge Bx\tau)). \\
(A \overset{L}{i} B)^{j2}(\xi) & \equiv & \exists x(\exists\tau(Ex\tau \wedge Aa\tau) \wedge \forall\tau(Ex\tau \wedge Bx\tau)). \\
(A \overset{M}{i} B)^{j2}(\xi) & \equiv & \exists x(\exists\tau(Ex\tau \wedge Aa\tau) \wedge \exists\tau(Ex\tau \wedge Bx\tau)). \\
(A \overset{L}{o} B)^{j2}(\xi) & \equiv & (\exists x(\exists\tau(Ex\tau \wedge Aa\tau) \wedge \forall\tau\neg(Ex\tau \wedge Bx\tau)) \\
 & & \vee\ \forall x\forall\tau\neg(Ex\tau \wedge Ax\tau)). \\
(A \overset{M}{o} B)^{j2}(\xi) & \equiv & (\exists x(\exists\tau(Ex\tau \wedge Aa\tau) \wedge \exists\tau\neg(Ex\tau \wedge Bx\tau)) \\
 & & \vee\ \forall x\forall\tau\neg(Ex\tau \wedge Ax\tau)). \\
(AaB)^{j2}(\xi) & \equiv & \forall x((Ex\xi \wedge Ax\xi) \to (Ex\xi \wedge Bx\xi)) \\
 & & \wedge\ \exists x(Ex\xi \wedge Ax\xi)). \\
(AeB)^{j2}(\xi) & \equiv & \forall x((Ex\xi \wedge Ax\xi) \to \neg(Ex\xi \wedge Bx\xi)). \\
(AiB)^{j2}(\xi) & \equiv & \exists x((Ex\xi \wedge Ax\xi) \wedge (Ex\xi \wedge Bx\xi)). \\
(AoB)^{j2}(\xi) & \equiv & (\exists x((Ex\xi \wedge Ax\xi) \wedge \neg(Ex\xi \wedge Bx\xi)) \\
 & & \vee\ \forall x\neg(Ex\xi \wedge Ax\xi)).
\end{array}
$$

Figure 12.2: Translation of the Johnston semantics

Definition 12.3.4 For each Buridan sentence ϕ we define $\phi^{j3}(\xi)$ to be the same formula as $\phi^{j2}(\xi)$ but with each conjunction $(Ex\tau \wedge \psi)$ replaced by ψ. In view of the next Theorem, we call the mapping $\phi \mapsto \phi^{j3}(\xi)$ the *Avicenna-Johnston semantics*.

Theorem 12.3.5 *By a standard model-theoretic device, the statement that an element a satisfies a formula $\phi(\xi)$ can be translated into a statement that $\phi(\delta)$ is true, where δ is a constant symbol whose interpretation is a. We adopt this device, using the same constant δ as in the (z) sentences of Figure 9.1. The sentences $\phi^{j3}(\delta)$, as ϕ ranges through the Buridan sentences, are exactly the sentences of the E-free dtz fragment. Moreover if T is a set of Buridan sentences and we write T^{j1} for the set $\{\phi^{j1}(\delta) : \phi \in T\}$ and T^{j3} for the set $\{\phi^{j3}(\delta) : \phi \in T\}$, then T^{j1} is consistent if and only if T^{j3} is consistent.* □

So the Avicenna-Johnston semantics sets up a bijection between the Buridan sentences and the sentences of the E-free dtz fragment. Johnston's semantic analysis of Buridan's sentences could equally well have been carried out in the E-free dtz fragment rather than in Kripke structures—the difference between the two is purely notational.

History 12.3.6 If Johnston had used ϕ^{j3} instead of ϕ^{j1}, then A in a formula ϕ^{j3} would stand for something different from A in ϕ^{j1}; in ϕ^{j3}, A would stand for 'is an A in the $j1$ sense *and is actual*'. One thing that this shows is that the mathematical formalism doesn't completely pin down the conceptual content. For example Johnston's semantics doesn't by itself throw any light on whether or not Buridan was operating under a meaning postulate that 'Everything is in all cases actual', because this meaning postulate would make no difference to which syllogisms count as valid according to the semantics. In the same vein, the fact that the valid syllogisms of the dt fragment exactly correspond to the divided modal syllogisms validated by Johnston's semantics doesn't entail that Ibn Sīnā's 'E' expresses the same concept as Johnston's 'actual'.

History 12.3.7 For the dt fragment in the first three figures, we noted in (10.3.1)–10.3.3) that Ibn Sīnā accepts the following moods:

First Figure: (d, d, d), (d, t, t), (t, d, d), (t, t, t).
Second Figure: (d, d, d), (d, t, d), (t, d, d).
Third Figure: (d, d, d), (d, t, t), (t, d, d), (t, t, t).

Read [16] pp. 41–44 includes a corresponding list of Buridan's verdicts on validity of syllogisms whose sentences are all divided modal. (Read lists major premise before minor; I have adjusted to Ibn Sīnā's convention of putting the minor first.) According to this list, Buridan accepts the following moods:

First Figure: (L, L, L), (L, M, M), (M, L, L), (M, L, M), (M, M, M).
Second Figure: (L, L, L), (L, M, L), (L, M, M), (M, L, L), (M, L, M).
Third Figure: (L, L, L), (L, M, M), (M, L, L), (M, L, M), (M, M, M).

A comparison shows that Ibn Sīnā and Buridan agree with each other (and with us—no surprise there) about which moods are productive, and about what the optimal conclusions are. The only difference in the lists is that Buridan lists some but not all of the non-optimal conclusions. Both Ibn Sīnā and Buridan give proofs, but a direct comparison of these is difficult because Buridan doesn't follow the axiomatic style that Ibn Sīnā uses.

We will see in Chapters BELOW that Skolem methods give a complete account of the productive moods and their conclusions in both the dtz fragment and the E-free dtz fragment. So by the route of Theorem 12.3.5 above, they tell us exactly what moods Buridan should have accepted according to Johnston's semantics. But this turns out to be a rather unpractical approach. The reason is that the Skolem method requires us to distinguish several refutation types according to their nominator sentences; for the dtz fragment in both versions, a dozen or so types are needed. The Orthogonality Principle, Corollary 10.2.4, fails for the dtz fragment both with and without E (cf. Exercise 12.7 below), and one reason for this is that the different refutation types give differing verdicts on the avicennan forms. The most practical approach seems to be the one followed by Johnston [64], namely to list the relevant moods systematically and then test each one for validity using bare hands. Alternatively a computer program to test the moods should be within reach of a Masters' project.

12.4 Exercises

12.1. Show that E is eliminable in the $atnd+z$ fragment. Hint: Consider the equivalences (9.1.6).

12.2. Prove: Let T be a core two-dimensional theory. If every connected component of T is two-dimensional consistent then T is two-dimensional consistent. [Exercise 3.11 doesn't serve as it stands, because it requires that the only relation symbols that occur in the sentences are their subject and predicate symbols, a condition not met when E appears.]

Solution. But the proof is not essentially different. We use Theorem 12.2.2(b) to restrict to the case where E is true everywhere in the models M_i involved, and then E is no barrier to combining these M_i into a single structure.

12.3. Theorem 12.2.2(b) left to right holds for all 2D sentences; show this. Also determine the entailments between pairs consisting of an E-free core and an E-free (z).

12.4. By Theorem 12.2.2, the E-free analogue of Theorem 9.3.1 holds. Show that Theorem 9.3.2 also transfers to the E-free case.

Solution. Entailments between pairs of (z) sentences are unaffected since every (z) sentence is equal to its E-free counterpart. For the rest: COMPLETE.

12.5. ON TWO INCOMPARABLE OPTIMAL CONCLUSIONS. Buridan offers *Cesare(d,z,t)* and *Cesare(d,z,z)*. (Read [16] p. 42.) Discuss. (See Exercise 9.3 above.)

Solution:

Every sometime-C is always a B, and there is a sometime-C;
no A-now is a B-now.
Therefore no sometime-C is always an A.
OR: Therefore no C-now is an A-now.

In the dtz fragment this translates to:

Every sometime-C is a B all the time it exists, and there is a sometime-C;
no A-now is a B-now.
Therefore no sometime-C is an A all the time it exists.
OR: Therefore no C-now is an A-now.

In this version the first conclusion fails, but the second still holds.

12.6. Verify:

(a) $(i\text{-}z)(D, C), (a\text{-}d)(C, B), (a\text{-}z)(B, A) \vdash (a\text{-}z)(C, A)$ is valid.

(b) $(i\text{-}z/E)(D, C), (a\text{-}d/E)(C, B), (a\text{-}z/E)(B, A) \vdash (a\text{-}z/E)(C, A)$ is valid.

(c) $(a\text{-}d)(C, B), (a\text{-}z)(B, A) \vdash (a\text{-}z)(C, A)$ is invalid. (This is *Barbara(d,z,z)*.)

(d) $(a\text{-}d/E)(C, B), (a\text{-}z/E)(B, A) \vdash (a\text{-}z/E)(C, A)$ is invalid. (This is E-free *Barbara(d,z,z)*.)

(e) $(a\text{-}d)(C, B), (e\text{-}z)(B, A) \vdash (e\text{-}z)(C, A)$ is valid. (This is *Celarent(d,z,z)*.)

(f) $(a\text{-}d/E)(C, B), (e\text{-}z/E)(B, A) \vdash (e\text{-}z/E)(C, A)$ is valid. (This is E-free *Celarent(d,z,z)*.)

Draw the subject-predicate digraph of the antilogism of (a).

Solution. (a) Suppose some object a is a D now and a C now. It follows that there is an A now (the augment of the conclusion). Let b be any object that is a C now. Then by the theory of E, b exists now, so by the second premise, b is a B now, and hence by the third premise, b is an A now.

(b) The proof is the same as for (a), except that there is no need to verify that b exists now.

(c) and (d): Nothing in the premises requires that any object is a C now, so we can't deduce the augment of the conclusion.

(e) Let b be any object that is a C now. Then by the theory of E, b exists now, so by the first premise, b is a B now. Therefore by the second premise, b is an A now.

(f) The proof is the same as for (e) except that there is no need to verify that b exists now.

12.7. Buridan claims ([16] p. 44) that

$(B \overset{M}{i} C), (BaA)$ is unproductive, but
$B \overset{M}{a} C), (BiA) \vdash (C \overset{M}{i} A$ is valid.

(a) Use the Avicenna-Johnston semantics to translate Buridan's claims into claims about premise-sequences in the E-free dtz fragment, and verify these claims.

(b) Show from (a), or otherwise, that the Orthogonality Principle, Corollary 10.2.4, fails for the E-free dtz fragment.

Solution. (a) The Avicenna-Johnston semantics translates the first claim to the claim that

$<(I\text{-}t/E)(B, A), (a\text{-}z/E)(B, A)> [C, A]$ is sterile.

The first premise says that some sometime-B is a sometime-A, and the second premise says that there is a B-now and every B-now is an A-now. But there could be a sometime-B that is not a B-now, so the two premises fail to connect and hence are sterile. OR COULD GIVE A PROOF OF STERILITY BY PSEUDOCONCLUSIONS.

The Avicenna-Johnston semantics translates the second claim to the claim that

$<(a\text{-}t/E)(B, C), (i\text{-}z/E)(B, A)> [C, A] \rhd (i\text{-}t/E)(C, A).$

The first premise says that some object is a sometime-B and every sometime-B is a sometime-C; the second says that some object a is a B-now and an A-now. It follows that a is a sometime-C and a sometime-A, so the premise-sequence is productive and entails $(i\text{-}t)(C, A)$. But EASY COUNTEREXAMPLES show that it doesn't entail any of $(a\text{-}t)(C, A)$, $(i\text{-}z)(C, A)$ or $(i\text{-}d)(C, A)$.

(b) In both premise-sequences the assertoric projections of the premises form a productive assertoric premise-pair (respectively *Datisi* and *Disamis*, and the avicennan forms agree in both cases, but the second premise-pair is productive and the first is not. Hence Orthogonality fails for the E-free dtz fragment.

(c) By Exercise 12.1, both premise-sequences lie in a fragment where the presence or absence of E is irrelevant to logical entailments, so the same examples show that Orthogonality fails also for the dtz fragment.

Chapter 13

Remainder of 2D logic

These chapters contain the mathematical theory to support the claims of Part III. The work is done, but I haven't included it in this draft because it needs some shaking down.

Part IV

Propositional

Chapter 14

Propositional logic

As for us, without seeking any help
we worked out all the syllogisms that
yield propositional compound goals,
and this without needing to reduce
them to predicative syllogisms;
and we enumerated all the
propositional compound propositions.
We invite those of our contemporaries
who claim to practise the art of logic
to do likewise,
and to compare all of their findings
to all of ours.
Ibn Sīnā, *Masā'il* 103.12–14

14.1 Propositional logic?

In Books v–viii of *Qiyās*, Ibn Sīnā moves away from the temporal and modal sentences that he has studied in Books ii–iv, and discusses what modern writers have called 'Propositional logic' (as in the title of Shehaby's translation [93]). That's a convenient name, but it might be only partially correct. The sentences under consideration are Arabic sentences formed by combining other Arabic sentences, rather than by combining terms as with the assertoric sentences. But the forms of composition are not necessarily ones that we would recognise today as propositional. For example many of them translate into English as

(14.1.1) Whenever p then q.

so that the sentence as a whole is in the scope of a universal quantifier over times.

History 14.1.1 When Ibn Sīnā illustrates a point with an example like

> At a time when humans exist, two is even too.

(*Qiyās* 235.7f in Book v), it's perverse not to think of examples that he used earlier in *Qiyās* with a wide-scope quantification over time, such as at *Qiyās* 193.12:

> There is nothing to prevent its being true at some time that every moving thing is a human.

At *Qiyās* 143.13 Ibn Sīnā says that he knows of no followers of Aristotle who adopted wide-scope time quantification; he implies that one should avoid it in expositions of Peripatetic modal logic. But Books v–viii of *Qiyās* are not an exposition of modal logic, and evidently Ibn Sīnā feels free to be more experimental here.

So at least some of the contents of *Qiyās* v–viii is a development of his two-dimensional logic in a new direction. We should aim to read *Qiyās* as a unity, bound together by the two-dimensional logic of Book i.

But certainly there are other strands. Just as Books ii-iv owe much of their content to Aristotle's alethic modal logic, so Books v–viii contain material that goes back to earlier Peripatetic ideas about 'conditional' (*šarṭī*) propositions. In the Arabic texts our best source for this material before Ibn Sīnā is the *Categories* [4] of Al-Fārābī. What follows is a summary of some relevant parts of Al-Fārābī's presentation there. I use Dunlop's translations [23] pp. 50–52.

Al-Fārābī describes the topic as *al-mutalāzimāni*, which probably means 'pairs of propositions that are inferentially related'; later he introduces the notion *al-muta*^{c}*ānidāni*, which must mean 'pairs of propositions that are in conflict with each other'. He gives several examples. Each of them consists of a single sentence with two subclauses, and the subclauses represent the pair of propositions that have the relationship in question.

For the 'inferentially related' he offers

(14.1.2)

(i) If Zaid comes, ^{c}Amr departs.

(ii) When Sirius rises in the morning, the heat will be severe and the rains will cease.

(iii) When man exists, animal exists necessarily.

(iv) If this number is even, it is not odd.

Dunlop translates (iii) literally, but Al-Fārābī may have in mind a sentence of the form 'If X is human then X is an animal'. A reference to 'the rising of the sun and its being day' ([4] 127.13, missing in Dunlop) must indicate the well-known example

(14.1.3) (v) When (or if) the sun is up, it's day.

For the 'conflicting' pairs Al-Fārābī offers no examples, but his discussion of (iv) above implies that it paraphrases a conflicting pair as in

(14.1.4) (vi) Either this number is even or this number is odd.

Al-Fārābī tells us that the sentences expressing the inferential relation are *muttaṣil*, literally 'connected', and the sentences expressing conflict are *munfaṣil*, literally 'separated'.

Al-Fārābī makes no suggestion that all the *muttaṣil* sentences have the same logical features, or all the *munfaṣil*. On the contrary, his main efforts go into cataloguing the different types. He has two main principles of classification. One is by the inferences that these sentences enter into. For example he believes that all *muttaṣil* sentences 'When (or if) p then q' support the inference

(14.1.5)
When p then q. p. Therefore q.
When p then q. Not q. Therefore not p.

But only some of them support the inferences in the other direction:

(14.1.6)
When p then q. q. Therefore p.
When p then q. Not p. Therefore not q.

He describes the ones that are symmetrical (*yatakāfa'āni*) in this way as 'complete' (*tāmm*), and the others as 'not complete'. Likewise all *munfaṣil* sentences 'Either p or q' support the inferences

(14.1.7) Either p or q. p. Therefore not q.
Either p or q. q. Therefore not p,

But only some of them support the opposite inferences

(14.1.8) Either p or q. Not p. Therefore q.
Either p or q. Not q. Therefore p.

Again he distinguishes the ones that support the opposite inferences as 'complete' (*tāmm*).

The second principle of classification is in terms of the source of the inference. Here he is a little vague, but broadly he distinguishes those cases where 'If p then q' is true 'essentially' (*bil ḏāt*), for example where p being true *causes* q to be true, and those cases where in some sense the relation holds 'accidentally' (*bil ᶜaraḍ*).

He also distinguishes those cases where the relationship holds 'for the most part' (*ᶜalā l-'akṯar*) from those where it holds always. Curiously he allows that some 'essential' relations hold only for the most part.

Al-Fārābī's whole discussion of these sentences is rather rough-hewn. When the *muttaṣil* sentence means 'whenever' rather than 'if', (14.1.5) could stand for any of several inferences, for example:

(14.1.9)

- Whenever the sun is up it's day. The sun is up now. Therefore it's day now.
- Whenever the sun is up it's day. Sometimes the sun is up. Therefore sometimes it's day.
- Whenever the sun is up it's day. The sun is often up. Therefore it's often day.
- Whenever the sun is up it's day. The sun is always up. Therefore it's always day. (Valid!)

Al-Fārābī shows no inclination to distinguish between these inferences. In this sense he is blind to the fact that some of his *muttaṣil* sentences contain a universal quantifier over times.

Al-Fārābī has further remarks on these sentences in his *Syllogism* [5]. Ostensibly he develops their proof theory. But this expectation is disappointed; there is no further development. The proof rules are just those already given in [4], except that in the *munfaṣil* sentences he now allows a disjunction of more than two clauses. This seems to me to indicate that we shouldn't expect anything new about formal propositional logic in his missing longer commentary on the *Qiyās*, if it ever comes to light.

