# Matrices and Modalities:
# On the Logic of Two-Dimensional Semantics

**MSc Thesis** (*Afstudeerscriptie*)

written by

**Peter Fritz**
(born March 4, 1984 in Ludwigsburg, Germany)

under the supervision of **Dr Paul Dekker** and **Prof Dr Yde Venema**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

**MSc in Logic**

at the *Universiteit van Amsterdam*.

**Date of the public defense:**
*June 29, 2011*

**Members of the Thesis Committee:**
Dr Paul Dekker
Dr Emar Maier
Dr Alessandra Palmigiano
Prof Dr Frank Veltman
Prof Dr Yde Venema

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

**Abstract**

Two-dimensional semantics is a theory in the philosophy of language that provides an account of meaning which is sensitive to the distinction between necessity and apriority. Usually, this theory is presented in an informal manner. In this thesis, I take first steps in formalizing it, and use the formalization to present some considerations in favor of two-dimensional semantics. To do so, I define a semantics for a propositional modal logic with operators for the modalities of necessity, actuality, and apriority that captures the relevant ideas of two-dimensional semantics. I use this to show that some criticisms of two-dimensional semantics that claim that the theory is incoherent are not justified. I also axiomatize the logic, and compare it to the most important proposals in the literature that define similar logics. To indicate that two-dimensional semantics is a plausible semantic theory, I give an argument that shows that all theorems of the logic can be philosophically justified independently of two-dimensional semantics.

# Acknowledgements

I thank my supervisors Paul Dekker and Yde Venema for their help and encouragement in preparing this thesis. I am grateful for Paul's advice on philosophical matters and Yde's advice on logical matters, but most of all, I thank them for their willingness to discuss issues that are both logical and philosophical. I also thank Emar Maier, Alessandra Palmigiano, and Frank Veltman for being on my thesis committee and for their questions and comments during the defense.

For helpful discussions on two-dimensional semantics, I thank the members of the reading group on two-dimensional semantics at the Australian National University in 2009. I am also grateful for comments by the audiences at my talks at the *Graduate Conference on Meaning, Modality and Apriority*, University of Cologne, May 2010, and the *Logic Tea*, University of Amsterdam, April 2011.

In addition to the above, I thank David Chalmers, Kasper Højbjerg Christensen, Catarina Dutilh Novaes, Bruno Jacinto, Thomas Krödel, Johannes Marti, Greg Restall, David Ripley, Laura Schroeter, and Wolfgang Schwarz for discussions and comments on the material in this thesis. I am especially grateful to Robert Michels for comments on a draft of the thesis as well as many valuable discussions.

The appendix is based on an essay written for the course *Philosophy of Logic* at the University of Amsterdam in spring 2011.

# Contents

# Chapter 1

# Introduction

Two-dimensional semantics is a theory in the philosophy of language that deals explicitly with the notions of necessity and apriority. The aim of this thesis is to construct and discuss a logic that captures some of the essential features of this theory, and use it to draw philosophical conclusions about two-dimensional semantics.

In this chapter, I will introduce the necessary philosophical background. I will start by explaining the distinction between necessity and apriority as it is drawn since Kripke (1980 [1972]), and the semantics of indexicals as it was developed in Kaplan (1989b). This will be followed by an introduction to the central ideas of two-dimensional semantics. For reasons of brevity, these explanations will be rather sketchy, but I will provide references to more detailed treatments. In the final section of this chapter, I will give an overview of the thesis and its aims.

## 1.1 Kripke on Necessity and Apriority

In Kripke (1980 [1972]), Kripke argues that proper names do not function semantically like abbreviated descriptions, a view often attributed to Russell (1905) and Frege (1892). This leads him to the claim that the notions of necessity and apriority are distinct, and do not even have the same extension. For this thesis, only Kripke's claim about necessity and apriority is central, so I will focus on this, without going into detail about his work on proper names.

### 1.1.1 The Concepts of Necessity and Apriority

Kripke (1980 [1972], pp. 34–38) notes that although the notions of necessity and apriority have often been treated as being interchangeable, they are *prima facie* distinct notions. Following Kant, Kripke calls those truths *a priori* that can be known independently of any experience. Although he leaves it open what kind of modality "can" expresses in this analysis, it is clear that the notion of apriority essentially deals with knowledge, and is therefore an epistemological notion. To explain the notion of necessity, he says that those truths are necessary that could not have been false. This explanation is not very informative, as it uses the modal expression "could", which may be just as unclear as "necessary". But the explanation makes clear that in contrast to the notion of apriority, the notion of necessity is not about states of knowledge, but about how *things* could be, that is, about states of the world. Therefore, necessity is not an epistemological notion. Kripke calls necessity a *metaphysical* notion, and in the philosophical literature, this specific conception of necessity is also often referred to as "metaphysical necessity".

As acknowledged by Kripke, the intelligibility of this notion of necessity has been called into question repeatedly, e.g in Quine (1976 [1953]). Although the notion is used in many discussions in contemporary analytical philosophy, there are still reasons to be skeptical about it, a point that is made e.g. in Fine (2005, p. 7). In the following, I will just assume that it is intelligible. However, this does not mean that if it isn't, all of the following work is irrelevant. Even if there is no clear concept of metaphysical necessity, many of the issues raised by Kripke and discussed in the following can be made if instead of metaphysical necessity, other non-epistemic notions of necessity are used, such as nomological necessity (necessity according to the laws of nature). See, e.g., Edgington (2004) for such an interpretation of Kripke's discussion.

### 1.1.2 Necessary A Posteriori and Contingent A Priori

Beyond arguing that necessity and apriority are distinct notions, Kripke also presents examples of truths for which these notions do not coincide. As usual, I will call a truth or falsity *contingent* if neither it nor its negation is necessary, and *a posteriori* if neither it nor its negation is a priori. Kripke's examples are therefore cases of necessary a posteriori or contingent a priori truths.

One of his examples for a truth that is necessary and a posteriori is

(HP) Hesperus is Phosphorus.

"Hesperus" is the Greek name of the first celestial body visible in the evening sky, and "Phosphorus" is the Greek name of the last celestial body visible in the morning sky. In fact, both are the planet Venus. (For simplicity, I'm not counting the moon among the celestial bodies.) That (HP) is a posteriori is very plausible: the only way to find out that the celestial bodies are one and the same is by empirical observation. That (HP) should be necessary might be more surprising. Kripke argues as follows: Hesperus and Phosphorus are in fact identical. For (HP) to be contingent, it would have to be possible that Hesperus and Phosphorus are not identical. But it is impossible for this to be the case, since one object cannot possibly not be identical to itself. As it stands, this reasoning may sound spurious, but Kripke substantiates it by an explanation of how we evaluate modal claims involving proper names. This is related to Kripke's other main claim, the one concerning the semantics of proper names.