Al-Fārābī explains in *Qiyās* one curious piece of terminology that he uses for inferences like (14.1.5) or (14.1.6). In Arabic grammatical terminology an 'exception' (*istiṯnā'*) is an expression like

(14.1.10)
All my brothers came except Zayd.

The sense is to exclude Zayd from the list of my brothers who came, and limit the list to 'the remainder' (*al-bāqī*). The word translated as 'except', usually *'illā*, is called a 'particle of exception' (*ḥarfu l-istiṯnā'*). Apparently Al-Fārābī saw an analogy with inferences like

(14.1.11)
This body is animal, vegetable or mineral.
Except that this body is not mineral.
Therefore this body is animal or vegetable.

So he called this kind of inference an *istiṯnā'*, and he described the clause 'This body is mineral' as the 'excepted' (*mustaṭnā*). The analogy is thin; in the linguistic case the original statement is corrected by the exception, but the *mustaṭnā* doesn't in any sense correct the *munfaṣil* premise. For *muttaṣil* there is no convincing analogy at all.

History 14.1.2 Ibn Sīnā took over the terminology of *istiṯnā'*. Fortunately there is a reading of it that saves us the embarrassment of Al-Fārābī's weak analogy, and may well have been how Ibn Sīnā himself saw it (though we have no way of checking this). Etymologically *istiṯnā'* can be read as 'taking twice', so we can speak of inferences like (14.1.5) or (14.1.6) as *duplicative* because a part of the *muttaṣil* or *munfaṣil* sentence is 'duplicated' by the second premise. In Ibn Sīnā's usual terminology the sentence is duplicated either as *ᶜayn* 'the same' or as *naqīḍ* 'contradictory negation', and likewise the conclusion.

Kamran Karimullah's recent excellent survey [66] of Al-Fārābī's views on 'conditionals' studies very different issues from the ones considered above. But where we do converge on the same questions, we seem to give compatible answers too. Thus 'Strictly speaking Alfarabi cannot be said to propound a proper logical doctrine of conditional propositions' (p. 212), '… for Alfarabi, there is no single "correct" reading of conditionals of the form 'if P, then q'. Rather, the sort of implication expressed by conditional sentences, divided according to the weakness of strength of the connection between the antecedent and consequent, depends crucially on the argumentative context in which the conditional is deployed' (p. 242f).

We need translations for *muttaṣil* and *munfaṣil*. This is a fairly standard contrasting pair in Arabic, and Shehaby's [93] 'connective' and 'separative' is a very acceptable translation in general. But it gives no indication of the logical content. Rescher [88], followed by Ahmed [2], went straight to logic and used 'conjunctive' and 'disjunctive'. 'Disjunctive' is fine for Al-Fārābī's *munfaṣil* sentences, but we will see that it doesn't work for Ibn Sīnā's use of the term. 'Conjunctive' for Al-Fārābī's *muttaṣil* is wrong; these sentences are not conjunctions at all. Ibn Sīnā will introduce (o) *muttaṣil* sentences which do contain a conjunction, but in fact his (i) *munfaṣil* sentences contain a conjunction too. I doubt that there are translations that work for all purposes. I will use *meet-like* for *muttaṣil* and *difference-like* for *munfaṣil*. The logical terms meet/difference are a good match for the Arabic *waṣl*/*faṣl* (*differentia* comes into Arabic as *faṣl*), and the '-like' establishes a safe distance between the two Arabic words and any precise logical meaning.

14.2 *Qiyās* vi.1: meet-like syllogisms

In *Qiyās* vi.1 Ibn Sīnā extends the class of *muttaṣil* (meet-like) sentences so that they fall into four forms like the assertoric sentences: (a), (e), (i) and (o). As with the assertorics, these letters are not Ibn Sīnā's own names; rather he describes the (a) sentences as 'universal affirmative', and likewise with the other three forms.

Definition 14.2.1 The meet-like propositions thus form a subject-predicate logic in the sense of Definition 3.1.5, except that they have clauses in place of terms. We write the four forms as (a, mt), (e, mt), (i, mt) and (o, mt) (where mt is for *muttaṣil*). Thus $(a, mt)(p, q)$ is the meet-like (a) sentence with first clause p and second clause q. The first clause is also called the *antecedent* (*muqaddam*) and the second is called the *consequent* (*tālī*). But one should be wary of these two expressions, because the relationship between the first and second clauses may not be what you would expect of an antecedent and a consequent.

History 14.2.2 We are writing p, q for the clauses. Ibn Sīnā almost never uses variable letters for propositions (*Qiyās* 544.18f is a rare counterexample). Instead he writes the clauses in a short assertoric style, for example 'C is D'. Should we assume that he intends all the clauses to be assertoric? In some places he explicitly says he does, for example when he talks of 'difference-like propositions formed from predicative propositions' (*Qiyās* 361.7). But often the nature of the clauses seems to us irrelevant to the arguments that he is studying, and *Qiyās* vi.1 is a case in point. He says at *Qiyās* 296.2 that 'A is a B' can stand for any predicative proposition; but does he mean permissively that it doesn't have to be literally 'A is B', or does he mean restrictively that it must be predicative? An instance that supports the permissive reading is at *Qiyās* 301.13, where he specifies a particular situation—or perhaps a class of situations—for an ecthetic argument. He specifies it by a sentence 'A is B'. But the context gives not the slightest reason to believe that the situation is describable by a predicative sentence, so at least here the 'A is B' seems to be a pure propositional variable. But one should note that Ibn Sīnā sometimes uses the forms of the clauses to control the sentence form; for example in *Qiyās* vi.2 he lets the fact that the clauses in a disjunction are affirmative be an indication that the disjunction is exclusive.

Ibn Sīnā explains that each of his meet-like sentences has two clauses, which can be referred to as 'terms' (*Qiyās* 295.7). So a pair of meet-like sentences with one term in common will fall into one of the predicative 'figures' (*Qiyās* 295.7). Disregarding the fourth figure as we did with the predicative syllogisms, we can ask whether a pair of meet-like sentences in a figure yields a consequence of the appropriate form, just as with assertoric syllogisms. Ibn Sīnā answers this question positively, and sets the facts out in detail.

His account of the valid moods for these sentences is point for point the same as the account he gave for assertoric syllogisms. Not only does he count exactly the analogous moods as valid, but he gives the same justifications in terms of conversion, contraposition and so on. This is documented in detail in Appendix A below, which catalogues the moods and justifications given for the asserorics in *Najāt*, *Qiyās* (ii for assertoric, vi for meet-like), *Dānešnāmeh* and *Išārāt*. At *Qiyās* 295.10f he gives the same general productivity conditions as he gave for the assertorics. The conclusion, which Ibn Sīnā leaves the reader to draw, is that the syllogistic logic of meet-like sentences obeys *exactly* the same rules as those of assertoric syllogisms.

This is a startling piece of work. I make a sequence of numbered comments on it.

1. One effect of extending the class of meet-like sentences to include all four types (a), (e), (i) and (o) is that the class is now closed under contradictory negation. This severs the link between meet-like sentences and the notion of a 'conditional' sentence. The contradictory negation of a sentence of the form 'If ... then' is not of this form, or of any form like it.

This raises a question about how we should understand the name *šarṭī*, literally 'conditional', when it is used for the sentences studied in propositional logic. Ibn Sīnā is ahead of us on this. In *Mašriqiyyūn* 61.7–12 he points out first that 'conditional' is more appropriate for the meet-like sentences than for the difference-like ones (and here he must mean those of the Farabian logic of the previous section rather than his own expanded meet-likes). He suggests that *šarṭī* should be understood as referring to sentences containing subclauses that don't have a truth value. The suggestion is not straightforward, but here is not the right place to discuss it. My guess is that he chiefly means sentences containing subclauses that are not understood as being asserted when the sentence is asserted. I count this as partial support for translating *šarṭī* as 'propositional compound', which I will do.

2. For Ibn Sīnā, logic handles sentences in a natural language. So he needs to have Arabic sentence forms that express his four kinds of meet-like sentence. The four forms that he generally uses are as follows:

(14.2.1)

(a, mt) *kullamā kāna* p *fa* q.

(e, mt) *laysa albatta 'iḏā kāna* p *fa* q.

(i, mt) *qad yakūnu 'iḏā kāna* p *fa* q.

(o, mt) *laysa kullamā kāna* p *fa* q.

These formulations have some curious features that Arabic scholars will want to take note of. The form for (o) is simply the form for (a) with an initial sentence negation *laysa*. But (e) and (i) have a different pattern. The *laysa* at the beginning of (e) can't have scope the whole sentence; in fact the only item negated is the final q, so *laysa* has been raised a long way from its logical position.

History 14.2.3 In fact I haven't yet found any examples of the phrase *laysa al-batta* ('not ever') in the literature outside *Qiyās*. This strongly suggests that Ibn Sīnā is now abandoning the position, which had guided him in his predicative logic, that logicians should study the forms of expression actually used in the relevant scientific communities. He indicates something of the sort in *Išārāt* immediately before he introduces the form of meet-like logic that runs parallel to the assertorics: 'We will mention some of these [propositional] syllogisms, but we will avoid those that are not close to our natural [ways of thinking]' (*Išārāt* 157.3, /432d/); cf. *Najāt* 84.12 '[these forms of argument are] very remote from our natures'. In effect he adopts a form like that for (e) above as a piece of technical vocabulary.

History 14.2.4 A curious feature of the (i) form is that Ibn Sīnā manages to introduce 'when' (*'iḏā*) in such a way that its logical context is not 'When p then q'. In English we can render his phrase as 'It can be, when p, that q'—which means that it can be the case that p and q. Readers interested in the formal semantics of this usage might note the analysis of a similar sentence in Schubert and Pelletier [90] p. 218. Evidently Ibn Sīnā wants to keep 'when' in the forms even where it is not logically appropriate. Did he really think he had to sneak in a 'when' to keep the Peripatetics happy, or did he take this as an entertaining mental challenge?

3. A modern logician might aim to explain the fact that meet-like logic follows the rules of assertoric logic by reducing one to the other. There is a natural way to do this, as George Boole taught us ([13] pp. 162–164). Paraphrase the four meet-like sentences as follows:

(14.2.2)

(a, mt) Every time when p holds is a time when q holds.

(e, mt) No time when p holds is a time when q holds.

(i, mt) Some time when p holds is a time when q holds.

(o, mt) Not every time when p holds is a time when q holds.

Under these paraphrases, the four sentences are simply assertoric sentences, so they automatically follow the rules of validity for assertorics. Strictly it still should be checked that they follow no other rules.

Unlike Boole, Ibn Sīnā shows no sign whatever of going down this road. His claim is always that the meet-like sentences are 'in analogy with' (*ᶜalā qiyās*, *Najāt* 83.7,9) the assertorics, or that they satisfy 'the same' (*wāḥid*, *Najāt* 83.7; *miṯl*, *Qiyās* 296.1) conditions, or that they 'take the same form' (*ᶜalā hay'a*, *Qiyās* 295.7), or that they are 'as the predicatives' (*ka*, *Išārāt* 157.6, /432d/). It will not do to say that Ibn Sīnā lacked the knowhow or the inclination to spell out a reduction, because we will see in the next section that he *precisely* reduces his difference-like logic to his meet-like logic, by spelling out how to paraphrase the former into the latter.

History 14.2.5 Ibn Sīnā's name for reduction by paraphrase is *rujūᶜ*. Thus he talks of reduction of third figure syllogisms to first figure by conversion, *Qiyās* 111.1, 302.12; reduction of contingency statements from affirmative to negative or vice versa, *Qiyās* 174.16; reduction of difference-like propositions to meet-like, *Qiyās* 305.10; reduction of affirmative to negative meet-like propositions, or vice versa, by metathesis, *Qiyās* 366.2, 508.10; reduction of syllogisms by absurdity to direct syllogisms, *Qiyās* 451.14f. This is only a sample of the places where he uses the notion of *rujūᶜ*. But he never applies it to 'reducing' meet-like syllogisms to assertoric ones.

ALSO NOTE REFERENCES ON *radd*. THERE ARE ONE IN QIYAS AND ONE IN MUKHTASAR THAT COULD BE RELEVANT.

In short, Ibn Sīnā's position with meet-like syllogisms is that they obey the same rules (*qawānīn*) as the assertoric ones. The *qawānīn* in question are those that he spells out in *Qiyās* vi.1, for example the conditions of productivity and the specific rules for each mood. This is a more radical position than the claim that meet-like logic 'reduces to' assertoric logic. The radical feature is that Ibn Sīnā offers, presumably for the first time in logic, two different forms of logic that obey the same formalism. Of course all logic from Aristotle onwards has used variables that are interpreted differently in different arguments. The radical new feature here is that Ibn Sīnā interprets the *logical* notions differently; for example he switches from an object quantification expressed by 'every' to a time quantifier expressed by 'whenever'. Here we reach the point that we anticipated in Chapter ch:11 above.

How could Ibn Sīnā come to such a position, if not by reducing the meet-like sentences to assertoric ones? He doesn't tell us. But earlier I pointed to a likely route: having extended assertoric logic to two-dimensional logic, he noticed that the time quantification could be used to play the same role that the object quantification plays in the assertorics. So he noticed a formal symmetry rather than a reduction.

Another possible source is Peripatetic discussion of possibility sentences (*ᶜalā sabīli l-'imkān*) like

(14.2.3) If this is an animal then it can be human.

(*Qiyās* 397.16.) Ibn Sīnā could have experimented with ways of making sense of such a sentence, as for example 'There is a possible situation in which this is an animal and this is human'. This could have reinforced his rearrangement of two-dimensional sentences by giving him some relevant raw material.

> **History 14.2.6** In this context, *Qiyās* 398.3 in viii.1 is one of the few places in *Qiyās* where Ibn Sīnā mentions the notion of substance, *jawhar*. The notion is mentioned as defining one kind of possibility ('possibility in terms of the mind and not in terms of the facts', *'imkān bi-ḥasabi l-ḏihni lā bi-ḥasabi l-'amr*). There is a similar use of *jawhar* at *Qiyās* 235.8, listing types of relation of 'following' (*luzūm*). At *Qiyās* 22.3 *jawhar* is mentioned as an alternative name for essence (*ḏāt*); this seems to be a Farabian usage. All the other references to *jawhar* in *Qiyās* are in connection with the matters of particular examples: 61.5–10, 106.7, 308.9–309.3, 317.2f, 450.6, 469.7–471.13, 508.11f, 544.5f.

4. The validity of the moods *Darapti* and *Felapton* rests on the fact that (a) sentences carry an existential augment. In fact for both of these moods Ibn Sīnā offers a justification that depends on converting $(a, mt)(p, q)$ to $(i, mt)(q, p)$ (*Qiyās* 302.12, 303.1), a conversion that doesn't work without an existential augment on the (a) sentence. So in *Qiyās* vi.1 Ibn Sīnā is supposing that (a, mt) sentences carry the analogue of the existential augment; in other words, $(a, mt)(p, q)$ is taken as false unless there is a time at which p is true. (See Movahed [79] for another angle; Movahed argues that this supposition is a mistake on Ibn Sīnā's part.)

Readers of Ibn Sīnā's propositional logic easily get the impression that Ibn Sīnā is careless about existential augments, and that he has no principled rule for including or excluding them. But arguably it's unwise to assume from the start that he has no rule. He does refer to the existential augment on (a, mt) sentences in several places, so it is certainly not below the level of his radar. Thus for example at *Qiyās* 337.6 he spells out the existential augment ('the condition that the first clause is one that can be true'). At *Qiyās* 368.14–17 he says that two (a, mt) sentences with contradictory second clauses (and, presumably, the same first clause) are pairwise contrary 'potentially' (*fī quwwa*, Shehaby [93] p. 168 translates 'can be considered as')—but they are only contrary with the augment. I read *Qiyās* 263.7 as saying explicitly that $(a, mt)(p, q)$ commits us to there being some time at which p holds; but I note that Shehaby [93] p. 62 reads the passage differently.

5. Intertwined with the discussion of the four sentence forms, this section contains some remarks about the classification of meet-like sentences as *ittifāqī* or *luzūmī* (e.g. *Qiyās* 298.2). This looks like a relative of Al-Fārābī's distinction between propositional compounds that hold 'accidentally' and those that hold 'essentially', or maybe an attempt to give some formal content to that distinction. The word *luzūmī* is from *luzūm*, 'entailment' or 'following'. It will be wise to hold back on translating the word *ittifāqī* until we see what Ibn Sīnā does with it.

In *Qiyās* 297.8–14 Ibn Sīnā discusses meet-like sentences that are true for reasons of *ittibā'* as opposed to *luzūm*. He clearly means *ittifāq* (e.g. at *Qiyās* 298.2,10), and possibly the text should be amended. The sentences that he describes as true by *ittibā'* are, if I read him right, meet-like sentences of the form (a) that are taken to be true because their second clause is true. He observes that if ϕ is a sentence with this property, and we use ϕ as a premise in a syllogism whose conclusion is the second clause of ϕ, then the syllogism has added no new information and is redundant. This comment makes logical sense if we suppose that he is describing a sequence of steps in an argument, where two consecutive steps are

$$\text{(14.2.4)}\qquad \dfrac{\text{Always } p \qquad \dfrac{\text{Always } q}{\text{Always if } p \text{ then } q}\ \textit{Step one}}{\text{Always } q}\ \textit{Step two}$$

He is saying that these two steps can be left out without loss of information. It's worth remarking that this elimination has the same form as the 'reductions' that Dag Prawitz uses for normalising natural deduction proofs (see his [81] pp. 35–38, particularly the rule at the top of page 37). But of course Ibn Sīnā has no apparatus of formal derivations; he is merely remarking that the combination of these two steps is pointless.

History 14.2.7 In this passage it seems that Ibn Sīnā understands an *ittifāqī* sentence $(a, mt)(p, q)$ to be one which is taken to be true on the basis that 'Always q' is true. If so, then describing an (a, mt) sentence as *ittifāqī* has nothing to do with the meaning of the sentence; it refers to the way the sentence has been introduced in an argument. When he moves on to (e, mt) sentences, he speaks of these as 'denying the *ittifāq*' (*sālibata l-muwāfaqa*, *Qiyās* 299.4); though the facts are not completely clear, he can be read as saying that a sentence $(e, mt)(p, q)$ 'denies the *ittifāq*' if it is taken as true on the grounds that q is always false. If these readings are correct, then *ittifāq* has to mean something like 'agreement with the assumed facts'. This is plausible, because 'agreement' is one of the common meanings of *ittifāq* in Arabic. But *ittifāq* can also mean accident or coincidence, and on this basis Ibn Sīnā's *ittifāqī* has often been taken as referring to a 'chance connection' between the two clauses. One should look through Ibn Sīnā's concrete examples. To my eye they overwhelmingly support the account given above; 'chance' is irrelevant to them. The one example difficult for the account above is at *Qiyās* 405.14; I think this is an accident of terminology, and for this particular example Ibn Sīnā is saying that a certain sentence is true *because of an agreement (ittifāq) between the two people mentioned*. The example works no better for the 'chance' interpretation than any of his other examples do.