Kripke's main example for a truth that is contingent and a priori is

(MS) The stick that was used to define the unit of measurement *meter* was one meter long at the time of definition.

Kripke argues that (MS) is a priori, since no empirical investigations are necessary to establish its truth – in particular, we do not have to measure the particular stick at any time. However, someone could have cut off a piece of the stick prior to the time of definition, which would have made it less than a meter long. So the stick might have been shorter than one meter, and therefore (MS) is not necessary.

These examples conclude my brief sketch of Kripke's position on necessity and apriority.

## 1.2 The Semantics of Indexicals

Besides the ideas by Kripke which I have just described, the account of necessity and apriority given by two-dimensional semantics is deeply influenced by the most widely accepted semantics of indexicals, which is given in Kaplan (1989b). Indexicals are expressions whose meaning is sensitive to the context in which they are uttered. E.g. "I" is an indexical since it refers to whoever uttered it, which varies between contexts. Other examples are "now", "yesterday", "here", "that", and "he".

### 1.2.1 A General Theory

Kaplan's semantics builds on intensional semantics, where the meaning of an expression is modeled as a function from points of evaluation to extensions. What the extensions of expressions are depends on the syntactic type of the expression. E.g. a proper name would be given an individual as its extension at a point of evaluation, a predicate a set of objects, and a sentential expression a truth value. What points of evaluation represent will depend on the intensional features of the fragment of the language that is modeled. If we want to model temporal expressions such as "it has been the case that", it will be points in time; if we want to model modal expressions such as "possibly", it will be possible worlds. For simplicity, I focus here on the latter case.

As noted above, what an indexical expression refers to depends on the context in which it is uttered. We can think of a context of utterance as a possible world, in which a spatio-temporal location is singled out as the location of the utterance. These are often called *centered worlds*. The central idea of Kaplan's semantics is to relativize the picture of intensional semantics to contexts of utterance. So he models the meaning of an expression by a function that takes a context of utterance and returns a function from possible worlds to extensions, as before. It is useful to distinguish two layers of meaning at this point: the function representing the meaning of *expressions* that was just mentioned is called its *character* by Kaplan, and the function returned by a character for a certain context of utterance he calls its *content*. Therefore, the character is a function from contexts of utterance to contents.

Characters can be illustrated nicely by tables in which the rows represent utterance contexts, the columns possible worlds, and the cells contain the appropriate extension. E.g., consider the indexical singular term "I", and the predicate "happy". In a very simple model, we might have two possible worlds, $w$ and $v$. Further, we might have two contexts of utterance $w^*$ and $v^*$, where $w^*$ is $w$, centered on Mary, and $v^*$ is $v$, centered on John. If Mary is happy in $w$ and $v$, but John is only happy in $w$, then we model the meanings of the expressions "I" and "happy" by functions that can be displayed as matrices as follows, where T stands for *true* and F for *false*:

| "I" | | |
|---|---|---|
| | $w$ | $v$ |
| $w^*$ | Mary | Mary |
| $v^*$ | John | John |

| "happy" | | |
|---|---|---|
| | $w$ | $v$ |
| $w^*$ | {Mary, John} | {Mary} |
| $v^*$ | {Mary, John} | {Mary} |

| "I am happy" | | |
|---|---|---|
| | $w$ | $v$ |
| $w^*$ | T | T |
| $v^*$ | T | F |

### 1.2.2 Indexical Operators

In the philosophical discussion of indexicals, two expressions are important that can be understood as unary sentential operators: "now" and "actually". "Actually" will be central in the following, but it is easiest to understand its intended meaning in analogy to "now". As an example for the use of "now" as a sentential operator, consider the following sentence:

(NH) It will at some point be the case that everyone who is now sad is happy.

On one natural reading of (NH), the quantifier "everyone" ranges over individuals existing at some future moment, but restricting it to the individuals that are now sad requires us to evaluate whether they are sad at the moment of utterance of (NH).

As proposed in Lewis (1970), "actually" can be seen as a modal analog to "now". So while modal expressions like "necessarily" and "possibly" require us to evaluate the expressions they operate on in other possible worlds, the function of "actually" is to evaluate what it operates on at the world of utterance. In analogy to (NH), we can illustrate this with

(AH) It could have been the case that everyone who is actually sad is happy.

As before, while the expression "it could have been that" takes us to alternative possibilities, to see whether some individual is actually sad, we have to consider whether they are sad in the actual world.

We can give a more systematic account of this in Kaplan's semantics. Here, I treat "necessarily" and "actually", assuming that the points of evaluation of the formal model are possible worlds. The basic idea is of course that necessity is truth in all possible worlds, and actuality is truth in the world of utterance. Let $p$ be a sentential expression, and $c$ its character.

(N) At any context of utterance $u$, the content of "necessarily $p$" is a constant function to *true* if $c(u)$ is a constant function to *true*, and a constant function to *false* otherwise.

(A) At any context of utterance $u$, the content of "actually $p$" is a constant function to *true* if $c(u)(w)$ is *true*, and a constant function to *false* otherwise, where $w$ is the possible world of $u$.

As before, this is best illustrated using matrices, using some worlds $w, v, u$ and centerings $w^*, v^*, u^*$ of these worlds:

| $p$ | | | |
|---|---|---|---|
| | $w$ | $v$ | $u$ |
| $w^*$ | T | T | T |
| $v^*$ | T | T | F |
| $u^*$ | T | F | F |

| "necessarily $p$" | | | |
|---|---|---|---|
| | $w$ | $v$ | $u$ |
| $w^*$ | T | T | T |
| $v^*$ | F | F | F |
| $u^*$ | F | F | F |

| "actually $p$" | | | |
|---|---|---|---|
| | $w$ | $v$ | $u$ |
| $w^*$ | T | T | T |
| $v^*$ | T | T | T |
| $u^*$ | F | F | F |

It should be noted that historically, the development of the logic and semantics of "now" and "actually" preceded the general development of the semantics of indexicals as described above, e.g., in papers like Prior (1968a) and Kamp (1971).

Note that if a matrix has *true* in every cell on the diagonal from the upper left to the lower right, then the sentence is true in the world of utterance at every utterance context; we might say that the sentence is *true-whenever-uttered.*

### 1.2.3 Logics of Indexicals

So far, I have presented a general picture of the semantics of indexicals and applied it to the indexical operators "now" and "actually". It is not difficult to see how this can be formalized, e.g. in something similar to intensional type theory (as in Gallin (1975)), in the sense of giving a formal definition of syntax, models, and an interpretation relation between them, except that this relation is now relativized to two indices. Something of the sort is in fact done in Kaplan (1989b).