There are also references in this section to meet-like sentences that are *luzūmī*. As far as I can see, the notion has no formal content at all; the sentences that are *luzūmī* don't obey or fail to obey any rules as a result of being *luzūmī*. Possibly Ibn Sīnā is using the expression just to mean 'not *ittifāqī*'. At *Qiyās* 306.6 (in vi.2) he says that a conclusion is *luzūmī* 'because' it follows validly from the premises.

14.3 vi.2: mixed meet-like and difference-like syllogisms

In the next section of *Qiyās*, section vi.2, Ibn Sīnā moves on from meet-like syllogisms; now he considers syllogisms where one of the premises is meet-like and the other is difference-like. Just as with the meet-like sentences in section vi.1, he expands the class of difference-like sentences so that each of them has one of the forms (a), (e), (i) and (o), and the class is closed under taking contradictory negations.

Definition 14.3.1 Here we have another subject-predicate logic. We write the four sentence forms of this logic as (a, mn), (e, mn), (i, mn) and (o, mn) (where mn is for *munfaṣil*).

Again Ibn Sīnā has to introduce Arabic expressions to represent these four forms. His preferred expressions are as follows:

(14.3.1)
(a, mn) *dā'iman 'immā 'an yakūna* p *'aw* q.

(e, mn) *laysa albatta 'immā* p *wa-'immā* q.

(i, mn) *qad yakūnu 'immā 'an yakūna* p *'aw* q.

(o, mn) *laysa dā'iman 'immā* p *wa-'immā* q.

Here (a) reads 'Always either p or q' and (o) reads 'It is not the case that always either p or q'. I won't offer translations of the Arabic (e) and (i) sentences because I don't believe they make enough sense in Arabic to justify any translation. If anybody tells you otherwise, get him or her to tell you exactly what they do mean and then compare with the meanings that we extract below from Ibn Sīnā's text.

History 14.3.2 The point made in History 14.2.3 above holds here with even greater force. The expressions for (e, mn) and (i, mn) can only be regarded as the equivalent of a pair of technical symbols. I thank Amirouche Moktefi for raising this issue.

Ibn Sīnā justifies his mixed meet-like/difference-like syllogisms by translating the difference-like premises into meet-like ones, drawing a conclusion in meet-like logic, and then offering a translation of the conclusion back into a difference-like sentence. As a result of this approach, he gives us in *Qiyās* vi.2 a large number of translations of difference-like sentences into meet-like ones, and a large number of translations in the opposite direction too. These provide us with plenty of information for inferring what he takes the difference-like forms to mean. Besides the translations, there are other clues that we can use. For example it's a virtual certainty that

(14.3.2) (a, mn) and (o, mn) are contradictory negations of each other, and (e, mn) and (i, mn) are contradictory negations of each other.

(In fact Ibn Sīnā tells us at *Qiyās* 381.7, 381.16 and 382.15 that (e, mn) is the contradictory negation of (i, mn).) Also, given the origins of the difference-like sentences in disjunctions, it would be very surprising if we didn't have:

(14.3.3) (a, mn) is symmetric in the sense that $(a, mn)(p, q)$ and $(a, mn)(q, p)$ are logically equivalent.

In these translations, Ibn Sīnā frequently negates either the first or the second clause.

Definition 14.3.3 We write (a, mn) for the (a) form of difference-like sentence. With first clause p and second clause q, this form yields the sentence $(a, mn)(p, q)$. We write $\bar{p}$ for the negation of the clause p; negation of clauses as opposed to whole sentences is called *metathetic negation*. So for example $(a, mn)(\bar{p}, \bar{q})$ means the difference-like sentence of form (a) whose first clause is the metathetic negation of p and whose second clause is the metathetic negation of q. Similarly with the forms (e, mn), (i, mn) and (o, mn).

History 14.3.4 In ^{c}*Ibāra* ii.1 Ibn Sīnā distinguishes two kinds of 'negation' (*salb*) of an assertoric sentence of the form $(a)(B, A)$, neither of them being contradictory negation. The first kind of negation is to replace (a) by its negative opposite number, which is (e), so that we get $(e)(B, A)$. The second is to keep the affirmative form but add a negation as part of the predicate A: $(a)(B, \text{not-}A)$. He calls this second kind of negation *metathesis* (^{c}*udūl*) (^{c}*Ibāra* 82.4), as opposed to the first which is 'simple negation' (*sālib basīṭ*). The two forms $(e)(B, A)$ and $(a)(B, \text{not-}A)$ are not quite equivalent, because only the second of the two counts as an affirmative proposition. So the existential augment applies to the second form and not to the first; if there are no Bs, then the first proposition is true and the second is false (^{c}*Ibāra* 82.14f). The analogy between assertorics and meet-like propositions allows us to carry over this distinction between simple negation and metathesis to propositional compounds, at least when the form is clearly stated. (In practice it is not so easy, unless the Arabic sentence is phrased using one of Ibn Sīnā's standard expressions for the sentence forms.)

Besides the translations between meet-like and difference-like, Ibn Sīnā also offers a number of alternative meet-like forms. Prima facie these should all be covered by the results in the previous section. But there will be no harm if we check them out. Here is a list of the translations that he offers between meet-like propositions, with references to the text of *Qiyās*:

				Qiyās	Note
1.	$(a, mt)(p, q)$	$\to$	$(i, mt)(q, p)$	311.13	
2.	$(a, mt)(p, q)$	$\to$	$(e, mt)(p, \bar{q})$	318.6	
3.	$(a, mt)(p, \bar{q})$	$\to$	$(e, mt)(p, q)$	308.3, 317.15	
4.	$(e, mt)(p, q)$	$\to$	$(a, mt)(p, \bar{q})$	310.16, 311.16	
5.	$(e, mt)(p, \bar{q})$	$\to$	$(a, mt)(p, q)$	310.10	
6.	$(e, mt)(\bar{p}, q)$	$\to$	$(a, mt)(p, q)$	312.14	(1)
7.	$(i, mt)(p, q)$	$\to$	$(i, mt)(q, p)$	317.10	
8.	$(i, mt)(p, \bar{q})$	$\to$	$(i, mt)(\bar{q}, p)$	316.16	
9.	$(i, mt)(\bar{p}, \bar{q})$	$\to$	$(o, mt)(\bar{p}, q)$	310.14, 310.18	
10.	$(o, mt)(p, q)$	$\to$	$(i, mt)(p, \bar{q})$	316.15?	
11.	$(o, mt)(p, \bar{q})$	$\to$	$(i, mt)(p, q)$	317.9	

Note (1): Correct the text so that $(e, mt)(\bar{p}, q)$ becomes $(e, mt)(p, \bar{q})$.

Figure 14.1: Translations between meet-likes

We comment on these translations. Translation 1, from a universal to an existential sentence, is not invertible, but all of the remaining translations are from universal to universal or from existential to existential. Translations 7 and 8 are within the same aristotelian form and express that (i) is symmetric.

The remaining translations relate an affirmative aristotelian form and the corresponding negative form, or vice versa. The recipe is to add or remove a metathetic negation on the second clause. Most of them are unproblematic. But we note that 4–6 take an (e) form to an (a) form, which implies that Ibn Sīnā is not including the existential augment with the (a) form. In case 4 it could be argued that the metathetic negation interferes with the augment, though this would go against Ibn Sīnā's clear statement in ^{c}*Ibāra* that an (a) sentence with metathetic negation of the second term counts as affirmative. This excuse is not available in case 5, where neither of the clauses is negated. In case 5 Ibn Sīnā has simply abandoned the existential augment. The same point applies in cases 10 and 11, where he deduces an (i) from an (o), discarding any universal augment on the (o) sentence; and again in case 10 the (o) sentence is pure negative with no metathesis.

There remains case 6, which is clearly an error in the text.

In sum: apart from the translation from (a) to (i), all these translations are paraphrases from a sentence to a logically equivalent sentence. The equivalences require that the sentences carry no augments, either existential or universal.

We turn to the translations from difference-like to meet-like.

				Qiyās	Notes
1.	$(a, mn)(p, q)$	$\rightarrow$	$(a, mt)(\bar{p}, q)$	313.10, 314.3, 317.5	
2.	$(a, mn)(p, q)$	$\rightarrow$	$(a, mt)(\bar{q}, p)$	310.2	
3.	$(a, mn)(p, \bar{q})$	$\rightarrow$	$(a, mt)(\bar{p}, \bar{q})$	308.1, 315.11, 316.2	
4.	$(a, mn)(p, \bar{q})$	$\rightarrow$	$(a, mt)(q, p)$	311.12	
5.	$(a, mn)(\bar{p}, q)$	$\rightarrow$	$(a, mt)(p, q)$	314.12, 315.4, 318.5	
6.	$(a, mn)(\bar{p}, q)$	$\rightarrow$	$(a, mt)(\bar{q}, \bar{p})$	312.10	
7.	$(a, mn)(\bar{p}, q)$	$\rightarrow$	$(a, mt)(p, \bar{q})$	317.14	(1)
8.	$(a, mn)(p, q)$	$\rightarrow$	$(a, mt)(p, \bar{q})$	309.13	(2)
9.	$(a, mn)(p, q)$	$\rightarrow$	$(a, mt)(q, \bar{p})$	305.10, 313.12	(2)
10.	$(e, mn)(p, q)$	$\rightarrow$	$(a, mt)(p, q)$	306.5	
11.	$(i, mn)(p, q)$	$\rightarrow$	$(i, mt)(p, \bar{q})$	306.11	

Notes: (1) is clearly out of line but there is no obvious emendation of the text. (2) Ibn Sīnā specifies in these cases that the difference-like sentence is 'strict' (*Qiyās* 305.7, 309.9, 313.6).

Figure 14.2: Translations from difference-like to meet-like

We begin with the translations of (a, mn). In numbers 8 and 9 Ibn Sīnā tells us that the disjunction is 'strict', i.e. exclusive, and the translations agree with this. Presumably these two translations are not invertible.

Translation 1 takes 'Always either p or q' to 'Always if not p then q', which is as we would expect. All of 2–7, except for the rogue 7, can be derived from 1 either by changing the variables or by contraposing in the (a, mt) formula, or both. We note that at the very least in 4, 5 we would expect an existential augment on the meet-like formula in the context of the logic of *Qiyās* vi.1. But it seems nobody has suggested any reason for putting a corresponding existential augment on the (a, mn) sentence. So we have confirmation that Ibn Sīnā has now abandoned the augments, and we also have a reason why: the augments don't fit with the difference-like sentences.

We come to (e, mn). What translation should we expect for this form? Analogy with the assertorics, the two-dimensional sentences and the meet-like sentences suggests that the negation of $(a, mn)(p, q)$ should be—ignoring augments—either $(a, mn)(\bar{p}, q)$ or $(a, mn)(p, \bar{q})$. The symmetry of (a, mn)

makes it hard to see why Ibn Sīnā should prefer either of these to the other. So we test them both against his translations. If $(e, mn)(p, q)$ is $(a, mn)(\bar{p}, q)$, then we can use (14.3.2) and (14.3.3) to get a complete set of equivalences between meet-like and difference-like forms, as follows (where $\sharp$ means contradictory negation):

OPTION A:

$$\begin{array}{ccccccc} (a, mt)(p, q) & \Leftrightarrow & (e, mt)(p, \bar{q}) & \Leftrightarrow & (a, mn)(\bar{p}, q) & \Leftrightarrow & (e, mn)(p, q) \\ \sharp & & \sharp & & \sharp & & \sharp \\ (o, mt)(p, q) & \Leftrightarrow & (i, mt)(p, \bar{q}) & \Leftrightarrow & (o, mn)(\bar{p}, q) & \Leftrightarrow & (i, mn)(p, q) \end{array}$$

On the other hand if $(e, mn)(p, q)$ is $(a, mn)(p, \bar{q})$ then we have

OPTION B:

$$\begin{array}{ccccccc} (a, mt)(p, q) & \Leftrightarrow & (e, mt)(p, \bar{q}) & \Leftrightarrow & (a, mn)(\bar{p}, q) & \Leftrightarrow & (e, mn)(\bar{p}, \bar{q}) \\ \sharp & & \sharp & & \sharp & & \sharp \\ (o, mt)(p, q) & \Leftrightarrow & (i, mt)(p, \bar{q}) & \Leftrightarrow & (o, mn)(\bar{p}, q) & \Leftrightarrow & (i, mn)(\bar{p}, \bar{q}) \end{array}$$

Translations 10 and 11 are both votes for Option A against Option B. But this is thin evidence. The translations from meet-like to difference-like will carry us further.

				Qiyās
1.	$(a, mt)(p, q)$	$\rightarrow$	$(e, mn)(p, q)$	306.8?, 310.10, 312.15, 314.10
2.	$(a, mt)(p, \bar{q})$	$\rightarrow$	$(e, mn)(p, \bar{q})$	311.17?
3.	$(a, mt)(\bar{p}, q)$	$\rightarrow$	$(e, mn)(\bar{p}, q)$	315.10
4.	$(a, mt)(\bar{p}, q)$	$\rightarrow$	$(a, mn)(p, q)$	313.9
5.	$(e, mt)(p, q)$	$\rightarrow$	$(e, mn)(\bar{p}, q)$	307.17, 315.3, 316.12
6.	$(e, mt)(\bar{p}, q)$	$\rightarrow$	$(o, mn)(p, q)$	314.6?
7.	$(e, mt)(\bar{p}, q)$	$\rightarrow$	$(e, mn)(p, q)$	317.6
8.	$(e, mt)(\bar{p}, q)$	$\rightarrow$	$(e, mn)(\bar{p}, q)$	316.4
9.	$(i, mt)(p, q)$	$\rightarrow$	$(o, mn)(p, q)$	311.10, 317.10, 324.4
10.	$(i, mt)(p, \bar{q})$	$\rightarrow$	$(o, mn)(p, \bar{q})$	312.9
11.	$(i, mt)(p, \bar{q})$	$\rightarrow$	$(o, mn)(\bar{p}, q)$	309.12
12.	$(i, mt)(\bar{p}, q)$	$\rightarrow$	$(e, mn)(\bar{p}, q)$	316.17

Figure 14.3: Translations from meet-like to difference-like

Here translations 1–3, four definite and two probable, all confirm Option A. But unfortunately this is not the end of the story, because 5 and 7 (four passages) all go with Option B. Translation 8 is wrong for both options. Translation 6 is also wrong for both options, and suspiciously translates from a universal to an existential.

Of the other translations from meet-like to difference-like, some are irrelevant to the question because they don't invove (e, mn) or (i, mn). These are translations 4 and 11 (clearly correct), and translations 9 and 10 (equally clearly incorrect, though they represent four passages). Translation 12 does mention (e, mn), but something is certainly wrong with it since it translates a particular as a universal.

Another passage provides further evidence. At *Qiyās* 384.1–5, in vii.2, Ibn Sīnā tells us that (e, mn) and (o, mn) sentences are entailed by three kinds of sentences as follows:

$$\begin{array}{lllll} & 1. & (a,mn)(\bar{p},q) & \to & (e,mn)(p,q) \\ & 2. & (i,mn)(\bar{p},q) & \to & (o,mn)(p,q) \\ & 3. & (a,mt)(p,q) & \to & (e,mn)(p,q) \\ (14.3.4) & 4. & (i,mt)(p,q) & \to & (o,mn)(p,q) \\ & 5. & (e,mt)(p,\bar{q}) & \to & (e,mn)(p,q) \\ & 6. & (o,mt)(p,\bar{q}) & \to & (o,mn)(p,q) \end{array}$$

Items 1, 3, and 5 all agree with Option A and not with Option B. (Items 2, 4 and 6 don't have any bearing on Options A and B. We note that item 4 is wrong.)

In short there is quite a lot of noise in the data, but the balance of the evidence is definitely in favour of Option A. Option A allows us to render the four difference-like forms as:

(14.3.5)

(a, mn) At all times t, at least one of p and q is true at t.

(e, mn) At all times t, if p is true at t then q is true at t.

(i, mn) There is a time at which p is true and q is not true.

(o, mn) There is a time at which neither p nor q is true.

There is one further source of evidence for these translations, but it is very murky and it raises more questions than it answers. This is Ibn Sīnā's choice of sentences for proving sterility, cf. Section se:4.5. In these proofs he gives sentences which are required to be true, and to be instances of certain sentence forms. We can examine the sentences that he offers as true sentences of the forms (e, mn) and (i, mn), and try to deduce what he must be taking these two forms to be. The next chapter is partly devoted to asking why this data is so hard to make sense of. Here let me simply list the sentences. In some cases the text may be corrupt.

First there are the sentences offered as instances of (e, mn):

306.14–17 (e, mn)(It is a number. It is a multiplicity divisible by two.)
(e, mn)(It is a number. It is a multiplicity not divisible by two.)

308.7–11 (e, mn)(It has a carrier. It is not a substance.)
(e, mn)(It has a carrier. Not every dimension is finite.)

309.1–4 (e, mn)(This is not a substance. This is in a subject.)
(e, mn)(This is an accident. No dimension is actually infinite.)

311.1–6 (e, mn)(This is even. This is a number.)
(e, mn)(This is even. The vacuum exists.)

312.3–5 (e, mn)(Zayd is drowning. Zayd is not flying.)
(e, mn)(Zayd is drowning. There is no vacuum.)

313.1–2 (e, mn)(Zayd is not not drowning. Zayd is not flying.)
(e, mn)(Zayd is not not drowning. There is no vacuum.)

313.15–314.2 (e, mn)(This is a vacuum. This is even.)
(e, mn)(This is even times even. This is even.)

314.16–18 (e, mn)(This is not rational. This is human.)
(e, mn)(This is a vacuum. This is human.)

315.15–17 (e, mn)(The human is not a body. The human is not mobile.)
(e, mn)(The human is not a vacuum. The human is not mobile.)

316.5–8 (e, mn)(It is a vacuum. It is not an even number.)
(e, mn)(It is not divisible by two. It is not an even number.)

317.1–3 (e, mn)(It is moving. It is a substance.)

318.1–3 (e, mn)(It is not at rest. It is a substance.)