However, to have a *logic* of indexicals, a definition of the consequence relation is needed. There are two natural proposal for this: firstly, we could say that a conclusion follows from some premises if the conclusion is true in every model at every context of utterance and possible world at which all the premises are true. Secondly, we could say that a conclusion follows from some premises if the conclusion is true in every model at every context of utterance and *its* world at which all the premises are true. Adapting terminology from Crossley and Humberstone (1977), I call the first *general consequence*, and the second *real-world consequence.* Given that a formula is valid iff it is a consequence of the empty set, this also gives us two corresponding notions of validity, which we can call *general validity* and *real-world validity.*

An important difference between the two is that arguments that follow by the semantics of indexicals are declared logically valid by real-world consequence, but not by general consequence. The following is an example for such an argument: It is raining now; therefore it is raining. Real-world consequence is mostly accepted as the correct definition, see e.g. Vlach (1973), Kaplan (1989b), or Williamson (2006), but some disagree. In the following, I will also assume real-world consequence. In the appendix, I discuss this matter in detail and argue for the choice of real-world consequence.

## 1.3 Two-Dimensional Semantics

I will now introduce two-dimensional semantics. The word "two-dimensional semantics" can be used for a number of related ideas. E.g., Schroeter (2010) uses it for any kind of double-indexing semantics, including Kaplan's semantics of indexicals, as well as the pragmatic theory described in Stalnaker (1978). Here, I will use "two-dimensional semantics" more restrictedly as referring to the theory called "epistemic two-dimensional semantics" in Chalmers (2004). This theory was first proposed in Chalmers (1996), and is most fully developed in Chalmers (2006).

### 1.3.1 A Sketch of the Theory

Using the last two sections, we can understand two-dimensional semantics as giving an account of Kripke's observations on the distinction between necessity and apriority using a semantic framework inspired by Kaplan's semantics for indexicals. In fact, the connections between two-dimensional semantics and these two topics runs much deeper, but to fully explain them, I would have to say much more, especially about Kripke's and Kaplan's views on reference. Since this will not play a big role in the following, I will skip these issues.

To motivate two-dimensional semantics, it is best to apply the picture of intensional semantics as mentioned above to Kripke's example (HP) of a necessary and a posteriori truth. Since (HP) is necessary, according to intensional semantics, its meaning is the constant function to *true*. The following is also a necessary truth:

(HH) Hesperus is Hesperus.

And since it is also necessary, its meaning is also the constant function to *true*. Therefore, according to intensional semantics, (HP) and (HH) have the same meaning. But (HH) is obviously a priori, while (HP) is not. This poses a problem: how can two truths have the same meaning, while one is a priori and the other is not? Note that since neither (HP) nor (HH) contains indexicals, complicating the picture by using characters instead of intensions does not change anything – (HP) and (HH) have the same constant character that maps every context to the constant function to *true*.

Two-dimensional semantics takes this to show that there is more to meaning than the function from possible worlds to extensions postulated by intensional semantics. In particular, it claims that there is also an *epistemic* component of meaning. According to two-dimensional semantics, the familiar intension is only the *metaphysical* component of the meaning, which is called the *secondary intension*. The epistemic component is to be captured with another function, the *primary intension*. This primary intension is to stand in the same relation to apriority as the secondary intension stands in relation to necessity. Therefore, a statement is a priori iff its primary intension is a constant function to *true*. So, a possible worlds analysis of apriority is needed. Therefore, the primary intension is a function from epistemically possible worlds to extensions. If we understand metaphysically possible worlds (henceforth just called *worlds*) as *ways the world could have been*, then we can analogously understand epistemically possible worlds (henceforth called *scenarios*) as *ways the world can turn out to be, given what can be known a priori*.

What kinds of things are scenarios? Maybe surprisingly, one version of two-dimensional semantics claims that they are centered possible worlds. (Another claims that they are maximal hypotheses in an idealized language, but I don't consider this option here. See Chalmers (2004, section 3.4) for more on the two versions.) This proposal raises a further question: what determines the extension of an expression at a scenario? One answer is to say that every expression is associated with a cognitive or conceptual role, and the extension of an expression at a scenario is given by what would fit that role if we would be in the respective scenario. Since a centered world can be seen as a world (a metaphysical possibility) by ignoring the centering, as well as a scenario (an epistemic possibility), it is useful to introduce the following terminology: if we talk about a (centered) world as a metaphysical possibility, we say that we consider it *as counterfactual*. If we talk about a centered world as an epistemic possibility, we say that we consider it *as actual*.

So far, we have only evaluated expressions relative to a scenario (considered as actual) and relative to a world (considered as counterfactual). But to get to the full semantic picture of two-dimensional semantics, we also have to combine these, and evaluate expressions relative to a scenario *and* a world. Roughly speaking, the scenario considered as actual fixes the referents of the simple components, whereas the world considered as counterfactual is the world relative to which we evaluate the expression, given these referents. This gives us a natural way of associating with each expression a function from scenario/world pairs to extensions, which we can call its *two-dimensional matrix*. From this matrix, we can recover the primary as well as the secondary intension. For the second, we just keep our current scenario fixed, and vary the world considered as counterfactual. For the first, we consider each scenario as actual as well as counterfactual.

Structurally, this picture is very similar to Kaplan's semantics of indexicals. We still have possible worlds, but now we have scenarios instead of utterance contexts. What was the character is now the two-dimensional matrix. Although the former is a function to functions and the second is a function taking two arguments, these representations can obviously be converted into each other (one direction is sometimes called *Schönfinkeling* or *Currying*). Given the understanding of scenarios as centered possible worlds, the semantics of indexicals straightforwardly integrates into two-dimensional semantics. That is, we can take the matrix of indexical expressions to be given by their characters. In particular, I will make use of this for the semantics of "actually", as described above.

Furthermore, just like characters, two-dimensional matrices can be visualized as tables, where the scenarios are listed on the vertical and worlds on the horizontal dimension. Given this picture, necessity turns out to be truth throughout the horizontal (of the current scenario), and apriority truth throughout the diagonal of the matrix. So what was truth-whenever-uttered before is now apriority. Just as possibility is generally understood as the dual of necessity – something is possible iff its negation is not necessary – according to two-dimensional semantics, *conceivability* is understood as the dual of apriority – something is conceivable iff its negation is not a priori. Therefore, something is possible iff there is a world considered as counterfactual in which it is true, and conceivable iff there is a scenario considered as actual in which it is true.