If we reckon that the first sentence implies the second, this counts for Option A. If we reckon that the second sentence implies the first, this counts for Option B. On my reckoning it comes out roughly two to one in favour of Option A on a simple count of sentences, and if we count only the completely unambiguous cases then Option A wins hands down.

Second there are the sentences offered as instances of (i, mn):

305.12–306.2 (i, mn)(Zayd is changing place. Zayd is abstaining from walking.)
(i, mn)(It is coloured black. It has a pleasant smell.)

307.9–13 (i, mn)(It has a will. It is not moving.)
(i, mn)(It has a will. It is not at rest.)

(There should be another example at *Qiyās* 308.16f, but here Ibn Sīnā tells us to find our own sentences.) If we reckon that the first sentence can be true and the second false at the same time, this counts for Option A; if the other way round, this counts for Option B.

14.4 Three phases of propositional logic

The preceding two sections give us some leverage for distinguishing different stages within Ibn Sīnā's propositional logic. To speak of 'stages' suggests movement in time, and it looks to me overwhelmingly likely that the stages do represent different moments in Ibn Sīnā's exploration of propositional logic. But all of the stages appear in *Qiyās*, so it's clear that Ibn Sīnā himself felt that each of them made sense on its own. Nevertheless we can point to inconsistencies between any two of them, and some (but not all) of these inconsistencies are pointed out by Ibn Sīnā himself.

In Section 12.1 we described a propositional logic with the following features:

(14.4.1)

- The logic uses two sentence forms: 'If p then q' and 'Either p or q', and possibly also negative forms 'If p then not q' and 'Either p or not q'.
- These sentence forms can, but need not, be read as involving a universal quantification over time; the formal logic is developed without any reference to the quantification.
- There is no classification of propositional sentences into (a), (e), (i) or (o) forms.
- There is no suggestion that the form 'If p then q' is understood as implying that p is true, or sometimes true.

We found these features in Al-Fārābī's accounts of propositional logic. But they are also present in some sections of Ibn Sīnā's *Qiyās*. A prime example is *Qiyās* viii.1,2, a pair of sections on duplicative (*istiṯnāī*) syllogisms. The main formal content of these sections is a collection of inferences very much like those offered by Al-Fārābī. A typical example is

(14.4.2)
If the human is an animal then the human is a body.
But the human is an animal.
It follows that the human is a body.

(*Qiyās* 389.11–13 in viii.1.) Note that the compound is of the simple 'If p then q' form, and that no time quantification plays any role.

At *Qiyās* 392.15–17, still in viii.1, Ibn Sīnā attacks Peripatetic logicians who fail to distinguish properly between form and matter in meet-like sentences. He comments that for purposes of syllogism one needs to consider what are the first and second clauses but ignore what their content ('matter', *mādda*) is. Nowhere in this discussion does Ibn Sīnā suggest that one needs to consider whether the first and second clauses are bound together by a quantification over time, or indeed whether they are bound affirmatively or negatively. The implication is that he is thinking of sentences of the form 'If p then q' in the first instance, and not of the (a), (e), (i) and (o) forms of Section 12.2.

Al-Fārābī freely takes the contradictory negations of the separate clauses of a propositional compound, but he hardly ever refers to the contradictory negation of a whole compound. The only case where he does so is when an exclusive disjunction has three or more clauses, so that excepting one of them yields the negation of the disjunction of the remaining two or more clauses (*'antaja muqābilāti l-bāqiyya*, [5] 33.3). This is exactly the context in which Ibn Sīnā refers to contradictory negations of propositional compounds in *Qiyās* 401.13f (*yuntija naqīḍa l-munfaṣilati llatī tutammu mina l-bāqiyyatayn*).

Very significant in this same passage is Ibn Sīnā's use of *laysa albatta 'immā* p *wa-'immā* q to mean

(14.4.3) In no case is p true and in no case is q true.

(*Qiyās* 403.8f.) This reading of the (e) sentence in (14.3.1) is quite different from the reading that we found at (14.3.5) in Section 14.3 above (following Option A—but it's equally incompatible with the Option B reading too).

So both the underlying assumptions and the vocabulary in *Qiyās* viii.1,2 match the Farabian material in Section 12.1, in the ways listed in (14.4.1). They don't match the contents of *Qiyās* vi.1, which we reviewed in Section 12.2.

So I will distinguish propositional logic in the style of Section 1 and *Qiyās* viii.1, 2 as *PL1*, and propositional logic in the style of Section 2 and *Qiyās* vi.1 as *PL2*. The distinguishing criteria are listed in (14.4.1).

We saw that there are differences between *Qiyās* vi.1 and *Qiyās* vi.2. Above all:

(14.4.4)

- Though both sections classify the meet-like propositions into forms (a), (e), (i) and (o), only *Qiyās* vi.2 also classifies the difference-like propositions into these four forms.
- In *Qiyās* vi.2 there are no existential or universal augments on the (a) or (o) meet-like sentences.
- In *Qiyās* vi.2 there is free use of negation of the clauses.

I will distinguish propositional logic in the style of Section 3 and *Qiyās* vi.2 as *PL3*. The distinguishing criteria between PL2 and PL3 are those listed above. Note that PL1 and PL3 agree with each other, and against PL2, in not using augments on meet-like sentences.

We can test this classification by asking how it applies to Ibn Sīnā's other material on propositional logic. In the next section I will do exactly that with *Qiyās* vii.1, showing that the bulk of it belongs in PL3. On the other hand Ibn Sīnā's sequent rules, which we examine in Chapter 16 below, bear signs of PL1.

The metathetic negation in PL3 has a dramatic effect on the forms of proofs. We will examine this in Chapter 15. There are strong indications that Ibn Sīnā never caught up with the implications of this change. It seems he made serious mistakes of logic as a result. This might be one reason why almost none of PL3 is represented in *Išārāt*.

History 14.4.1 At the end of *Qiyās* vi.6, 356.7–357.15, Ibn Sīnā distinguishes his own approach to propositional logic from one that he found in a book. The main distinctions that he makes are in fact between PL1 (which he ascribes to the book) and PL2–3 (which he claims as his own view). Thus he complains that the book doesn't classify the sentences correctly as universal and particular, affirmative and negative. He adds that the book gives an incorrect account of contrariety, contradiction and equivalence. (On his own account, this second fault is a consequence of the first; 'if you know about affirmative, negative, universal and existential, then you already know about contradiction and contrariety ...', *Qiyās* 362.5f.)

14.5 *Qiyās* vii.1 and some readings of Ibn Sīnā

In Book vii.1 of *Qiyās*, 361–372, Ibn Sīnā collects up the various forms of meet-like sentence with assertoric clauses, and discusses the logical relationships between them. For example he says (*Qiyās* 366.1–3) that each kind of (e, mt) sentence is reducible to an (a, mt) sentence and vice versa.

(14.5.1) The mode of reduction is that you keep the quantity of the proposition as it was and you alter the quality, and the first clause is kept as it was but is followed by the contradictory negation of the second clause.

In other words, $(a, mt)(p, q)$ is equivalent to $(e, mt)(p, \bar{q})$ and $(e, mt)(p, q)$ is equivalent to $(a, mt)(p, \bar{q})$. We met both of these equivalences Section 14.3; they are common to Option A and Option B.

In the terminology of History 14.3.4 above, altering the quality is simple negation, and taking the contradictory negation of the second clause is metathesis. The use of metathesis is one of the marks of PL3.

Another mark of PL3 is that there are no augments on (a) or (o) meet-like sentences. The equivalence stated above wouldn't hold if the (a) sentence was required to have an existential augment. We can pin down the exact point where this matters in Ibn Sīnā's derivation of $(a, mt)(p, q)$ from $(e, mt)(p, \bar{q})$. He assumes $(e, mt)(p, \bar{q})$ and the contradictory negation of $(a, mt)(p, q)$, which is $(o, mt)(p, q)$, and he simply states that these contradict each other. This makes sense only if he is taking literally, and without universal augment, his statement

(14.5.2) In this case [i.e. assuming $(o, mt)(p, q)$] we will have that p and together with it q.

(*Qiyās* 367.4f.) In fact the equivalence is fully in accord with the conventions in *Qiyās* vi.2, though not with those in *Qiyās* vi.1. So PL3 it is.

In [68] p. xxxiiif, Khaled El-Rouayheb discusses this passage together with Khūnajī's reaction to it. He describes the reduction from (a) to (e) as 'a principle attributed to Boethius'. From his reference to a paper of Chris Martin [76] we learn that the principle in question states

(14.5.3) $$\vdash \ (P \to Q) \to -(P \to -Q).$$

Straight away we see that Boethius' principle has no time quantifiers, so if it represents anything in Ibn Sīnā's thinking, it would have to be something in PL1 and not in PL3. We can turn $(P \to Q)$ into $(a, mt)(P, Q)$ by adding a universal quantification over time. But $(e, mt)(P, -Q)$ is the universal time quantification of $(P \to - - Q)$, not of $-(P \to -Q)$. I suggest that this misunderstanding could never have arisen if Ibn Sīnā's text had been read in the vocabulary of PL2 or PL3 rather than that of PL1.

In brief, Ibn Sīnā's point in this passage has nothing to do with Boethius' principle. The equivalence that Ibn Sīnā states is immediate and hardly controversial, given what he takes the sentences to mean. It would make sense to go back to Khūnajī's objections to see whether Khūnajī has misunderstood Ibn Sīnā, or whether he is deliberately striking out in a new direction. There is no doubt that Khūnajī is raising some significant philosophical problems, even if they are not problems about *Qiyās* vii.1 as Ibn Sīnā intended it.

One moral of the discussion above is that a reader of Ibn Sīnā's propositional logic needs to be aware of the differences in assumptions and vocabulary between different strands of Ibn Sīnā's writing.

It's worth mentioning that one of the passages we quoted earlier as evidence of Ibn Sīnā's recognition of existential augments in propositional compounds was *Qiyās* 368.14–17, which is in the middle of *Qiyās* vii.1.

Maybe this is why Ibn Sīnā said 'potentially', meaning that the PL2 option is open though not forced on us. Maybe there is some other explanation.

In his pioneering article [88], written before the Cairo edition of *Qiyās* was available, Nicholas Rescher proposes formalisations of Ibn Sīnā's meet-like and difference-like sentences. His evidence is taken mainly from *Išārāt* and *Dānešnāmeh*. For the meet-like he offers ([88] p. 80):

(14.5.4)
$$(a)\ \forall t\ (At \to Ct).$$
$$(e)\ \forall t\ \neg(At \wedge Ct).$$
$$(i)\ \exists t\ (At \wedge Ct).$$
$$(o)\ \exists t\ (At \wedge \neg Ct).$$

(The notation is mine, not Rescher's.) These all agree with our findings above for PL3. The lack of augments on (a) and (o) disagrees with PL2. For the difference-like Rescher gives ([88] p. 82):

(14.5.5)
$$(a)\ \forall t\ (At \vee Ct).$$
$$(e)\ \forall t\ \neg(At \vee Ct).$$
$$(i)\ \exists t\ (At \vee Ct).$$
$$(o)\ \exists t\ \neg(At \vee Ct).$$

The formulas for (a) and (o) agree with our findings, but those for (e) and (i) are doubly wrong. He has disjunctions where Ibn Sīnā has conjunctions (or vice versa), and he incorrectly guesses that these two forms should be symmetric between A and C.

Apart from the difference-like forms (e) and (i), where frankly he had almost no information to call on, Rescher's formalisations are insightful and more accurate than much of the recent literature. But Rescher himself is partly to blame for these later mishaps, through some indefensible English translations that he offers in [88]. Thus for (i, mt) he suggests

(14.5.6) Sometimes: when A, then (also) C.

This implies, quite wrongly, that the (i, mt) form in PL2 and PL3 has a component of the form 'When p then q' (cf. History 14.2.4). His English for

(i, mn) has a similar mistake. For (e, mt) he has

(14.5.7) Never: when A, then (also) C.

This wrongly suggests that A lies within the scope of the negation, creating the same error as in El-Rouayheb's reading of (14.5.1) above (and indeed El-Rouayheb does use Rescher's translation of (e, mt) here).

Rescher's translations were taken up also by Shehaby [93]. We can see an effect of this in Shehaby's translation of *Qiyās* 330.10 ([93] p. 130), where Shehaby correctly calculates that a negation of the second clause of an (e, mt) sentence is needed, so he adds it, not realising that Ibn Sīnā has already expressed it with *laysa albatta*.

14.6 Exercises

1.

Chapter 15

Metathetic logic

The use of metathesis has quite a profound effect on the logic. Seen from Ibn Sīnā's end, it implies that every sentence is convertible (reversibly). Also some new moods become available, and justifications have to be found for these. Seen from our end, the effect is that the logic is no longer Horn, and hence we can't construct models by a simple closing-off operation; choices have to be made. But the language is still translatable into Krom formulas, a low-complexity fragment of first-order propositional logic that has been widely studied in the computer science logic literature (e.g. [7]).

15.1 Language and reductions

Definition 15.1.1 We introduce a logic that we call *metathetic logic*, $\mathcal{L}_{meta}$.

(a) The *admissible signatures* of $\mathcal{L}_{meta}$ are monadic relational signatures, with the feature that the relation symbols come in *polar pairs* of the form A, $\bar{A}$. We distinguish A and $\bar{A}$ as *positive* and *negative* relation symbols, and we say that they have *opposite polarities*. We say that two relation symbols R, S are *alphabetically distinct* if R and S are not both in the same polar pair.

(b) There is no negation symbol in the languages of $\mathcal{L}_{meta}$, but as a metalanguage symbol we use $\neg$ in front of relation symbols, with the meanings

$$\neg A = \bar{A}, \quad \neg\bar{A} = A$$

where A is any positive relation symbol. Note that $\neg\neg R = R$ for any relation symbol R.

(c) $\mathcal{L}_{meta}$ has two forms of sentence:

$$\forall\tau(R\tau \vee S\tau), \quad \exists\tau(R\tau \wedge S\tau)$$

where R and S are alphabetically distinct relation symbols. The first of these sentences is *universal* and the second is *existential*. We write them as $(a, me)(R, S)$ and $(i, me)(R, S)$ respectively; where the context allows, we abbreviate (a, me) and (i, me) to (a), (i) respectively. So metathetic logic has just the two sentence forms (a, me) and (i, me). The relation symbol R and S in these two sentences are known as the *subject relation symbol* and the *predicate relation symbol* respectively.

(d) Thus metathetic logic has no negative sentence forms; instead we negate a sentence by switching between positive and negative relation symbols. Thus the *contradictory negation* of $(a)(R, S)$ is $(i)(\neg R, \neg S)$, and the *contradictory negation* of $(i)(R, S)$ is $(a)(\neg R, \neg S)$.

(e) The semantics of metathetic logic $\mathcal{L}_{meta}$ is as for first-order logic with the added rule that if Σ is an admissible signature and M is a Σ-structure, then $\bar{A}^M = \text{dom}(M) \setminus A^M$ for every positive relation symbol A in Σ. Thanks to this rule, we can treat $\neg$ as a symbol of the language; the condition for $(\neg A)a$ to be true is the same as the usual condition for $\neg(Aa)$ to be true. As before, a set of sentences of $L(\Sigma)$ is *consistent* if some Σ-structure is a model of it, and *inconsistent* otherwise.

History 15.1.2 But note that if ϕ is an atomic sentence, then $\neg\neg\phi$ actually is the sentence ϕ, not just a sentence elementarily equivalent to it. Here we are following the Aristotelian convention for handling contradictory negation (*naqīḍ*).

Definition 15.1.3 Thanks to Definition 15.1.1(c) above, each metathetic sentence has a subject symbol and a predicate symbol. Two things result from this.

(a) We can speak of the *figure* of a syllogism in metathetic logic. But note that for the definition to be comparable with our earlier uses, we have to regard A and $\bar{A}$ as being the same relation symbol, occurring positively in the one case and negatively in the other.

(b) Every finite set of metathetic sentences has a subject-predicate digraph; cf. Section se:4.5 above. Again for this to agree with earlier definitions, we need to regard a vertex of the digraph as representing both A and $\bar{A}$ simultaneously. We will find uses for these digraphs, but they are much less important than they were for assertoric and two-dimensional theories. The reason is that the digraph of a minimal inconsistent theory is no longer a clue to a proof of inconsistency. For example the arrow representing a sentence $\forall x(Bx \to Ax)$ points from B to A and indicates that we reason from B to A. But with $\forall\tau(B\tau \vee A\tau)$ there is no preferred direction of reasoning; the disjunction is symmetric.

Lemma 15.1.4 *Every sentence of metathetic logic converts. Thus $(a)(R,S)$ converts to $(a)(S,R)$ and $(i)(R,S)$ converts to $(i)(S,R)$.* □

The sentence $(a, me)(R,S)$ expresses the same as $(a, mn)(R,S)$ expressed in Section 14.3 above. Likewise $(i, me)(R,S)$ expresses the same as $(i, mt)(R,S)$ of Section 12.2 above. By the equivalences of Section 14.3, going with Option A, this allows any two of the universal forms (a, me), (a, mt), (e, mt), (a, mn) and (e, mn) to be translated into each other, and likewise any two of the existential forms (i, me), (i, mt), (o, mt), (i, mn) and (o, mn). **Caution**: In metathetic logic the default is that there are no existential or universal augments, so (a, mt) and (o, mt) of Section 12.2 must have their augments removed before we use them in these translations.

The translations between the sentence forms of Section 12.2 and (a, me) and (i, me) keep the same subject and predicate relation symbols. So Lemma 15.1.4 implies that all these forms convert, thanks to metathetic negation. But the conversions are not all obvious. For example $(a, mt)(P, Q)$ translates to $(a, me)(\neg P, Q)$, which converts to $(a, me)(Q, \neg P)$, and this translates back to $(a, mt)(\neg Q, \neg P)$.

In the later sections of this chapter we will be building up metathetic logic as a self-sufficient logical system. But Ibn Sīnā justified syllogisms in metathetic logic by translating them back into meet-like logic (normally unaugmented). More precisely he would translate the premises into meet-like logic, draw a conclusion in meet-like logic, and then translate the conclusion back. Normally there will be no augments, so he is deprived of the use of meet-like *Darapti* and *Felapton*, but he can call on any of the other twelve meet-like moods in figures one to three. We will add the two moods in fourth figure.

This procedure of translating into meet-like syllogisms needs some justification. Certainly if the meet-like translation is valid, the original metathetic syllogism is valid, because validity is preserved by meaning-preserving translations. But the other direction is not so clear, because even when a metathetic syllogism is valid, it may translate into a meet-like syllogism with negative terms. The meet-like logic that we studied in Section 12.2 had no such terms.

The following devices allow us to handle some negative terms.

(1) $(a, mt)(R, \neg S)$ is equivalent to $(e, mt)(R, S)$, and $(i, mt)(R, \neg S)$ is equivalent to $(o, mt)(R, S)$. Likewise we can eliminate a negative predicate in an (e) sentence by switching to an (a) sentence, or from (o) to (i). (Since the mid nineteenth century these processes in traditional logic have been known as *obversion*.)

(2) If a term R occurs twice in the context $\neg R$, we can treat $\neg R$ as a new term—in effect switching between polar opposites. This is Ibn Sīnā's usual procedure.

(3) If a term occurs only once in subject position, and that one occurrence is negated, then likewise we can switch R to its polar opposite at both occurrences. This eliminates the negative term in subject position, and the occurrence in predicate position can be handled as in (1).

The remaining case, which won't go away, is where a term occurs twice in subject position, and only one of those occurrences is negated. This never happens in fourth figure, because of the shape of the fourth figure digraph. But it can and does happen in each of the other figures. Ibn Sīnā gives an example in each figure. In his examples, one of the premises is already a meet-like sentence and the other has to be translated from a difference-like form.

Ibn Sīnā provides an example in first figure at *Qiyās* 306.10 (a variant of the syllogism at *Qiyās* 306.3f). The syllogism is

(15.1.1) $$(e, mt)(R, Q), (i, mn)(Q, P) \vdash (o, mn)(R, P).$$

which translates to the meet-like syllogism

(15.1.2) $$(e, mt)(R, Q), (i, mt)(Q, \neg P) \vdash (o, mt)(\neg R, P).$$

Note that R is subject at both occurrences, but it is negated in only one of them. Note also that this is a first figure syllogism with an existential major premise, which violates the productivity conditions for assertoric syllogisms. Ibn Sīnā shows the validity of the syllogism by converting the minor premise to $(e, mt)(Q, R)$ and then obverting to $(a, mt)(Q, \neg R)$, and finally obverting the conclusion to $(i, mt)(\neg R, \neg P)$:

(15.1.3) $$(a, mt)(Q, \neg R), (i, mt)(Q, \neg P) \vdash (i, mt)(\neg R, \neg P).$$

This is a valid syllogism in *Disamis*.

Ibn Sīnā's example in second figure (*Qiyās* 316.14–16) is

(15.1.4) $$(i, mt)(P, Q), (a, mn)(\neg R, \neg Q) \vdash (i, mt)(\neg R, P).$$

(In fact for the second premise he has an exclusive disjunction of R and Q; we will come back to this.) In third figure (*Qiyās* 309.15, varying *Qiyās* 309.10) his example is

(15.1.5) $$(a, mt)(R, Q), (a, mn)(R, P) \vdash (e, mn)(Q, \neg P).$$

I leave it to the reader to check that these are examples (Exercise 13.1 below).

In these examples Ibn Sīnā shows, very likely for the first time, that there are metathetic syllogisms which are valid but can't be shown valid by translation of each sentence to an assertoric or meet-like sentence. These may also be the only occasions when he offers justifications by conversion that are not simply copied from Aristotle. However, there is no real problem about justifying these syllogisms. Since all metathetic sentences convert, the premises can be converted so that all the syllogisms are in fourth figure, and we saw that in fourth figure the metathesis is never a problem.

History 15.1.5 But Ibn Sīnā was not the only writer before modern times to give examples of valid metathetic syllogisms that can't be validated by translation to assertoric syllogisms. In *Quaestio 37* of his *In Libros Priorum Analyticorum Aristotelis Quaestiones* [91], Pseudo-Scotus observes that if a putative syllogism with negative terms can be translated into a valid assertoric syllogism by regarding each negative term as positive, then the syllogism is valid ('omnis modus alicuius figurae triorum figurarum valens ex terminis finitis, valet etiam ex terminis infinitis'). Here he is agreeing with move (2) above. But then he continues: some valid syllogisms with negative terms can't be justified by this method ('huiusmodi non potest fieri ex terminis finitis'). The example he gives is in third figure:

$$(e)(\neg R, Q), (e)(R, P) \vdash (e)(Q, P).$$

This is equivalent to Ibn Sīnā's third figure example but with R and $\neg R$ transposed.

Finally note:

Lemma 15.1.6 *There are no inference relations between any two of the metathetic sentences* $(a)(R, S)$, $(a)(\neg R, S)$, $(a)(R, \neg S)$, $(a)(\neg R, \neg S)$, $(i)(R, S)$, $(i)(\neg R, S)$, $(i)(R, \neg S)$ *and* $(i)(\neg R, \neg S)$.

Proof. Let Σ be an admissible signature for metathetic logic, and M a Σ-structure. Let R and S be two alphabetically distinct relation symbols in Σ. Then the domain of M is partitioned into four sets: $R^M \cap S^M$, $R^M \cap (\neg S)^M$, $(\neg R)^M \cap S^M$ and $(\neg R)^M \cap (\neg S)^M$. Each of the sentences in the lemma, when applied to M, expresses information about just one of these partition sets, and the information is either that the set is empty or that it is not empty. For example $(a)(R, S)$ expresses that $R^M \cap (\neg S)^M$ is empty. The pair of sentences relevant to one partition set are contradictory negations of each other, so there are no inferences from one to the other. For any two partition sets, we can choose them empty or non-empty independent of each other; hence there are no inference relations between any other pairs of sentences from the lemma. □

Corollary 15.1.7 *In metathetic logic every minimal inconsistent set is optimally minimal inconsistent.* □

15.2 Inconsistent metathetic theories

A weak form of the Law of Quantity (Section se:8.5) holds for metathetic logic.

Lemma 15.2.1 *(a) Every minimal inconsistent set of sentences contains at most one existential sentence.*

(b) There are minimal inconsistent sets of universal sentences.

Proof. (a) The proof is as Lemma le:8.5.2, but easier because we have only one domain to deal with.

(b) Consider the theory

(15.2.1) $\forall\tau(A\tau \vee B\tau),\ \forall\tau(A\tau \vee \neg B\tau),\ \forall\tau(\neg A\tau \vee B\tau),\ \forall\tau(\neg A\tau \vee \neg B\tau).$

Given the stipulation on nonempty domains, every model of this theory has an element a such that $(Aa \vee Ba)$, $(Aa \vee \neg Ba)$, $(\neg Aa \vee Ba)$ and $(\neg Aa \vee \neg Ba)$ are all true in the model. This is impossible. □

Definition 15.2.2 Let Σ be an admissible signature for metathetic logic. We adopt a constant symbol γ, and write $\Sigma(\gamma)$ for the result of adding this constant to Σ. The Skolem and Herbrand sentences of a metathetic theory T in $L(\Sigma)$ with at most one existential sentences are then sentences of $L(\Sigma(\gamma))$, as in the figure below. If T does contain an existential sentence then α is γ; if it doesn't, then we choose an arbitrary constant for α. As the Figure shows, all the Herbrand sentences have one of the two forms ϕ, $(\phi \vee \psi)$ where ϕ and ψ are atomic sentences. We refer to the set of sentences of these two forms as the *Herbrand fragment* of $L(\Sigma(\gamma))$.

form	Skolem sentences	Herbrand sentences
$\forall\tau(R\tau \vee S\tau)$	$\forall\tau(R\tau \vee S\tau)$	$(R\alpha \vee S\alpha)$
$\exists\tau(R\tau \wedge S\tau)$	$R\gamma, S\gamma$	$R\gamma, S\gamma$

Figure 15.1: Metathetic sentences, Skolem and Herbrand

Definition 15.2.3 Let ϕ and ψ be atomic sentences in $L(\Sigma(\gamma))$ as in Definition 15.2.2. We write

$$[\phi, \psi]$$

for any set of one or more sentences in the Herbrand fragment which can be arranged in a sequence (without repetition) as

$$(\phi_0 \vee \psi_0)^\star, (\phi_1 \vee \psi_1)^\star, \dots, (\phi_{n-1} \vee \psi_{n-1})^\star, (\phi_n \vee \psi_n)^\star \tag{15.2.2}$$

where

(a) each $(\eta \vee \theta)^\star$ is either $(\eta \vee \theta)$ or $(\theta \vee \eta)$,

(b) for each i ($0 \leqslant i < n$), ϕ_{i+1} is $\neg\psi_i$,

(c) ϕ is ϕ_0 and ψ is ψ_n.

Lemma 15.2.4 *Let Φ be a set of the form $[\phi, \psi]$, and suppose that in the corresponding sequence (15.2.2) there are $i < j$ such that both ϕ_i and ϕ_j are $R\gamma$. Then the set got from Φ by removing the sentences $(\phi_i, \psi_i)^\star, \dots, (\phi_{j-1}, \psi_{j-1})^\star$ is again of the form $[\phi, \psi]$.* □

Lemma 15.2.5 *(a) Any set $[\phi, \psi]$ is also a set $[\psi, \phi]$.*

(b) The union of sets $[\phi, \neg\psi]$ and $[\psi, \chi]$ contains a set $[\phi, \chi]$.

(c) $[\phi, \psi] \vdash (\phi \vee \psi)$.

Proof. (a) is immediate from the definition.

(b): If we concatenate the sequences represented by $[\phi, \neg\psi]$ and $[\psi, \chi]$, we get a sequence that fits the definition of $[\phi, \chi]$ except that some sentences may be repeated. Take the earliest sentence η which has two occurrences in the concatenated sequence, and delete one of the two occurrences and everything between them. The resulting sequence is of the form $[\phi, \chi]$.

(c) is by induction on the number of sentences in the sequence. If there is one, the result is trivial. Suppose the result is proved for n sentences, and $[\phi, \psi]$ has $n + 1$ sentences. Then for some atomic χ, $[\phi, \psi]$ can be written as $[\phi, \neg\chi]$ followed by either $(\chi \vee \psi)$ or $(\psi \vee \chi)$; without loss of generality assume the former. By induction hypothesis $[\phi, \neg\chi]$ entails $(\phi \vee \neg\chi)$. But $(\phi \vee \neg\chi)$ and $(\chi \vee \psi)$ together entail $(\phi \vee \psi)$. □

Definition 15.2.6 Let T be a set of sentences in the Herbrand fragment of $L(\Sigma(\gamma))$.

(a) Let ϕ be an atomic sentence. We say that T *declares* ϕ if T contains either ϕ or a set $[\phi, \phi]$; the sentence ϕ and the set $[\phi, \phi]$ are called *declarations* in T.

(b) We say that T is *metathetic inconsistent* if there are atomic sentences ϕ, ψ such that T declares ϕ and ψ, and either T contains a set $[\neg\phi, \neg\psi]$, or ϕ is $\neg\psi$.

If T is not metathetic inconsistent, we say that T is *metathetic consistent*.

Lemma 15.2.7 *Let T be a set of sentences as above, and suppose T is metathetic consistent. Let ϕ be any atomic sentence. Then either $T \cup \{\phi\}$ or $T \cup \{\neg\phi\}$ is metathetic consistent.*

Proof. Suppose for contradiction that T is metathetic consistent, but neither $T \cup \{\phi\}$ nor $T \cup \{\neg\phi\}$ is metathetic consistent. Since $T \cup \{\phi\}$ is not metathetic consistent, there must be an atomic sentence ψ such that T declares ψ and either T contains $[\neg\phi, \neg\psi]$ or ϕ is $\neg\psi$. Since $T \cup \{\neg\phi\}$ is not metathetic consistent, there must be an atomic sentence χ such that T declares χ and either T contains $[\phi, \neg\chi]$ or ϕ is χ. We consider the cases.

First, suppose T contains $[\neg\phi, \neg\psi]$ and $[\phi, \neg\chi]$. Then by Lemma 15.2.5(a), T contains $[\neg\psi, \neg\chi]$. But T declares both ψ and χ, and so T is already metathetic inconsistent.

Second, suppose T contains $[\neg\phi, \neg\psi]$ and ϕ is χ. Then T contains $[\neg\chi, \neg\psi]$, and again T is already metathetic inconsistent. The same argument works if ϕ is $\neg\psi$ and T contains $[\phi, \neg\chi]$.

Third, suppose ϕ is $\neg\psi$ and ϕ is χ. Then T declares both ϕ and $\neg\phi$, so again T is metathetic inconsistent. □

The property of being metathetic consistent is of finite character. So if T is metathetic consistent, then there is in $L(\Sigma(\gamma))$ a maximal metathetic consistent $T^+ \supseteq T$. (In this remark we are using a version of the Axiom of Choice. If metathetic logic had been within Horn logic this would not have been necessary.)

Lemma 15.2.8 *(a) For each relation symbol R, exactly one of $R\gamma$ and $\neg R\gamma$ is in T^+.*

(b) If $(\phi \vee \psi)$ is in T, then at least one of ϕ and ψ is in T^+.

Proof. (a) Since T^+ is maximal metathetic consistent, Lemma 15.2.7 tells us that either $R\gamma$ or $\neg R\gamma$ is in T^+. If both $R\gamma$ and $\neg R\gamma$ are in T^+ then it's immediate that T^+ is metathetic inconsistent.

(b) Suppose for contradiction that T contains $(\phi \vee \psi)$ but T^+ contains neither ϕ nor ψ. Then by (a), T^+ contains $\neg\phi$ and $\neg\psi$. But then T^+ is metathetic inconsistent, contradiction. □

Theorem 15.2.9 *Let T be a theory in the Herbrand fragment. Then T is consistent if and only if it is metathetic consistent.*

Proof. Suppose first that T is metathetic inconsistent. Then we can check from the cases of Definition 15.2.6(b) that T is inconsistent. For example if T contains sets $[\phi, \phi]$, $[\psi, \psi]$ and $[\neg\phi, \neg\psi]$, then we quote Lemma 15.2.5(c) to infer that T entails ϕ, ψ and $(\neg\phi \vee \neg\psi)$, making T inconsistent.

Second, suppose that T is metathetic consistent, in the language $L(\Sigma(\gamma))$. Then as above we can extent T to a set T^+ which is maximal metathetic consistent in the Herbrand fragment. We construct a $\Sigma(\gamma)$-structure M as follows. The domain of M consists of one element γ, which names itself. Let A be a positive relation symbol in Σ. By Lemma 15.2.8(a), exactly one of $A\gamma$, $\bar{A}\gamma$ is in T^+. We choose A^M to be $\{\gamma\}$ if and only if $A\gamma$ is in T^+. This choice guarantees that the atomic sentences true in M are exactly those in T^+. The remaining sentences in the Herbrand fragment are of the form $(\phi \vee \psi)$. By Lemma 15.2.8(b), if $(\phi \vee \psi)$ is in T then $(\phi \vee \psi)$ is true in M. Hence M is a model of T^+, and therefore of T too, so that T is consistent. hspace40pt □

Definition 15.2.10 We write

$$(a)[R, S]$$

for any set of one or more metathetic sentences which can be arranged in a sequence (without repetition) as

(15.2.3) $$(a)(R_0, S_0)^\star, (a)(R_1, S_1)^\star, \ldots, (a)(R_{n-1}, S_{n-1})^\star, (a)(R_n, S_n)^\star$$

where

(a) each $(a)(R_i, S_i)^\star$ is either $(a)(R_i, S_i)$ or $(a)(S_i, R_i)$,

(b) for each i ($0 \leqslant i < n$), R_{i+1} is $\neg S_i$,

(c) R is R_0 and S is S_n.

Theorem 15.2.11 *Let T be a set of metathetic sentences. Then T is inconsistent if and only if there are relation symbols R and S such that*

(a) either S is $\neg R$ or T contains a set $(a)[\neg R, \neg S]$; and

(b) either

(i) T contains the sentence $(i)(R, S)$ or $(i)(S, R)$, or

(ii) T contains a sentence $(i)(R, Q)$ or $(i)(Q, R)$ for some relation symbol Q, and a set $(a)[S, S]$, or

(iii) T contains sets $(a)[R, R]$ and $(a)[S, S]$.

Proof. If there are relation symbols R and S as stated, then the Herbrand theory of T is metathetic inconsistent, and hence inconsistent by Theorem 15.2.9. So T is inconsistent by Lemma 2.2.9.

Conversely if T is inconsistent, then by Lemma 2.2.9 again, the Herbrand theory of T is inconsistent, and hence metathetic inconsistent by Theorem 15.2.9 again. By examining what sentences in T the Herbrand sentences could have come from, we see that the condition on the relation symbols R and S holds. □

We consider the following situation. The metathetic theory T contains a set of the form $(a)[R, R]$ and a set of the form $(a)[\neg R, \neg S]$. In the style of ABOVE, we can display these two sets as a sequence

(15.2.4) $$(a)(R_0, S_0)^\star, (a)(R_1, S_1)^\star, \ldots, (a)(R_{n-1}, S_{n-1})^\star, (a)(R_n, S_n)^\star$$

where the sentences from $(a)(R_0, S_0)^\star$ to $(a)(R_m, S_m)^\star$ form $(a)[R, R]$ and the remaining sentences from $(a)(R_{m+1}, S_{m+1})^\star$ to $(a)(R_n, S_n)^\star$ form $(a)[\neg R, \neg S]$. We note two properties of this sequence:

(a) For each $i < n$, $S^i = \neg R_{i+1}$.

(b) There is k, $0 < k < n$, such that $R_0 = \neg R_{k+1}$.

If the sequence has these two properties, then the first $k+1$ sentences form a set of the form $(a)[R_0, R_0]$ (after removing repeated sentences if necessary), and the remaining $n - k$ sentences are a set of the form $(a)[R_{k+1}, S_n]$ (with the same proviso). Let us say that the sequence is *minimal* if there is no sequence got from the present one by removing one or more sentences, which has the same final symbol S_n and still satisfies conditions (a) and (b). Note that minimality guarantees that the provisos about repeated sentences are already met.

Lemma 15.2.12 *If the sequence above is minimal, then* $(0, k+1)$ *is the only pair of indices* (i, j) *with* $i < j < n$ *and* $R_i = \neg R_j$.

Proof. If $(0, h)$ is another such pair, say with $h < k+1 < n$, then properties (a) and (b) are preserved if we remove the sentences from h to k. If (i, h) is another such pair with $0 < i$, then properties (a) and (b) are preserved if we remove the sentences from 0 to $i-1$, replacing k by $h-1$. □

Lemma 15.2.13 *If the sequence above is minimal, then there are no indices* i, j *with* $0 \leqslant i < j \leqslant n$ *such that* $R_i = R_j$.