Before illustrating the sketch of two-dimensional semantics I have just given with an example, I want to point out that many aspects of the theory are controversial. E.g., it is very natural to wonder why truth on the diagonal of the two-dimensional matrix should coincide with apriority – see Chalmers (2006, section 3) for an extended discussion of this.

### 1.3.2 "Hesperus is Phosphorus" in 2D

It may be helpful to illustrate the ideas of two-dimensional semantics using Kripke's example of "Hesperus is Phosphorus" for a necessary a posteriori truth. Let $w$ be the actual world, and $v$ a possible world in which Mars is the first celestial body visible in the evening sky, and Venus is the last celestial body visible in the morning sky. Let $w^*$ and $v^*$ be scenarios based on these worlds with a center on Earth. Then from the perspective of $v^*$, Mars fits the cognitive role associated with "Hesperus", and Venus fits the role associated with "Phosphorus". Hence in this small model, we get the following matrices:

"Hesperus"

| | $w$ | $v$ |
|---|---|---|
| $w^*$ | Venus | Venus |
| $v^*$ | Mars | Mars |

"Phosphorus"

| | $w$ | $v$ |
|---|---|---|
| $w^*$ | Venus | Venus |
| $v^*$ | Venus | Venus |

"Hesperus is Phosphorus"

| | $w$ | $v$ |
|---|---|---|
| $w^*$ | T | T |
| $v^*$ | F | F |

Since $w^*$ is our current scenario, the secondary intension of "Hesperus is Phosphorus" is the constant function to *true*, which represents the fact that it is necessary. But the diagonal of its matrix contains a *false*, which represents the fact that it is not a priori.

### 1.3.3 Two-Dimensional Modal Logic?

It is sometimes remarked that two-dimensional semantics is in some way based on "two-dimensional modal logic", e.g., in Soames (2006). This may give the impression that there already is a logic for two-dimensional semantics, and therefore no such thing needs to be developed. However, this is not the case. There are three main strands of research that could be meant when using the phrase "two-dimensional modal logic" in such a claim, but none of them is concerned with capturing the central claim of two-dimensional semantics about the connection between necessity and apriority. The first of these is the logic of indexicals as discussed above; the second is the logic presented in Davies and Humberstone (1980); and the third is the use of double-indexing in Lewis (1973), Åqvist (1973), and Segerberg (1973).

As is clear from the above exposition, logics for indexical operators were not developed to formalize two-dimensional semantics. Furthermore, they could not be used for this, e.g. because they do not treat the central notion of apriority. Davies and Humberstone (1980) is also not meant to capture the notion of apriority, but it does formalize the related notion of deep necessity. Therefore, it is not directly applicable as a formalization of two-dimensional semantics, although it can be used for this with certain restrictions. I discuss this in more detail in section 4.1. Finally, the texts by Lewis, Åqvist and Segerberg are neither concerned with indexicality nor with apriority, but with counterfactuals. Lewis and Åqvist use double-indexing to solve a problem with the so-called operator analysis of counterfactuals, and Segerberg discusses technical problems posed by Åqvist's paper. It is clear from this that they do not intend to formalize two-dimensional semantics. In fact, the only reason I refer to them is because they are often mentioned in connection with two-dimensional semantics, e.g. in Davies and Stoljar (2004) and Schroeter (2010, section 1.2). Besides these three uses of the phrase "two-dimensional modal logic" I have just presented, there are some others, but it is easy to see that they are not relevant, as discussed in Humberstone (2004, section 1).

The only other text that is primarily concerned with a formalization of two-dimensional semantics is Restall (2010). Restall's paper is in fact very close to the research in this thesis. Specifically, it presents a semantics which is essentially the same as the one I will use, although it uses a completely different approach to its proof theory. I discuss Restall's paper and the reasons for developing a different proof system in section 4.2.

## 1.4 Formalizing Two-Dimensional Semantics

I will now describe which kind of logic I want to develop for two-dimensional semantics, and why I think that such an enterprise is interesting.

### 1.4.1 Choice of Language

The most important choice is that I will only consider propositional logics. The reason for this is just simplicity: it is sensible to start as simple as possible, and a propositional logic is much simpler than a quantified one. In the concluding chapter 7, I will briefly discuss the possibility of extending this to a quantified logic and explain why this would be philosophically interesting.

For a propositional language, we naturally start with propositional letters representing unanalyzed sentential expressions, as well as Boolean operators. Since the central concepts for two-dimensional semantics are necessity and apriority, it is natural to add these as unary modal operators. Furthermore, an important element of two-dimensional semantics is the relation of indexicals to these notions, which motivates the inclusion of an indexical expression. Given the existing literature on "actually" mentioned earlier, it is natural to choose "actually" as a representative indexical. For most parts of this thesis, this will be the language I will be working with: a propositional language with Boolean connectives and three unary sentential operators representing necessity, apriority, and actuality.

### 1.4.2 Aims of the Formalization

There are a number of reasons for constructing logics for two-dimensional semantics. Firstly, critics of two-dimensional semantics have suggested that the two-dimensionalist's claim that apriority and necessity operate on different intensions (i.e. primary and secondary) may give an incoherent semantic account. Schroeter (2010, section 2.4.2) provides references to works in which this worry is raised. A logic such as the one discussed below can be seen as a formalization of a fragment of English, and as such, it indicates the coherence of the basic ideas of two-dimensional semantics.

One specific criticism of two-dimensional semantics of this form is the so-called nesting problem, which derives from an argument in Soames (2005, pp. 278–279). Its core is the observation that two-dimensional semantics seems to carry a commitment to the instance of (1) and (2) for any sentence $p$, and that under the ordinary logical assumption that necessity distributes over implication, they entail the corresponding instance of (3):

(1) If it is a priori that $p$, then it is necessarily a priori that $p$.

(2) Necessarily, if it is a priori that $p$, then $p$.

(3) If it is a priori that $p$, then it is necessary that $p$.

Given the existence of contingent a priori truths, (3) must be rejected, so this constitutes an argument against two-dimensional semantics. A logic of two-dimensional semantics shows which of the premises of this argument should be rejected by a proponent of two-dimensional semantics, and thereby provides a principled way of responding to the nesting problem. This will be discussed in detail in section 3.2 and chapter 6.

A second reason for defining a logic of the kind indicated could be an interest in the interactions between the notions of necessity, apriority, and actuality, independently of two-dimensional semantics. Since two-dimensional semantics is one of the most explicit theories of how these notions relate, it is a natural candidate to formalize, among other such theories. I will discuss some of these interactions in section 3.5.

Thirdly, one can test the plausibility of two-dimensional semantics by investigating its commitments concerning the interactions between necessity, apriority, and actuality. This is best done using a formal logic, and it will be one of the central topics of this thesis, and the main concern of chapters 5 and 6.