Proof. If $j \leqslant k+1$ then (a) and (b) are preserved if we remove the sentences from i to $j-1$. If $i \geqslant k+1$ then again (a) and (b) are preserved if we remove the sentences from i to $j-1$. There remains the case where $i < k+1$ and $j > k+1$. In this case we note that we can reverse the ordering of the first $k+1$ sentences, relabelling as appropriate. The relation symbol S_{i-1} in the old ordering is $\neg R_i$; the new ordering makes this symbol an $R_{i'}$ with $i' > 0$. So we are in the situation of the previous lemma, and as in that lemma we can remove sentences from the sequence while preserving (a) and (b). □

This allows us to characterise the minimal inconsistent sets of types (i) and (ii). THIS WILL NEED proof that taking any sentence away allows finding a model. For this, in case (i) we want: A GRAPH-LINEAR THEORY HAS A MODEL. In case (ii) we need that the declaration is consistent on its own, and the track is consistent without the declaration. The latter follows from the LINEAR THEORY fact. The former needs further. NOTE THAT we could assume no repetitions in the declaration except at the ends. Can satisfy the declaration by making every sentence in it satisfied everywhere. Then can add part of the track, which has no letters in common with it.

15.3 Metathetic MITs and syllogisms

From Theorem ABOVE follows that if T is minimal inconsistent, then T contains the relevant sentences as described in (a) and one of (b)(i), (b)(ii) or (b)(iii) as in the theorem, and no other sentences. In general it doesn't follow that every such theory T is minimal inconsistent, because there could be some scope for removing some sentences (EXAMPLE IN EXERCISES).

Definition 15.3.1 In Theorem ABOVE, if T is minimal consistent:

(a) We call the set $(a)[\neg R, \neg S]$ the *track*, with *left end* R and *right end* S; in the case where S is $\neg R$ we say that *the track is empty*.

(b) We say that T is of the form **M** if (b)(i) applies, of the form **N** if (b)(ii) applies, and of the form **O** if (b)(iii) applies. NO, WE WANT A DESCRIPTION FROM WHICH WE CAN DEDUCE MINIMALITY.

Lemma 15.3.2 *If T is minimal inconsistent then no relation symbol occurs at two junctions in the track.*

Lemma 15.3.3 *If T is minimal inconsistent then no relation symbol occurs with opposite polarities at two junctions in the track.*

Proof. Suppose T consists of the sentence $(i)(R, S)$ together with sentences

$$(a)(R_0, S_0)^\star, (a)(R_1, S_1)^\star, \ldots, (a)(R_{n-1}, S_{n-1})^\star, (a)(R_n, S_n)^\star \tag{15.3.1}$$

where R_0 is $\neg R$ and S_0 is $\neg S$. By LEMMA, all of $R_0, \ldots, R_n$ are distinct relation symbols. Suppose now that there are $i < j$ such that R_i is $\neg R_j$. Then R_i is S_{j-1}, and so the sequence contains $(a)[R_i, R_i]$. Hence we can change to a set T' of the form **N** by removing all the sentences from $(a)(R_j, S_j)^\star$ onwards, and taking the track to finish at $(a)(R_{i-1}, S_{i-1})$. Can we assume T' is smaller than T? Suppose $i = 0$ and $j = n$. Then ϕ is $\neg\psi$, so by minimality there is no track, contradiction. Hence either $i > 0$ or $j < n$. Either way, after swapping the construction over from left to right if necessary, we can arrange that some sentence gets left out in T'. □

Lemma 15.3.4 *If T is of the form **M** (as above with no relation symbol at two junctions, with either polarity), then T is minimal inconsistent.*

Proof. If we leave off the existential sentence then we are not in any of the forms. If we leave out a universal sentence the track becomes either disconnected or loose from one of the endpoints, and either way again we are not in any of the three forms. □

Theorem 15.3.5 *If T has a circular graph, then it is minimal inconsistent if and only if it has the form*

$$(i)(R, S), (a)[\neg R, \neg S] \tag{15.3.2}$$

for two relation symbols R, S that are not alphabetically related.

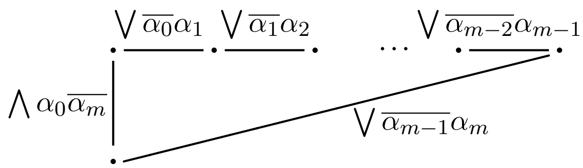
Proof. If it is minimal inconsistent then it is in one of the three forms. Only form **M** allows it to be graph-circular. Conversely if it has the form given, and is graph-circular, then by LEMMA above it is minimal inconsistent. □

Ibn Sīnā has two devices for getting more than we should do out a premise-pair. They both amount to treating a premise ϕ as a conjunction of ϕ and another sentence. The first is where ϕ is $\forall\tau(B\tau \vee A\tau)$, and Ibn Sīnā treats this as a conjunction with the sentence $\forall\tau(\neg B\tau \vee \neg A\tau)$; he calls this taking ϕ as *strict* (*ḥaqīqī*). The second is where ϕ is $\forall\tau(B\tau \rightarrow A\tau)$ and he adds an existential augment. Logically this is equivalent to conjoining ϕ with $\exists\tau(B\tau \wedge A\tau)$. A third device of this kind was available to him in the Peripatetic tradition, though he doesn't use it in *Qiyās* vi.2. This is to conjoin $\forall\tau(B\tau \rightarrow A\tau)$ with $\forall\tau(\neg B\tau \rightarrow \neg A\tau)$, which ELSEWHERE he calls taking $\forall\tau(B\tau \rightarrow A\tau)$ as *complete*. This third device translates into the first if we translate the (a, mt) into a metathetic (a, mn).

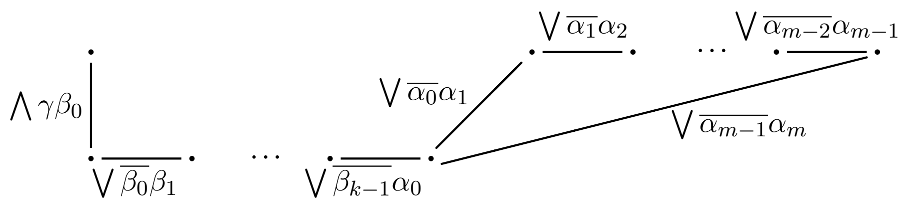
Ibn Sīnā always makes it explicit when he uses the first or second device. Also he is careful about the metathesis: whenever he uses them, he calls attention to the fact that B and A are not metathetically negated. The reason for this restriction is not clear to me. Does he think that it reflects some fact about normal scientific usage?

In each of these cases, inspection of the forms of minimal inconsistent set shows that at most one of the conjuncts can have any effect on the productivity. If the original premise-pair was productive, then adding the new conjunct produces no new conclusions. If the original premise-pair wasn't productive but the new premise-pair is, then the conclusion from the new premise-pair owes nothing to the old conjunct, and for logical purposes we should catalogue the premise-pair as if the new conjunct replaced the old one rather than being conjoined to it. Ibn Sīnā takes the opposite view and catalogues the premise-pair on the basis of the old conjunct. This accounts for some differences between Ibn Sīnā's cataloguing in this section and ours in Appendix C.

Form MA(m): not including the existential edge on the left, m edges along the top and n edges along the bottom. Parameters: $n \geqslant 1, m+n \geqslant 2$. $|\Gamma| = m+n+1$. CHECK THESE. This is the case where both confirmations are direct but starting at different relation symbols (which must come from the same $\bigwedge$ since they have the same constant). Any number of sentences from two upwards.

(15.3.3) 

Form MB(k, m): not including the existential edge on the left, m edges along the top and n edges along the bottom. Parameters: $n \geqslant 1, m + n \geqslant 2$. $|\Gamma| = m + n + 1$. CHECK THESE. This is the case where both confirmations are direct and they start at the same relation symbol, which must come from one side of a $\bigwedge$. This can also be described as: one confirmation direct and one indirect. The simplest case is three sentences.

(15.3.4) 

MC will be the case where both confirmations are indirect. This can be described either as two separate confirmations or as one confirmation covering both cases. This always involves at least four sentences.

Restrict now to three sentences.

A *(valid) mood* is an array derived from this form in the following two steps, which both involve adding some extra features. First we choose one of the three sentences and call its contradictory negation the *conclusion*. Then the other sentence which contains the subject relation symbol of the conclusion becomes the *minor premise*, and the remaining sentence becomes the *major premise*, exactly as with the assertorics and the two-dimensionals.

Thus each syllogism is in one of the four figures, as with assertoric syllogisms. As with assertoric syllogisms, Ibn Sīnā ignores the fourth figure, but we will carry it along for mathematical tidiness. We can distinguish forms according to which sentences are universal and which are existential. (Both (a) and (i) are affirmative, so there is not much use in trying to distinguish moods by affirmative versus negative.) So in all we have twelve moods.

We write M(m, n) for the mood in the m-th figure where the n-th sentence is existential. In this notation the twelve valid moods are as follows. At the right I list the assertoric moods that correspond.

(15.3.5)

$\mathbf{M}(1,1)$	$\bigvee(\bar{r},q), \bigwedge(\bar{q},p) \vdash \bigvee(\bar{r},p)$	$Darii, Ferio$
$\mathbf{M}(1,2)$	$\bigwedge(\bar{r},q), \bigvee(\bar{q},p) \vdash \bigvee(\bar{r},p)$	$-$
$\mathbf{M}(1,3)$	$\bigwedge(\bar{r},q), \bigwedge(\bar{q},p) \vdash \bigwedge(\bar{r},p)$	$Barbara,\ Celarent$
$\mathbf{M}(2,1)$	$\bigvee(\bar{r},q), \bigwedge(p,\bar{q}) \vdash \bigvee(\bar{r},p)$	$-$
$\mathbf{M}(2,2)$	$\bigwedge(\bar{r},q), \bigvee(p,\bar{q}) \vdash \bigvee(\bar{r},p)$	$Festino, Baroco$
$\mathbf{M}(2,3)$	$\bigwedge(\bar{r},q), \bigwedge(p,\bar{q}) \vdash \bigwedge(\bar{r},p)$	$Cesare, Camestres$
$\mathbf{M}(3,1)$	$\bigvee(q,\bar{r}), \bigwedge(\bar{q},p) \vdash \bigvee(\bar{r},p)$	$Datisi, Ferison$
$\mathbf{M}(3,2)$	$\bigwedge(q,\bar{r}), \bigvee(\bar{q},p) \vdash \bigvee(\bar{r},p)$	$Disamis, Bocardo$
$\mathbf{M}(3,3)$	$\bigwedge(q,\bar{r}), \bigwedge(\bar{q},p) \vdash \bigwedge(\bar{r},p)$	$-$
$\mathbf{M}(4,1)$	$\bigvee(q,\bar{r}), \bigwedge(p,\bar{q}) \vdash \bigvee(\bar{r},p)$	$Fresison$
$\mathbf{M}(4,2)$	$\bigwedge(q,\bar{r}), \bigvee(p,\bar{q}) \vdash \bigvee(\bar{r},p)$	$Dimatis$
$\mathbf{M}(4,3)$	$\bigwedge(q,\bar{r}), \bigwedge(p,\bar{q}) \vdash \bigwedge(\bar{r},p)$	$Calemes$

When we allow negative sentences to occur as subsentences, both as first clause and as second clause, it becomes less practical to maintain a distinction between affirmative and negative sentences. Leaving aside the augments, everything can be said using only affirmative sentences.

Still leaving aside the augments and the syllogistic moods that depend on them (*Darapti* and *Felapton*), abandoning the negative sentence forms cuts down the number of valid syllogisms. In first figure, *Celarent* and *Ferio* become respectively *Barbara* and *Darii*, but with negative major subsentences. Likewise in third figure, *Bocardo* and *Ferison* become *Disamis* and *Datisi* but with negative major subsentences. This cuts down the number of optimal assertoric forms from 12 to 8. Also it ceases to matter which subsentences are affirmative and which are negative; the only thing we need to know is whether two occurrences of the same subsentence are of the same or different qualities. This abolishes the distinction between *Cesare* and *Camestres*, and likewise the distinction between *Festino* and *Baroco*, both in second figure.

But three new forms become available, because we can put negative sentences as first clauses. The new forms are as follows:

(15.3.6)
First figure:
Whenever r then not q.
It can be the case, when q, that p.
Therefore it can be the case, when not r, that p.

Second figure:
(15.3.7) Whenever r then not q.
It can be the case, when p, that q.
Therefore it can be the case, when not r, that p.

Third figure:
(15.3.8) Whenever q then r.
Whenever not q then p.
Therefore whenever not r then p.

The existing three optimal moods in fourth figure, *Calemes*, *Dimatis* and *Fresison*, are not affected by allowing negative clauses. So we are left with three distinct moods in each of the four figures. In each figure the moods are distinguished as one with no existentially quantified premise, one with the minor premise existentially quantified and one with the major premise existentially quantified.

The new first figure mood appears at *Qiyās* 306.10. Ibn Sīnā proves it by converting the minor premise, taking the syllogism to a third figure syllogism

(15.3.9) Whenever q then (not r).
It can be the case, when q, that p.
Therefore it can be the case, when (not r), that p.

which is *Disamis* with a negative clause.

The new second figure mood appears at *Qiyās* 316.14. Ibn Sīnā proves it by switching the order of the premises and then converting the conclusion:
(15.3.10)
It can be the case, when p, that q.
Whenever r then not q.
Therefore it can be the case, when p, that not r.
Therefore it can be the case, when not r, that p.

The syllogism used here is *Festino*.

The new third figure mood appears at *Qiyās* 309.15. Curiously Ibn Sīnā first proves (at *Qiyās* 309.10) a mood with existentially quantified conclusion, that relies on the major premise being strict *munfaṣil*. But then he observes that we can get the new mood by taking the contrapositive of the minor premise:
(15.3.11)
Whenever (not r) then (not q).
Whenever (not q) then p.
Therefore whenever (not r) then p.

which is *Barbara* with two negative clauses.

These three arguments are entirely correct, and they are a novel solution to a novel problem (though not a particularly demanding one).

The breakdown of the distinction between affirmative and negative sentences is helpful to Ibn Sīnā in another context. In *Qiyās* viii.3 he gives an account of proofs by reductio ad absurdum. We can summarise his account as follows.

Suppose we are aiming to prove 'Not every C is a B'. Then we assume 'Not every C is a B' is false, and we deduce a contradiction. He offers the view that when we make this assumption, we enter an argument where the sentences depending on this assumption should all be understood to have 'If it is false that not every C is a B' as an antecedent, even when this antecedent is silent. On this view, the argument proceeds with sentences that are taken to be true absolutely, not subject to an assumption, because if they are stated in full they incorporate the assumption. Thus we begin the argument by stating 'If not not every C is a B then every C is a B' (*Qiyās* 408.13), and we finish it with a statement of the form 'If not not every C is a B then absurdity', which entails 'Not every C is a B' by one final logical step. To justify this view it has to be shown that if an inference step is valid, then it remains valid if 'If θ then ...' is added to the conclusion and one of the premises. Ibn Sīnā devotes *Qiyās* vii.4 to showing this in detail when the inference step is an assertoric syllogism.

Now there is a snag. Suppose the statement 'Not every C is a B' is a logical proof. Then 'Not not every C is a B' is an impossibility. Ibn Sīnā treats 'If p then q' as a variant of 'Whenever p then q'. But the affirmative (a) sentence 'Whenever p then q' is false if there is no situation where p is true, and in particular it is false if p is impossible. But it is not clear that the view expressed in the previous paragraph achieves anything useful if the sentences in the expanded version of the argument are not all true absolutely.

Strictly 'Not not every C is a B' is a negated sentence, since it is the negation of 'Not every C is a B'. I don't know of any place where Ibn Sīnā discusses whether a doubly negated sentence counts as affirmative or negative. But if the presence of a negation at the beginning of 'Not not every C is a B'' does count as nullifying the existential augment on the (a) sentence, then the whole problem disappears at once. Ibn Sīnā never says that he is following this route. But the complete absence of any explanation of how he copes with the existential augment in reductio arguments suggests that he has some rather straightforward and obvious way of handling it. I suggest that this cancellation of the existential augment in the case of negated subsentences is the obvious candidate.

When reductio is used to prove an affirmative sentence, say 'Every C is a B', then Ibn Sīnā's view would have us adding an assumed antecedent 'Not every C is a B' to the relevant steps of the argument. This antecedent is negative on any account. So it should certainly cancel the existential augment if we are reading Ibn Sīnā's procedures with *muttaṣil* and *munfaṣil* sentences correctly.

15.4 Productivity and sterility

DEFINE DISTRIBUTED.

Theorem 15.4.1 *Let $T(C, A)$ be a finite graph-linear metathetic theory in $L(\Sigma)$. Then $T(C, A)$ is productive if and only if the following hold:*

(a) T contains at most one existential sentence.

(b) Every relation symbol occurring in two sentences of T is distributed in exactly one of the sentences.

WHAT PRODUCTIVITY CONDITIONS DOES HE STATE?

15.5 Exercises

1.

(a) Translate (13.2), (13.4) and (13.5) into syllogisms using only the sentence forms (a, me) and (i, me).

(b) Translate each of these metathetic syllogisms into meet-like sentences, and convince yourself that the resulting syllogisms are all valid. Confirm also that none of these translations have the form of valid (non-metathetic) meet-like syllogisms, and that in each case the premise-pair fails to meet one of the productivity conditions for assertoric syllogisms.

(c) Find conversions that convert your meet-like translations of (13.4) and (13.5) into valid meet-like syllogisms.

2.

Part V

Global

Chapter 16

Sequent rules

16.1 Application to reductio ad absurdum

16.2 The potential of Ibn Sīnā's sequent rules

16.3 Exercises

1.

Chapter 17

Decidability

17.1 Why are Ibn Sīnā's logics decidable?

Suppose Σ is a finite signature, and $L(\Sigma)$ is the first-order language with signature Σ. A 'logic' that uses $L(\Sigma)$ will normally have a set S of sentences which is a subset (not necessary proper) of the set of sentences of $L(\Sigma)$, and an entailment relation $\vdash$ between these sentences. We say that the logic is *decidable* if the set of finite sequences of sentences of S

$$\{(\phi_0, \ldots, \phi_{n-1}, \chi) : \phi_0, \ldots, \phi_{n-1} \vdash \chi\}$$

is a computable (i.e. a recursive set, meaning that there is an algorithm which determines, for any finite sequence of symbols, whether or not it is in the given set).

In this sense Aristotle's assertoric logic, in any suitable signature, is decidable. This is a consequence of the fact that the suitable signatures consist of sets of 1-ary relation symbols, and if Σ is any such signature then $L(\Sigma)$ is decidable. However, all the better-defined logics introduced by Ibn Sīnā, such as his core two-dimensional logic and his propositional logic at level PL3, are decidable too. The two-dimensional logic needs 2-ary relation symbols. We know that first-order logic with 2-ary relation symbols is undecidable, even in the limited form that Ibn Sīnā uses (see Theorem .2.2 below). So we might suspect that the 1-ary relation symbols of assertoric logic are a red herring.

In fact all the logics of Ibn Sīnā that we have investigated so far in this book are decidable by the following theorem.

Theorem 17.1.1 (Mortimer's Theorem) *Let Σ be a relational signature, $L(\Sigma)$ the first-order language with signature Σ (and with equality =). Let S be a set* *of sentences of $L(\Sigma)$ closed under negation, such that every sentence of S uses at most two variables (though each variable may occur any number of times). Then every finite set of sentences of S that has a model has a finite model. Hence the set of finite sequences of sentences of S*

$$\{(\phi_0, \ldots, \phi_{n-1}, \chi) : \phi_0, \ldots, \phi_{n-1} \vdash \chi\}$$

is computable.

I am not yet entirely convinced that Mortimer's Theorem is the right one to quote here, because the best general proof of the theorem (Börger et al. [14] pp. 377–381) bears no clear relation to the distinctive features of Ibn Sīnā's languages. For example it would make more sense if it turned out that all Ibn Sīnā's decidable logics could be proved decidable because they all have some property along the lines of the 'guardedness' of Andréka, Németi and Van Benthem [6]. The remainder of this chapter is aimed at finding a meaningful dividing line between decidable logics of the Ibn Sīnā type and undecidable ones.

Ibn Sīnā's modal and temporal logics are open-ended: he concentrates on certain sentence forms, but he does mention others and he never draws a firm line around the sentences that he is willing to consider. For example he mentions two-dimensional sentences where the time quantification has wide scope, and he mentions an interpretation of possibility that refers to the future and hence to the linear ordering of time. Both of these examples could lead to sentences that no longer obey Mortimer's condition of being expressible using only two variables.

For example he might find himself considering a two-dimensional sentence along the lines

$$\forall\tau\exists x(\exists\sigma Bx\sigma \to \neg Ax\tau)$$

Thus: For every time τ there is an object x such that if x is ever a B, then it is not an A at time τ. The nested scopes of the variables make it impossible to write this with just two variables. But do sentences of this kind, added to Ibn Sīnā's two-dimensional logic, ever lead to undecidability?

Or again he might assume, in some argument about future possibilities, that time is linearly ordered. The axioms of linear ordering can't be written with just two variables, because there are first-order sentences about linear orderings that have models but no finite ones. The transitivity axiom

$$\forall\rho\forall\sigma\forall\tau(\rho \leqslant \sigma \wedge \sigma \leqslant \tau \to \rho \leqslant \tau)$$

uses three variables. The theory of linear orderings is decidable (by Ehrenfeucht [26]), but could it lead to an undecidable logic in the present two-dimensional context?

17.2 Undecidability on the horizon

17.3 Exercises

1.

Appendix A

Assertoric moods by figure

The numbering of the first, second and third figure moods is standard. Ibn Sīnā rejects the fourth figure and doesn't number its moods; the numbering below is taken from *Manṭiq al-kabīr* [83] attributed to Rāzī, pp. 162af.

Ibn Sīnā runs through these moods and their justifications in at least six places:

- *Muḵtaṣar* 49b9–53a6;
- *Najāt* 57.1–64.3;
- *Qiyās* ii.4, 108.12–119.8;
- *Qiyās* vi.1, 296.1–304.4 for meet-like (*muttaṣil*) sentences (recombinant propositional moods)
- *Dānešnāmeh* 67.5–80.2;
- *Išārāt* 142.10–153.9.

Sadly the relevant part of his earlier *Ḥikma al-ᶜArūḍiyya* is missing in the one surviving manuscript.

We list for each non-perfect mood the justifications that Ibn Sīnā gives for it. It is clear that the justifications given in *Qiyās* ii.4 for assertoric moods and in *Qiyās* vi.1 for recombinant propositional moods are almost identical. A comparison with Aristotle's *Prior Analytics* i.4–6 will show that Ibn Sīnā departs significantly from Aristotle in only one place, namely that he adds an ecthetic justification for *Baroco*.

Figure 1 mood 1, *Barbara*

(A.0.1)
$$\begin{array}{l}(\forall x(Cx \to Bx) \wedge \exists xCx)\\ (\forall x(Bx \to Ax) \wedge \exists xBx)\\ \vdash (\forall x(Cx \to Ax) \wedge \exists xCx)\end{array}$$

Aristotle	*Mukṯaṣar*	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
$25b37$	$49b13f$	57.5	109.16	296.3	67.5	143.3

Figure 1 mood 2, *Celarent*

(A.0.2)
$$\begin{array}{l}(\forall x(Cx \to Bx) \wedge \exists xCx)\\ \forall x(Bx \to \neg Ax)\\ \vdash \forall x(Cx \to \neg Ax)\end{array}$$

Aristotle	*Mukṯaṣar*	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
$25b40$	$49b14$	57.8	110.1	296.5	68.3	143.5

Figure 1 mood 3, *Darii*

(A.0.3)
$$\begin{array}{l}\exists x(Cx \wedge Bx)\\ (\forall x(Bx \to Ax) \wedge \exists xBx)\\ \vdash \exists x(Cx \to Ax)\end{array}$$

Aristotle	*Mukṯaṣar*	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
$26a17$	$49b15$	57.12	110.2	296.8	68.9	143.7

Figure 1 mood 4, *Ferio*

(A.0.4)
$$\begin{array}{l}\exists x(Cx \wedge Bx)\\ \forall x(Bx \to \neg Ax)\\ \vdash (\exists x(Cx \to \neg Ax) \vee \neg \exists xCx)\end{array}$$

Aristotle	*Mukṯaṣar*	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
$26a17$	$49b17$	57.15	110.3	296.11	69.4	143.7

Figure 2 mood 1, *Cesare*

(A.0.5)
$$(\forall x(Cx \to Bx) \wedge \exists x Cx)$$
$$\forall x(Ax \to \neg Bx)$$
$$\vdash \forall x(Cx \to \neg Ax)$$

	Aristotle	*Mukṯaṣar*	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
	$27a5$	$51a5f$	59.2	114.6	300.15	70.3	147.10
$\varepsilon \to$ Celarent	$27a6$	$51a6$	59.3	114.6	300.16	70.5	147.11
$\kappa^2 \to$ Ferio	$27a14$	$51a7$	59.5	114.8	300.17		

Figure 2 mood 2, *Camestres*

(A.0.6)
$$\forall x(Cx \to \neg Bx)$$
$$(\forall x(Ax \to Bx) \wedge \exists x Ax)$$
$$\vdash \forall x(Cx \to \neg Ax)$$

	Aristotle	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
	$27a9$	59.10	115.17	301.3	70.11	147.13
$\varepsilon \to$ Calemes	$27a13$	59.11	115.17	301.3	71.1	147.14
$\kappa \to$ Darii	$27a14$		116.2	301.5		

Figure 2 mood 3, *Festino*

(A.0.7)
$$\exists x(Cx \wedge Bx)$$
$$\forall x(Ax \to \neg Bx)$$
$$\vdash (\exists x(Cx \wedge \neg Ax) \vee \neg \exists x Cx)$$

	Aristotle	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
	$27a32$	60.3	116.4	301.9	71.6	148.3
$\varepsilon \to$ Ferio	$27a34$	60.5	116.4	301.10	71.8	
$\kappa \to$ Celarent		60.5	116.5	301.10		

Figure 2 mood 4, *Baroco*

(A.0.8)
$$(\exists x(Cx \wedge \neg Bx) \vee \neg\exists x Cx)$$
$$(\forall x(Ax \rightarrow Bx) \wedge \exists x Ax)$$
$$\vdash (\exists x(Cx \rightarrow Ax) \vee \neg\exists x Cx)$$

	Aristotle	*Muḵtaṣar*	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
	$27a36$		60.7	116.8	301.11	72.3	148.5
$\kappa \rightarrow$ Barbara	$27a38$		61.2	116.9	301.13	74.1	148.6
$\theta \rightarrow$ **B(1,2)** $\rightarrow$ **B(1,1)**		$51b3$	60.11	116.10	301.13	73.4	$148.8ch$

Figure 3 mood 1, *Darapti*

(A.0.9)
$$(\forall x(Bx \rightarrow Cx) \wedge \exists x Bx)$$
$$(\forall x(Bx \rightarrow Ax) \wedge \exists x Bx)$$
$$\vdash \exists x(Cx \wedge Ax)$$

	Aristotle	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
	$28a17$	61.10	117.6	302.10	75.3	151.3
$\downarrow\rightarrow$ Darii	$28a19$	61.11	117.10	302.12	75.5	151.5
$\kappa^2 \rightarrow$ Celaront	$28a22$	$62.2(Celarent)$	117.10	302.12		

Aristotle 28a24 gives a proof by ecthesis.

Figure 3 mood 2, *Felapton*

(A.0.10)
$$(\forall x(Bx \rightarrow Cx) \wedge \exists x Bx)$$
$$\forall x(Bx \rightarrow \neg Ax)$$
$$\vdash (\exists x(Cx \wedge \neg Ax) \vee \neg\exists x Cx)$$

	Aristotle	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
	$28a26$	62.3	117.15	302.15	75.9	153.7
$\downarrow\rightarrow$ Ferio	$28a28$	62.4	118.1	303.1	76.2	
$\kappa \rightarrow$ Camestros	$28a29$	$62.4pres$	118.1	303.1		

Figure 3 mood 3, *Datisi*

(A.0.11)
$$\begin{array}{l}\exists x(Bx \wedge Cx)\\ (\forall x(Bx \to Ax) \wedge \exists x Bx)\\ \vdash \exists x(Cx \wedge Ax)\end{array}$$

	Aristotle	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
	$28b7$	62.5	118.4	303.3	76.5	153.5
$\iota \to$ Darii	$28b9$	$62.6(likeDrpt)$	118.5	303.5	76.7	
$\kappa \to$ Ferio		$62.6(likeDrpt)$	$118.4?$	303.5		

Aristotle 28b14 also refers to a proof by ecthesis.

Figure 3 mood 4, *Disamis*

(A.0.12)
$$\begin{array}{l}(\forall x(Bx \to Cx) \wedge \exists x Bx)\\ \exists x(Bx \wedge Ax)\\ \vdash \exists x(Cx \wedge Ax)\end{array}$$

	Aristotle	*Muḵtaṣar*	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
	$28b11$	$52b5$	62.7	118.7	303.6	76.10	151.7
$\iota \to$ Dimatis	$28b13$	$52b5$	$62.8cmnt$	118.9	303.7	77.2	151.8
$\kappa^2 \to$ Celarent	$28b14$	$52b6$	$63.2pres$	118.11	303.8		–
$ec \to$ **B(1,2)**	$28b14$	–	$62.12f$	118.7	–		–

Najāt 62.13 refers to a proof by ecthesis for the non-convertible case; said it will be given elsewhere, but where?

Figure 3 mood 5, *Bocardo*

(A.0.13)
$$\begin{array}{l}(\forall x(Bx \to Cx) \wedge \exists x Bx)\\ (\exists x(Bx \wedge \neg Ax) \vee \neg\exists x Bx)\\ \vdash (\exists x(Cx \wedge \neg Ax) \vee \neg\exists x Cx)\end{array}$$

	Aristotle	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
	28*b*17	63.3	118.14	303.9	78.4	152.10
$\kappa \to$ Baroco	28*b*19		119.3	303.11		
$\kappa^2 \to$ Barbara		63.9			78.8	152.13
$\theta \to$ **D(3,1)**	28*b*20	63.6	119.1	303.11	78.4	152.15

Figure 3 mood 6, *Ferison*

(A.0.14)
$$\begin{array}{l} \exists x(Bx \wedge Cx) \\ \forall x(Bx \to \neg Ax) \\ \vdash (\exists x(Cx \wedge \neg Ax) \vee \forall x \neg Cx) \end{array}$$

	Aristotle	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
	28*b*33	63.12	119.5	304.1	79.4	153.8
$\iota \to$ Ferio	28*b*34	63.13	119.5	304.1	79.6	
$\kappa \to$ Camestres		63.13$pres$	119.7	304.1		

Figure 4 mood 1, *Bamalip*

(A.0.15)
$$\begin{array}{l} (\forall x(Bx \to Cx) \wedge \exists x Bx) \\ (\forall x(Ax \to Bx) \wedge \exists x Ax) \\ \vdash \exists x(Cx \wedge Ax) \end{array}$$

Figure 4 mood 2, *Dimatis*

(A.0.16)
$$\begin{array}{l} (\forall x(Bx \to Cx) \wedge \exists x Bx) \\ \exists x(Ax \wedge Bx) \\ \vdash \exists x(Cx \wedge Ax) \end{array}$$

	Aristotle	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
$\varepsilon! \to$ Darii		62.9	118.9	303.8		

Figure 4 mood 3, *Calemes*

(A.0.17)
$$\begin{array}{l} \forall x(Bx \to \neg Cx) \\ (\forall x(Ax \to Bx) \wedge \exists x Ax) \\ \vdash \forall x(Cx \to \neg Ax) \end{array}$$

	Aristotle	*Najāt*	*Qiyās* ii	*Qiyās* vi	DN	*Išārāt*
			116.1	301.5		
$\iota! \to$ Celarent		59.12	116.1	301.5		147.14

Figure 4 mood 4, *Fesapo*

(A.0.18)
$$\begin{array}{l} (\forall x(Bx \to Cx) \wedge \exists x Bx) \\ \forall x(Ax \to \neg Bx) \\ \vdash (\exists x(Cx \wedge \neg Ax) \vee \forall x \neg Cx) \end{array}$$

Figure 4 mood 5, *Fresison*

(A.0.19)
$$\begin{array}{l} \exists x(Bx \wedge Cx) \\ \forall x(Ax \to \neg Bx) \\ \vdash (\exists x(Cx \wedge \neg Ax) \vee \neg \exists x Cx) \end{array}$$

Ibn Sīnā gives no proofs of invalidity for conclusions from productive assertoric premise-pairs. For proofs of sterility in the assertoric case, he relies on stating conditions of productivity and then suggesting reasons why these conditions are necessary. These statements and arguments are at the following references. The references in *Išārāt* should be taken with a pinch of salt, because in that book Ibn Sīnā treats all predicative logics simultaneously.

	Muḵtaṣar	*Najāt*	*Qiyās* ii	*Qiyās* vi	*DN*	*Išārāt*
General	49b6–8	53.11–13	108.8f	295.10f	65.5–7	142.8
1st fig	49b11f,	57.4	109.8f	296.1	66.4–67.2	142.12–143.1
	50a6–50b4	58.3–5				
2nd fig	50b7f	58.8–10	111.10	299.10–12	69.10f	145.14–16
3rd fig	51b18f	61.7f	117.5	302.8	74.6–75.1	150.16f

Appendix B

Metathetic moods by figure

We write $\mathrm{M}(m, n)$ for the mood in the m-th figure where the n-th sentence is existential. In this notation the twelve valid moods are as follows. At the right I list the assertoric moods that correspond.

Figure 1, first premise $\exists$; *Darii, Ferio*

$$\begin{array}{l} \exists\tau(R\tau \wedge Q\tau) \\ \forall\tau(\neg Q\tau \vee P\tau) \\ \vdash \exists\tau(R\tau \wedge P\tau) \end{array}$$

Qiyās	method
where	how
where	how

Figure 1, second premise $\exists$; new

$$\begin{array}{l} \forall\tau(R\tau \vee Q\tau) \\ \exists\tau(\neg Q\tau \wedge P\tau) \\ \vdash \exists\tau(R\tau \wedge P\tau) \end{array}$$

Qiyās	form	verdict	method
where	whether	how	

Figure 1, conclusion $\forall$; *Barbara*

$$\forall\tau(R\tau \vee Q\tau), \forall\tau(\neg Q\tau \vee P\tau) \ \vdash \ \forall\tau(R\tau \vee P\tau)$$

Qiyās	form	verdict	method
305.8	$R-, Q+, P-$	Yes	meet-like *Barbara*

where how

(B.0.1)

$$\begin{array}{lll}
\mathbf{M}(1,1) & \bigvee(\bar{r},q), \bigwedge(\bar{q},p) \vdash \bigvee(\bar{r},p) & Darii, Ferio \\
\mathbf{M}(1,2) & \bigwedge(\bar{r},q), \bigvee(\bar{q},p) \vdash \bigvee(\bar{r},p) & - \\
\mathbf{M}(1,3) & \bigwedge(\bar{r},q), \bigwedge(\bar{q},p) \vdash \bigwedge(\bar{r},p) & Barbara,\ Celarent \\
\mathbf{M}(2,1) & \bigvee(\bar{r},q), \bigwedge(p,\bar{q}) \vdash \bigvee(\bar{r},p) & - \\
\mathbf{M}(2,2) & \bigwedge(\bar{r},q), \bigvee(p,\bar{q}) \vdash \bigvee(\bar{r},p) & Festino, Baroco \\
\mathbf{M}(2,3) & \bigwedge(\bar{r},q), \bigwedge(p,\bar{q}) \vdash \bigwedge(\bar{r},p) & Cesare, Camestres \\
\mathbf{M}(3,1) & \bigvee(q,\bar{r}), \bigwedge(\bar{q},p) \vdash \bigvee(\bar{r},p) & Datisi, Ferison \\
\mathbf{M}(3,2) & \bigwedge(q,\bar{r}), \bigvee(\bar{q},p) \vdash \bigvee(\bar{r},p) & Disamis, Bocardo \\
\mathbf{M}(3,3) & \bigwedge(q,\bar{r}), \bigwedge(\bar{q},p) \vdash \bigwedge(\bar{r},p) & - \\
\mathbf{M}(4,1) & \bigvee(q,\bar{r}), \bigwedge(p,\bar{q}) \vdash \bigvee(\bar{r},p) & Fresison \\
\mathbf{M}(4,2) & \bigwedge(q,\bar{r}), \bigvee(p,\bar{q}) \vdash \bigvee(\bar{r},p) & Dimatis \\
\mathbf{M}(4,3) & \bigwedge(q,\bar{r}), \bigwedge(p,\bar{q}) \vdash \bigwedge(\bar{r},p) & Calemes
\end{array}$$

Bibliography

[1] Mohammad Achena and Henri Massé, *Avicenne: Le Livre de Science: I Logique, Métaphysique, II Science Naturelle, Mathématiques*, Les Belles Lettres, Paris 2007.