### 1.4.3 Overview of the Thesis

In chapter 2 (*A Logical Toolkit*), I provide the necessary background in modal logic to develop the logic for two-dimensional semantics. Depending on how familiar this material is, it may be best to skip or skim this chapter, and use it as a reference. To facilitate this, the thesis contains a list of notation after the bibliography. The central chapter is chapter 3 (*A Logic for Two-Dimensional Semantics*). In it, I define the semantics of the logic of two-dimensional semantics, and apply it to the nesting problem that was described in the last section. Furthermore, I axiomatize it, and describe some of its properties. In chapter 4 (*Comparisons*), I compare this logic to the systems in Davies and Humberstone (1980) and Restall (2010). The rest of the thesis is mainly concerned with an argument that shows that the logic of two-dimensional semantics as defined in chapter 3 is plausible independently of two-dimensional semantics. First, in chapter 5 (*Logical Commitments of Two-Dimensional Semantics*), I compare the logic to a minimal logic that should be commonly accepted in philosophy to determine the additional logical commitments of two-dimensional semantics. Moreover, I argue that all of these commitments except one formula are plausible independently of two-dimensional semantics. It turns out that this remaining formula is the one representing premise (1) in the nesting argument. In chapter 6 (*The Nesting Problem*), I discuss the nesting problem independently of two-dimensional semantics, and argue that accepting (1) and rejecting (2) is the correct solution to it. Together with chapter 5, this shows that all theorems of the logic of two-dimensional semantics are independently plausible. Chapter 7 (*Conclusion*) sums up the results of the thesis and provides an outlook on possible ways to extend the logic. The appendix (*Consequence*) deals with general and real-world consequence, and argues for the latter.

# Chapter 2

# A Logical Toolkit

This chapter provides the definitions and facts concerning propositional modal logics that will be needed to develop the logic of two-dimensional semantics. Most research on modal logics today focuses on normal modal logics. But as is well-known, given the definition of real-world validity, a logic modeling "actually" may well be non-normal. Informally, although "If it is raining, then it is actually raining" is a logical truth, its generalization "Necessarily, if it is raining, then it is actually raining" is not. This violates the condition on normal modal logics that they must be closed under the rule of generalization. Therefore, I will also treat a class of modal logics that do not have to satisfy this condition, called *quasi-normal modal logics*.

Most of the theory described here can be found in standard textbooks, and therefore, I will not present it in full detail. I will adapt much from Blackburn et al. (2001), although in the final section, I will prove a new completeness theorem. In the rest of the thesis, I will mainly work with a specific language containing only three operators for necessity, apriority, and actuality. In this chapter, I will be a bit more general, and work with an arbitrary set $\alpha$ of unary modal operators.

## 2.1 Modal Logics

I start with a syntactic definition of the relevant classes of logics and corresponding proof systems.

### 2.1.1 Logics and Proof Systems

The *language* $\mathcal{L}$ is constructed inductively from proposition letters using Boolean operators and the modal operators in $\alpha$. For the purposes of proofs by induction, I will assume that $\neg$ and $\wedge$ are the primitive Boolean operators, and that all others are defined in terms of them. The primitive modal operators in $\alpha$ are all interpreted universally (i.e., as "boxes"). The elements of $\mathcal{L}$ are called *formulas*.

In later sections and chapters, I will work with different languages at the same time. In these cases, I will resolve ambiguities by specifying the set of operators as a subscript. E.g., when talking about two sets of operators $\alpha$ and $\beta$, the corresponding languages are $\mathcal{L}_\alpha$ and $\mathcal{L}_\beta$. This notation will also be used for many of the definitions given below. If the sets are given explicitly by listing operators, I will just list the operators instead of writing down the set. E.g., I will write $\mathcal{L}_\Box$ instead of $\mathcal{L}_{\{\Box\}}$.

A *logic* is a set of formulas. A *proof system* is given by a set of axioms and a set of rules. A formula $\varphi$ is a *theorem* of a proof system $P$ if $\varphi$ can be derived from the axioms of $P$ using the rules of $P$. (Obviously, I'm not being completely precise here – e.g., there must be an effective method of deciding whether a formula is an axiom. Any proof system considered here will be sufficiently standard that this will not be an issue.) The *logic* of $P$, written $L(P)$, is the set of theorems of $P$.

Let $\Lambda$ be a logic. A formula $\varphi$ is a $\Lambda$-*theorem*, written $\vdash_\Lambda \varphi$, if it is an element of $\Lambda$. $\varphi$ *follows* from a set of formulas $\Gamma$ in a logic $\Lambda$, written $\Gamma \vdash_\Lambda \varphi$, if there are $\psi_1, \dots \psi_n \in \Gamma$ such that $\vdash_\Lambda \left(\bigwedge_{i \leq n} \psi_i\right) \to \varphi$. If $P$ is a proof system, I will write $\vdash_P$ instead of $\vdash_{L(P)}$. A set of formulas is $\Lambda$-*consistent* if $\bot$ does not follow from it in $\Lambda$, and a $\Lambda$-*maximal consistent set* ($\Lambda$-MCS) if it is $\Lambda$-consistent but has no proper $\Lambda$-consistent extension. Formulas $\varphi$ and $\psi$ are $\Lambda$-*equivalent* if $\vdash_\Lambda \varphi \leftrightarrow \psi$.

### 2.1.2 Normal Modal Logics

A *normal modal logic* (NML) is a logic that includes all propositional tautologies and $K_\Box$ for all $\Box \in \alpha$, and is closed under the rules of modus ponens (*MP*), uniform substitution (*US*), and generalization (*Gen*), where these are the following:

- $K_\Box = \Box(p \to q) \to (\Box p \to \Box q)$
- *MP*: From $\varphi$ and $\varphi \to \psi$, derive $\psi$.
- *US*: From $\varphi$, derive any substitution instance of $\varphi$. (A substitution instance of $\varphi$ is the result of replacing every occurrence of some proposition letter in $\varphi$ uniformly by some formula.)
- *Gen*: From $\varphi$, derive any $\Box\varphi$ for $\Box \in \alpha$.

For any set of formulas $\Gamma$, the NML *axiomatized* by $\Gamma$, written $\oplus\Gamma$, is the smallest NML containing $\Gamma$. $N\Gamma$ is the proof system that contains as axioms the propositional tautologies, $K_\Box$ for all $\Box \in \alpha$, and the members of $\Gamma$, and as rules *MP*, *US*, and *Gen*. It is easy to see that $\oplus\Gamma = L(N\Gamma)$. **K** is defined to be the smallest normal modal logic.

The following fact is very useful: if $\Lambda$ is an NML and $\psi$ and $\chi$ are $\Lambda$-equivalent, then $\varphi$ is $\Lambda$-equivalent to any formula $\varphi'$ obtained by replacing an occurrence of $\psi$ in $\varphi$ by $\chi$. See, e.g., Hughes and Cresswell (1996, pp. 32–33, 46) for a proof. I mention this explicitly since it fails for the class of logics defined in the next section.

### 2.1.3 Quasi-Normal Modal Logics

A *quasi-normal modal logic* (QNML) is a logic that includes **K** and is closed under *MP* and *US*. Note that every NML is a QNML.

Similar to NMLs, the QNML *axiomatized* by a set of formulas $\Gamma$, written $+\Gamma$, is the smallest QNML containing $\Gamma$. $QN\Gamma$ is the proof system that contains as axioms the members of $\mathbf{K}$ and $\Gamma$, and as rules *MP* and *US*. Again, it is easy to see that $+\Gamma = L(QN\Gamma)$.

Although the replacement of equivalents does not hold in QNMLs, we get a similar, but somewhat weaker result. To state it, we define the *kernel of* a QNML $\Lambda$ to be $\ker(\Lambda) = \{\varphi : \heartsuit\varphi \in \Lambda \text{ for all } \heartsuit \in \alpha^*\}$, where $\alpha^*$ is the set of finite sequences on $\alpha$. We can then show the following:

**Proposition 2.1.** *Let $\Lambda$ be a* QNML. *If $\psi$ and $\chi$ are $\ker(\Lambda)$-equivalent, then $\varphi$ is $\Lambda$-equivalent to any formula $\varphi'$ obtained by replacing an occurrence of $\psi$ in $\varphi$ by $\chi$. In particular, if $\vdash_\Lambda \varphi$ then $\vdash_\Lambda \varphi'$.*

*Proof.* The set $\ker(\Lambda)$ is an NML. So as noted above, if $\psi$ and $\chi$ are $\ker(\Lambda)$-equivalent, then $\varphi$ and $\varphi'$ are $\ker(\Lambda)$-equivalent. Since $\ker(\Lambda) \subseteq \Lambda$, it also follows that $\varphi$ and $\varphi'$ are $\Lambda$-equivalent. So if $\vdash_\Lambda \varphi$, it follows by modus ponens that $\vdash_\Lambda \varphi'$. □

Note that for any QNML $\Lambda$, $\mathbf{K} \subseteq \ker(\Lambda)$. Hence replacement of $\mathbf{K}$-equivalents preserves theoremhood in any QNML.

### 2.1.4 Joins of Sets of Formulas

At a number of points in the following chapters, there will be two sets of formulas representing some logical commitments, and it will be interesting to say something about what happens if they are combined. For this, the notion of a minimal combination is helpful. The most straightforward formal representation of such a combination is the smallest logic (of the relevant kind) that contains both of the original sets of formulas. This motivates the following definitions:

The *normal join* of two sets of formulas $\Gamma$ and $\Delta$, written $\Gamma \oplus \Delta$, is the smallest normal modal logic containing both $\Gamma$ and $\Delta$. Similarly, the *quasi-normal join* $\Gamma + \Delta$ is the smallest quasi-normal modal logic containing them.

The name "join" for these is motivated by the fact that the set of NMLs ordered by the subset relation is a lattice, and $\oplus$, restricted to the members of this lattice, is its join operation. The analogous observation holds for QNMLs and $+$; see Chagrov and Zakharyaschev (1997, p. 113). Normal joins in which both of the sets are normal modal logics in languages with disjoint sets of operators are called *fusions* in the literature, see e.g. Kracht and Wolter (1991) or Gabbay et al. (2003, chapter 4).

These binary joins can also be generalized to joins of multiple sets of formulas. I will write $\oplus(\Gamma_1, \ldots, \Gamma_n)$ for the smallest NML containing $\Gamma_1, \ldots, \Gamma_n$ and $+(\Gamma_1, \ldots, \Gamma_n)$ for the smallest QNML containing $\Gamma_1, \ldots, \Gamma_n$.

A number of properties of the join operators are easily established. E.g., it follows immediately from the definitions that they are commutative, and it is also not difficult to see that they are associative. The following is another property that will be useful in the following:

**Proposition 2.2.** *If $\beta \subseteq \alpha$, then $(\Gamma +_\beta \Delta) +_\alpha \Theta =+ _\alpha(\Gamma, \Delta, \Theta)$.*

*Proof.* Note that these are the logics of the proof systems $QN_\alpha(L(QN_\beta(\Gamma \cup \Delta)) \cup \Theta)$ and $QN_\alpha(\Gamma \cup \Delta \cup \Theta)$. By inductions on these proof systems, we can show that the logics are the same. □

## 2.2 Relational Semantics

As usual in philosophical discussions of modal logic, I will primarily deal with its relational semantics, which is sometimes also called *Kripke semantics*. It will turn out that the most natural way of formalizing the treatment of necessity, actuality and apriority by two-dimensional semantics is as a class of frames if we assume general validity, and as a class of frames with distinguished elements if we assume real-world validity. Frames with distinguished elements are a generalization of frames; this is a standard tool in the literature on quasi-normal modal logics. In this section, I define both kinds of semantic structures, and some useful transformations between them. To mark the difference between the two kinds, I will use different fonts; e.g. $\mathfrak{F}$ and $\mathfrak{M}$ for regular structures, and $\mathcal{F}$ and $\mathcal{M}$ for structures with distinguished elements.

### 2.2.1 Frames

A *frame* is a tuple $\langle W, R_\Box \rangle_{\Box \in \alpha}$ such that $W$ is a non-empty set and every $R_\Box$ is a binary relation on $W$. The elements of $W$ are called the *points* of the frame. Note that the points of a frame need not represent possible worlds. In fact, in the semantics to be presented in the next chapter, they will represent tuples of possible worlds. A *model* based on such a frame is a tuple $\langle W, R_\Box, V \rangle_{\Box \in \alpha}$ such that $V$ is a function from proposition letters to subsets of $W$. $V$ is called the *valuation* of the model.