[2] Asad Q. Ahmed, *Avicenna's Deliverance: Logic* with introduction by Tony Street, Oxford University Press, Oxford 2011.

[3] Al-Fārābī, *al-Manṭiq ^{c}inda l-Fārābī* (3 volumes), edited by R. al-^{c}Ajam, Dār al-Mashriq, Beirut 1985, 1986, 1986.

[4] Al-Fārābī, *Kitāb al-maqūlāt*, in [3] Vol. 1, pp. 89–131.

[5] Al-Fārābī, *Kitāb al-qiyās*, in [3] Vol. 2, pp. 11–64.

[6] Hajnal Andréka, István Németi and Johan van Benthem, 'Modal languages and bounded fragments of predicate logic', *Journal of Philosophical Logic* 27 (1998) 217–274.

[7] Bengt Aspvall, Michael F. Plass and Robert Endre Tarjan, 'A linear-time algorithm for testing the truth of certain quantified boolean formulas', *Information Processing Letters* 8.3 (1979) 121-123.

[8] Allan Bäck, on affirmative and existential augment.

[9] Allan Bäck, *Al-'Ibāra: Avicenna's Commentary on Aristotle's De Interpretatione: part one and part two*, trans. Bäck, Philosophia, Munich 2013.

[10] Abū al-Barakāt, *Kitāb al-mu^{c}tabar fī al-ḥikmat al-ilāhiyya*, Byblion, Jbeil Lebanon 2007.

[11] Patrick Blackburn, Maarten de Rijke and Yde Venema, *Modal Logic*, Cambridge University Press, Cambridge 2002.

[12] I. M. Bocheński, *A History of Formal Logic*, trans. Ivo Thomas, Chelsea, New York 1970.

[13] George Boole, *An Investigation of the Laws of Thought on Which are Founded the Mathematical Theories of Logic and Probabilities*, Dover, New York 1958 (original 1854).

[14] Egon Börger, Erich Grädel and Yuri Gurevich, *The Classical Decision Problem*, Perspectives in Mathematical Logic, Springer, Berlin 1997.

[15] John Buridan, *Summulae: De Suppositionibus*, ed. Ria van der Lecq, Artistarium vol. 10-4, Ingenium, Nijmegen 1998; trans. Gyula Klima in *John Buridan, Summulae de Dialectica*, Yale University Press, New Haven 2001.

[16] John Buridan, *Treatise on Consequences*, trans. and introduced by Stephen Read, Fordham University Press, New York 2015.

[17] Rudolf Carnap, 'Meaning postulates', *Philosophical Studies* 3 (1952) 65–73; reprinted in Rudolf Carnap, *Meaning and Necessity: a Study in Semantics and Modal Logic*, University of Chicago Press, Chicago !L 1956, pp. 212–220.

[18] Gary Chartrand and Ping Zhang, *A First Course in Graph Theory*, Dover 2012.

[19] Saloua Chatti, 'Existential import in Avicenna's modal logic', *Arabic Sciences and Philosophy* 26 (2016) 45–71.

[20] René Descartes, *Descartes: Regulae ad Directionem Ingenii*, ed. G. Crapulli, Nijhoff, The Hague 1966.

[21] Joseph Dichy, 'Al-Ḫaiīl's conjecture: How the first comprehensive dictionary in history was invented', in Manuela E. B. Giolfo (ed.), *Arab and Arabic Linguistics: Traditional and New Theoretical Approaches*, *Journal of Semitic Studies* Supplement 34, Oxford University Press, Oxford 2014, pp. 43–68.

[22] Randall R. Dipert, 'The life and logical contributions of O. H. Mitchell, Peirce's gifted student', *Transactions of the Charles S. Peirce Society* 30 (3) (1994) 515–542.

[23] D. M. Dunlop, 'Al-Fārābī's paraphrase of the Categories of Aristotle [Part 2]', *Islamic Quarterly* 5 (1959) 21–54.

[24] Khaled El-Rouayheb, *Relational Syllogisms and the History of Arabic Logic, 900–1900*, Brill, Leiden 2010.

[25] Khaled El-Rouayheb, 'Post-Avicennan logicians on the subject matter of logic: some thirteenth- and fourteenth-century discussions', *Arabic Sciences and Philosophy* 22 (1) (2012) 69–90.

[26] Andrzej Ehrenfeucht, 'Decidability of the theory of one linear ordering relation', *Notices of American Mathematical Society* 6 (1959) 268–9.

[27] Asadollah Fallahi, 'A critical edition of the risāla 'Illuminating the fourth figure' by Nūšatuh Majd al-Dīn Jīlī (Persian, Jīlī's text in Arabic)', *Falsafe va Kalam-e Eslami* 48 (2) (2015/6) 201–244.

[28] Al-Fārābī, *Kitāb al-Ḥurūf*, ed. Mohsin Mahdi, Dar El-Machreq, Beirut 1990.

[29] Gottlob Frege, *Begriffsschrift*, Nebert, Halle 1879; trans. Stefan Bauer-Mengelberg in *From Frege to Gödel*, ed. Jan van Heijenoort, Harvard University Press, Cambridge MA 1967, pp. 5–82.

[30] Gottlob Frege, 'Über die Grundlagen der Geometrie'. *Jahresbericht der Deutschen Mathematikervereinigung* 15 (1906) 293–309, 377–403, 423–430.

[31] Jean H. Gallier, *Logic for Computer Science: Foundations of Automatic Theorem Proving*, Harper and Row, New York 1986.

[32] Rodolphus Goclenius, *Isagoge in Organum Aristotelis*, Palthenius, Frankfurt 1598.

[33] A.-M. Goichon, *Lexique de la Langue Philosophique d'Ibn Sīnā*, Desclée de Brouwer, Paris 1938.

[34] Dimitri Gutas, *Avicenna and the Aristotelian Tradition: Introduction to Reading Avicenna's Philosophical Works, Second Edition*, Brill, Leiden 2014.

[35] Ahmad Hasnawi, 'Avicenna on the quantification of the predicate (with an appendix on [Ibn Zur^{c}a])', in S. Rahman et al. (eds.), *The Unity of Science in the Arabic Tradition*, Springer 2008, pp. 295–328.

[36] Ahmad Hasnawi and Wilfrid Hodges, chapter on early Arabic logic in forthcoming *Handbook of Medieval Logic*, ed. Catarina Dutilh and Stephen Read, Cambridge University Press (forthcoming).

[37] D. Hilbert and W. Ackermann, *Grundzüge der Theoretischen Logik*, Springer, Berlin 1928; trans. of second edition (1937) by L. M. Hammond et al., *Principles of Mathematical Logic*, Chelsea, New York 1950.

[38] Wilfrid Hodges, *A Shorter Model Theory*, Cambridge University Press, Cambridge 1997.

[39] Wilfrid Hodges, 'The laws of distribution for syllogisms', *Notre Dame Journal of Formal Logic* 39 (1998) 221–230.

[40] Wilfrid Hodges, 'Traditional logic, modern logic and natural language', *Journal of Philosophical Logic* 38, 589–606 (2009).

[41] Wilfrid Hodges, 'Ibn Sina on analysis: 1. Proof search. Or: Abstract State Machines as a tool for history of logic', in *Fields of Logic and Computation: Essays Dedicated to Yuri Gurevich on the Occasion of his 70th Birthday*, ed. A. Blass, N. Dershowitz and W. Reisig, Lecture Notes in Computer Science 6300, Springer, Heidelberg 2010, pp. 354–404.

[42] Wilfrid Hodges, 'Affirmative and negative in Ibn Sina', in *Insolubles and Consequences: Essays in honour of Stephen Read*, ed. Catarina Dutilh Novaes and Ole Thomassen Hjortland, College Publications, London 2012, pp. 119–134.

[43] Wilfrid Hodges, 'The move from one to two quantifiers', in *The Road to Universal Logic: Festschrift for 50th Birthday of Jean-Yves Béziau, Vol. 1*, ed. Arnold Koslow and Arthur Buchsbaum, Birkhäuser, Basel 2015, pp. 221–240.

[44] Wilfrid Hodges, 'Ibn Sīnā on reductio ad absurdum', *Review of Symbolic Logic* (to appear).

[45] Wilfrid Hodges, *A Manual of Ibn Sīnā's Formal Logic* (in preparation).

[46] Wilfrid Hodges, 'Ibn Sīnā applies his temporal logic to a thesis of Aristotle' (preprint).

[47] Wilfrid Hodges and Amirouche Moktefi, *Ibn Sīnā and the practice of logic*, in preparation.

[48] Wilfrid Hodges and Spencer Johnston, 'Ibn Sīnā's two-dimensional syllogisms and Buridan's divided modal syllogisms' (provisional title, for IfCoLog volume in memory of Grisha Mints).

[49] G. E. Hughes and M. J. Cresswell, *A New Introduction to Modal Logic*, Routledge, London 1996.

[50] Shams C. Inati, *Ibn Sīnā Remarks and Admonitions, Part One: Logic*, Pontifical Institute of Mediaeval Studies, Toronto Ontario 1984.

[51] Avicenna, *Avicennae opera in lucem redacta: ac nuper per canonicos emendata. Logyca. Sufficientia. De Celo et Mundo. De Anima. De Animalibus. De Intelligentiis. Alpharabius de Intelligentiis. Philosophia prima*, ed. Caecilius Fabrianensis, Per Bonetum Locatellum, Venice 1508.

[52] Ibn Sīnā, *Lettre au Vizir Abū Sa^{c}d*, ed. Yahya Michot, Éditions al-Bouraq, Beirut 2000.

[53] Ibn Sīnā, *Al-Maqūlāt*, ed. M. El-Khodeiri, G. C. Anawati et al., al-Hay'ah al-^{c}Āmmah li-Shu'ūn al-Maṭābi' al-Amīyah, Cairo 1959.

[54] Ibn Sīnā, *Al-^{c}ibāra*, ed. M. El-Khodeiri et al., Dār al-Kātib al-^{c}Arabī lil-Ṭabā^c wal-Našr, Cairo 1970.

[55] Ibn Sīnā, *Al-qiyās (Syllogism)*. Ed. S. Zayed, Cairo 1964.

[56] Ibn Sīnā, *Al-nafs*, ed. G. Anawati and S. Zayed, al-Hay'a al-Miṣriyya al-^{c}Amma lil-Kitāb, Cairo 1983.

[57] Ibn Sīnā, *Kitāb al-najāt*, ed. M. Danishpazuh, Tehran University Press, Tehran 1364h (1945).

[58] Ibn Sīnā, *Al-muḵtaṣar al-awsaṭ*, REF.

[59] Ibn Sīnā, *Manṭiq al-mašriqiyyīn*, Al-Maktaba al-Salafiyya, Cairo 1910.

[60] Ibn Sīnā, *Resālah manṭeq dānešnāmeh ^{c}alā'ī*, ed. Mohammad Mo^{c}īn and Sayyed Mohammad Meshkāt, Hamadan 2004.

[61] Ibn Sīnā, *Al-išārāt wal-tanbīhāt*, ed. Mojtaba Zāre^{c}ī, Būstān-e Ketab-e Qom, Qum 2000. The logical part is translated: S. C. Inati, *Ibn Sīnā, Remarks and Admonitions, Part One: Logic*, Pontifical Institute of Mediaeval Studies, Toronto 1984.

[62] Ibn Sīnā, *Kitāb al-ḥudūd (Book of Definitions)*, ed. Muḥammad Mahdī Fūlādvand, Soroush Press, Tehran 1987.

[63] Ibn Sīnā, *^{c}Uyūn al-ḥikma*, ed. M. F. al-Jabr, Dār al-Yanābī, Damascus 1996.

[64] Spencer C. Johnston, *Essentialism, Nominalism, and Modality: the Modal Theories of Robert Kilwardby and John Buridan*, PhD Thesis, University of St Andrews, 2015.

[65] Alexander Kalbarczyk, 'The *Kitāb al-Maqūlāt* of the *Muhtaṣar al-awsat fī l-manṭiq*: a hitherto unknown source for studying Ibn Sīnā's reception of Aristotle's *Categories*', *Oriens* 40 (2012) 305–354.

[66] Kamran Karimullah, 'Alfarabi on conditionals', *Arabic Sciences and Philosophy* 24 (2014) 211–267.

[67] John Neville Keynes, *Studies and Exercises in Formal Logic*, 4th edition, Macmillan, London 1906.

[68] Afḍal al-Dīn al-Khūnajī, *Kashf al-asrār ᶜan ghawāmiḍ al-afkār*, ed. and introduced by Khaled El-Rouayheb, Iranian Institute of Philosophy, Tehran 2010.

[69] Hermann Landolt, 'Khwāja Naṣīr al-Dīn al-Ṭūsī (597/1201–672/1274). Ismā'īlism, and Ishrāqī philosophy', in N. Pourjavady and Ž. Vesel (eds.), *Naṣ=ır al-Dīn Ṭūsī, Philosophe et Savant du XIIIe Siècle*, Presses Universitaires d'Iran, Tehran 2000, pp. 13–30.

[70] Norman Kretzmann, *William of Sherwood: Introduction to Logic*, University of Minnesota Press, Minneapolis Minnesota 1966.

[71] G. W. Leibniz, *New Essays on Human Understanding*, ed. and trans. Peter Remnant and Jonathan Bennett, Cambridge University Press, Cambridge 1996.

[72] David Londey and Carmen Johanson, *The Logic of Apuleius*, Brill, Leiden 1987.

[73] Jan ukasiewicz, *Aristotle's Syllogistic from the Standpoint of Modern Formal Logic*, Second Edition, Clarendon Press, Oxford 1957.

[74] Penelope Maddy, *Defending the Axioms: on the Philosophical Foundations of Set Theory*. Oxford University Press, Oxford 2011.

[75] Miklos Maróth, *Ibn Sīnā und die peripatetische "Aussagenlogik"*, Brill, Leiden1989.

[76] Chris Martin, 'The logic of negation in Boethius', *Phronesis* 36 (1991) 277–304.

[77] O. H. Mitchell, 'On a new algebra of logic', in *Studies in Logic*, ed. Charles Sanders Peirce et al., Little Brown and Company, Boston Mass. 1983, pp. 72–106.

[78] Zia Movahed, 'Ibn-Sina's anticipation of the formulas of Buridan and Barcan', in *Logic in Tehran*, ed. Ali Enayat et al., Association for Symbolic Logic and A. K. Peters, Wellesley Mass. 2003, pp. 248–255.

[79] Zia Movahed, 'A critical examination of Ibn Sīnā's theory of the conditional syllogism', *Sophia Perennis* 1 (1) (2009) pp. 5–19. www.ensani.ir/storage/Files/20120507101758-9055-5.pdf.

[80] John Philoponus, *In Aristotelis Analytica Priora Commentaria*, ed. M. Wallies, Reimer, Berlin 1905.

[81] Dag Prawitz, *Natural Deduction*, Almqvist & Wiksell, Stockholm 1965.

[82] Fak̠r al-Dīn al-Rāzī, *Al-manṭiq al-mulak̠k̠aṣ*,

[83] Fak̠r al-Dīn al-Rāzī, *Al-manṭiq al-kabīr*, ms Ahmad III 3401.

[84] Fak̠r al-Dīn al-Rāzī, *Lubāb*,

[85] Stephen Read, 'John Buridans theory of consequence and his octagons of opposition', in *Around and beyond the square of opposition*, ed. Jean-Yves Beziau and Dale Jacquetter, Springer, Basel 2012, pp. 93–110.

[86] Stephen Read, 'Aristotle's theory of the assertoric syllogism',

[87] David Reisman, *The Making of the Avicennan Tradition: The Transmission, Contents, and Structure of Ibn Sīnā's Al-Mubāḥat̠āt (The Discussions)*, Brill, Leiden 2002.

[88] Nicholas Rescher, 'Avicenna on the Logic of "Conditional" propositions', in Nicholas Rescher, *Studies in the History of Arabic Logic*, University of Pittsburgh Press, Liverpool 1983, pp. 76–86. GIVE ORIGINAL. Notre Dame journal of formal logic 4.1 (1963): 48-58.

[89] Nicholas Rescher and Arnold van der Nat, 'The theory of modal syllogistic in medieval Arabic philosophy', in Nicholas Rescher, *Studies in Modality*, American Philosophical Quarterly Monograph 8, Blackwell, Oxford 1974, pp. 17–56.

[90] Lenhart K. Schubert and Francis Jeffry Pelletier, 'Generically speaking, or, using discourse representation theory to interpret generics', in G. Chierchia, B. H. Partee and R. Turner eds., *Properties, Types and Meanings II*, Kluwer 1989, pp. 193–268.

[91] John Duns Scotus (now attributed to Pseudo-Scotus), *In Libros Priorum Analyticorum Aristotelis Quaestiones*, in *Ioannis Duns Scoti Doctoris Subtilis Ordinis Minorum Opera Omnia* Vol. I, ed. L. Wadding, Lyon 1639.

[92] Stewart Shapiro, 'Classical logic', *The Stanford Encyclopedia of Philosophy*, (Winter 2013 Edition), ed. Edward N. Zalta, `http://plato.stanford.edu/entries/logic-classical`.

[93] Nabil Shehaby, *The Propositional Logic of Avicenna*, Reidel, Dordrecht 1973.

[94] William of Sherwood, *Einführung in die Logik. Introductiones in Logicam*, ed. Hartmut Brands and Christoph Kann, Meiner, Hamburg 2013. Translated in [70].

[95] Joseph R. Shoenfield, *Mathematical Logic*, Addison-Wesley, Reading Mass. 1967.

[96] Tony Street, 'Avicenna and Tusi on the Contradiction and Conversion of the Absolute', *History and Philosophy of Logic* 21 (1) (2000) 45–56.

[97] Tony Street, 'An outline of Avicenna's syllogistic', *Archiv für Geschichte der Philosophie* 84 (2), 129–160 (2002).

[98] Tony Street, 'Arabic logic', *Handbook of the History of Logic* ed. Dov M. Gabbay and John Woods, Elsevier 2004, pp. 523–596.

[99] Paul Thom, *The Syllogism*, Philosophia Verlag, Munich 1981.

[100] Naṣīr al-Dīn al-Ṭūsī, *Asās al-iqtibās* (Persian), Ferdows, Tehran 2010.

[101] Robin Wilson, *Introduction to Graph Theory*, Pearson, Harlow 1996.

[102] F. Zabeeh, *Avicenna's Treatise on Logic: Part One of Danesh-name Alai (a concise philosophical encyclopaedia and autobiography)*, Nijhoff, The Hague 1971.

Arabic terms

English terms

Cited passages of Ibn Sīnā

People