*Truth* of a formula $\varphi$ in a model $\mathfrak{M} = \langle W, R_\Box, V \rangle_{\Box \in \alpha}$ at a point $w$ is written as $\mathfrak{M}, w \Vdash \varphi$, and inductively defined as follows:

$\mathfrak{M}, w \Vdash p$ iff $w \in V(p)$

$\mathfrak{M}, w \Vdash \neg\varphi$ iff not $\mathfrak{M}, w \Vdash \varphi$

$\mathfrak{M}, w \Vdash \varphi \wedge \psi$ iff $\mathfrak{M}, w \Vdash \varphi$ and $\mathfrak{M}, w \Vdash \psi$

$\mathfrak{M}, w \Vdash \Box\varphi$ iff $\mathfrak{M}, v \Vdash \varphi$ for all $v$ such that $wR_\Box v$

A set of formulas is *true* in a model at a point if all of its members are. From the definition of truth, a number of further notions are derived:

A formula $\varphi$ is *valid* in a model $\mathfrak{M}$, written $\mathfrak{M} \Vdash \varphi$, if it is true at all of the points in $\mathfrak{M}$. $\varphi$ is *valid* at a point $w$ in a frame $\mathfrak{F}$, written $\mathfrak{F}, w \Vdash \varphi$, if it is true at that point in all models based on $\mathfrak{F}$. $\varphi$ is *valid* in a frame $\mathfrak{F}$, written $\mathfrak{F} \Vdash \varphi$, if it is valid in all models based on $\mathfrak{F}$. $\varphi$ is *valid* on a class of frames or models $\mathsf{C}$, written $\mathsf{C} \Vdash \varphi$, if it is valid in all elements of $\mathsf{C}$. $\Gamma$ is *valid* in any of these senses if all of its members are.

$\varphi$ is a *consequence* of $\Gamma$ on $\mathsf{C}$, written $\Gamma \vDash_{\mathsf{C}} \varphi$, if $\mathfrak{M}, w \Vdash \varphi$ for every model $\mathfrak{M}$ based on a frame in $\mathsf{C}$ and point $w$ in $\mathfrak{M}$ such that $\mathfrak{M}, w \Vdash \psi$ for all $\psi \in \Gamma$. Let $X$ be a formula or set of formulas. $X$ is *satisfiable* at a point $w$ in a frame $\mathfrak{F}$ if $X$ is true at $w$ in some model based on $\mathfrak{F}$. $X$ is *satisfiable* in $\mathfrak{F}$ if $X$ is satisfiable at some point in $\mathfrak{F}$. $X$ is *satisfiable* on $\mathsf{C}$ if it is satisfiable in some frame in $\mathsf{C}$.

### 2.2.2 Frames With Distinguished Elements

To give a semantics for QNMLs, frames have to be enriched by distinguishing some points. A *frame with distinguished elements* (FWDE) is a tuple

$\langle W, R_\Box, D\rangle_{\Box\in\alpha}$ such that $\langle W, R_\Box\rangle_{\Box\in\alpha}$ is a frame and $D \subseteq W$. $D$ are the *distinguished points* of the frame. As with frames, a *model* is obtained by adding a valuation function.

Truth in a model based on an FWDE at a point is defined as for models based on frames. The only difference is in the definition of validity, consequence, and satisfiability. Here, the idea is that only the distinguished points matter. So a formula is *valid* in a model based on an FWDE if it is true at all distinguished points in this model. The other notions of validity are derived from this as in the case of frames.

Similarly, the consequence relation on a class of FWDEs $\mathsf{C}$ is defined as follows: $\varphi$ is a *consequence* of $\Gamma$ on $\mathsf{C}$, written $\Gamma \vDash_{\mathsf{C}} \varphi$, if $\mathcal{M}, w \Vdash \varphi$ for every model $\mathcal{M}$ based on an FWDE in $\mathsf{C}$ and distinguished point $w$ in $\mathcal{M}$ such that $\mathcal{M}, w \Vdash \psi$ for all $\psi \in \Gamma$. Finally, a formula or set of formulas $X$ is *satisfiable* at a point $w$ in an FWDE $\mathcal{F}$ if $X$ is true at $w$ in some model based on $\mathcal{F}$. $X$ is *satisfiable* in $\mathcal{F}$ if $X$ is satisfiable at some distinguished element in $\mathcal{F}$. $X$ is *satisfiable* on $\mathsf{C}$ if it is satisfiable in some FWDE in $\mathsf{C}$.

### 2.2.3 Structural Transformations

Two well-known transformations of frames will be needed, mainly for the completeness proofs: generated subframes and bounded morphisms. Each of these relates two frames, and comes with a preservation result. Their definitions can be found in Blackburn et al. (2001, section 3.3). It is straightforward to extend the definitions of generated subframes to FWDEs. For any structure $F$ and set of points $X$, I will write $F_X$ for the substructure of $F$ generated by $X$. For a point $w$, I will write $F_w$ for $F_{\{w\}}$; such a substructure is called *point-generated.*

It is also useful to define a kind of transformation that takes a structure for some set of operators, and returns a structure for a subset of these by leaving out irrelevant relations. For $\beta \subseteq \alpha$, I will write this restriction as $|\beta$. E.g., for a frame $\mathfrak{F} = \langle W, R_\Box\rangle_{\Box\in\alpha}$, the frame $\mathfrak{F}|\beta$ is defined as $\langle W, R_\Box\rangle_{\Box\in\beta}$. This is defined analogously for FWDEs, models, and classes of such structures.

## 2.3 Logics and Classes of Structures

As often in logic, much of the work in the following chapters will be concerned with establishing connections between syntactic and semantic notions such as being a theorem of a certain logic and being valid on a certain class of structures. This section provides the relevant definitions. Here, when I speak of a class of structures, it can be a class of frames, a class of FWDEs, or a class of models.

### 2.3.1 Characterization and Definition

I start with the standard notions of characterization and definition, which capture the idea that every class of structures determines a logic, and every logic determines a class of structures. The *logic characterized* by a class of structures $\mathsf{C}$, written $L(\mathsf{C})$, is the set of formulas valid on it. It can be proven that any class of frames characterizes an NML, and any class of FWDEs characterizes a QNML.

The class of frames *defined* by a set of formulas $\Gamma$, written $\mathsf{Fr}(\Gamma)$, is the class of frames in which $\Gamma$ is valid. Similarly, the class of FWDEs *defined* by $\Gamma$, written $\mathsf{FrD}(\Gamma)$, is the class of FWDEs in which $\Gamma$ is valid. It is easy to see that for any set of formulas $\Gamma$, $\mathsf{Fr}(\oplus\Gamma) = \mathsf{Fr}(\Gamma)$ and $\mathsf{FrD}(+\Gamma) = \mathsf{FrD}(\Gamma)$. I will call frames/FWDEs in which $\Gamma$ is valid $\Gamma$-*frames/*FWDE*s*.

In the course of the thesis, I will define some logics as the quasi-normal join of some other logics, some of them being normal modal logics axiomatized by a set of formulas. It is useful to have a result that describes the class of FWDEs defined by such a logic. To prove such a result, I first prove the following lemma:

**Lemma 2.3.** *Let $\beta \subseteq \alpha$, $\mathcal{F} = \langle \mathfrak{F}, D\rangle$ an $\alpha$-*FWDE*, and $\Lambda$ a $\beta$-*NML*. Then $\mathcal{F} \Vdash \Lambda$ iff $(\mathfrak{F}|\beta)_D \Vdash \Lambda$.*

*Proof.* First, assume that $\mathcal{F} \Vdash \Lambda$. Since $\Lambda \subseteq \mathcal{L}_\beta$, also $\mathcal{F}|\beta \Vdash \Lambda$. Consider any point $w$ in $(\mathfrak{F}|\beta)_D$ and $\varphi \in \Lambda$. Then there are $v \in D$ and points $v_1, \ldots, v_n$ in $\mathcal{F}$ such that $vR_{\Box_1}v_1, \ldots, v_nR_{\Box_{n+1}}w$ for some $\Box_1, \ldots, \Box_{n+1} \in \beta$. Since $\Lambda$ is normal, $\Box_1 \ldots \Box_{n+1}\varphi \in \Lambda$. So $\mathcal{F}|\beta \Vdash \Box_1 \ldots \Box_{n+1}\varphi$, and therefore, since $v \in D$, $\mathcal{F}|\beta, v \Vdash \Box_1 \ldots \Box_{n+1}\varphi$. So $\mathfrak{F}|\beta, w \Vdash \varphi$. By a preservation result for generated subframes, $(\mathfrak{F}|\beta)_D, w \Vdash \varphi$. Therefore $(\mathfrak{F}|\beta)_D \Vdash \Lambda$.

Now assume that $(\mathfrak{F}|\beta)_D \Vdash \Lambda$. Consider any $w \in D$. Then $(\mathfrak{F}|\beta)_D, w \Vdash \Lambda$, and therefore by a preservation result for generated subframes $\mathfrak{F}|\beta, w \Vdash \Lambda$. So also $\mathfrak{F}, w \Vdash \Lambda$, and therefore $\mathcal{F}, w \Vdash \Lambda$. Since $w$ was chosen arbitrarily among $D$, it follows that $\mathcal{F} \Vdash \Lambda$. □

**Proposition 2.4.** *Let $\Gamma \subseteq \mathcal{L}_\alpha$, $n \in \mathbb{N}$, and for every $i \leq n$, $\beta_i \subseteq \alpha$ and $\Delta_i \subseteq \mathcal{L}_{\beta_i}$. Then $\mathsf{FrD}_\alpha(+_\alpha(\Gamma, \oplus_{\beta_1}\Delta_1, \ldots, \oplus_{\beta_n}\Delta_n))$ is the class of $\alpha$-*FWDE*s $\mathcal{F} = \langle \mathfrak{F}, D\rangle$ such that*

- *$\mathcal{F} \Vdash \Gamma$, and*
- *$(\mathfrak{F}|\beta_i)_D \Vdash \Delta_i$ for all $i \leq n$.*

*Proof.* Consider any $\alpha$-FWDE $\mathcal{F}$. Then $\mathcal{F} \Vdash +_\alpha(\Gamma, \oplus_{\beta_1}\Delta_1, \ldots, \oplus_{\beta_n}\Delta_n)$ iff $\mathcal{F} \Vdash \Gamma$ and $\mathcal{F} \Vdash \oplus_{\beta_i}\Delta_i$ for all $i \leq n$. By Lemma 2.3, for any $i \leq n$, $\mathcal{F} \Vdash \oplus_{\beta_i}\Delta_i$ iff $(\mathfrak{F}|\beta_i)_D \Vdash \oplus_{\beta_i}\Delta_i$, which is the case iff $(\mathfrak{F}|\beta_i)_D \Vdash \Delta_i$, from which the claim follows. □

### 2.3.2 Soundness and Completeness

We also need the familiar notions of soundness and completeness of a logic with respect to a class of structures. Let $\Lambda$ be a logic and $\mathsf{C}$ a class of structures. $\Lambda$ is *sound* with respect to $\mathsf{C}$ if every formula in $\Lambda$ is valid on $\mathsf{C}$. $\Lambda$ is *weakly complete* with respect to $\mathsf{C}$ if every formula valid on $\mathsf{C}$ is in $\Lambda$. $\Lambda$ is *strongly complete* with respect to $\mathsf{C}$ if $\Gamma \vdash_\Lambda \varphi$ for all $\Gamma$ and $\varphi$ such that $\Gamma \vDash_{\mathsf{C}} \varphi$.

Unsurprisingly, strong completeness implies weak completeness. As usual, we can show that $\Lambda$ is *strongly complete* with respect to $\mathsf{C}$ if and only if every $\Lambda$-consistent set is satisfiable on $\mathsf{C}$. From this, it follows that if the same sets of formulas are satisfiable on two classes of structures and a logic is sound and strongly complete with respect to one of them, it is also sound and strongly complete with respect to the other.

A logic is *strongly frame-complete* if it is sound and strongly complete with respect to some class of frames, and *strongly* FWDE*-complete* if it is sound and strongly complete with respect to some class of FWDEs. It is not difficult to see that a logic is strongly frame/FWDE-complete if and only if it is strongly complete with respect to the class of frames/FWDEs it defines.

### 2.3.3 Conservativity

When comparing a logic in a language with another logic in a sublanguage, a common question is whether the two logics agree on the theorems in the smaller language. In this case, the larger logic is said to be a conservative extension of the smaller one. This issue will come up at a number of places, so I will give a formal definition for this notion: let $\beta \subseteq \alpha$. An $\alpha$-logic $\Lambda$ is a *conservative extension* of a $\beta$-logic $\Lambda'$ if $\Lambda \cap \mathcal{L}_\beta = \Lambda'$. For logics that are weakly complete with respect to some class of structures, the following proposition provides a useful semantic criterion for being a conservative extension:

**Proposition 2.5.** *Let $\Lambda$ be an $\alpha$-logic that is sound and weakly complete with respect to some class of $\alpha$-structures $\mathsf{C}$ and $\Lambda'$ a $\beta$-logic that is sound and weakly complete with respect to some class of $\beta$-structures $\mathsf{D}$. If $L(\mathsf{C}|\beta) = L(\mathsf{D})$, then $\Lambda$ is a conservative extension of $\Lambda'$.*

*Proof.* We have to show that $\Lambda \cap \mathcal{L}_\beta = \Lambda'$. By the weak completeness of the logics, $L(\mathsf{C}) = \Lambda$ and $L(\mathsf{D}) = \Lambda'$. Since by assumption $L(\mathsf{C}|\beta) = L(\mathsf{D})$, it suffices to show that $L(\mathsf{C}) \cap \mathcal{L}_\beta = L(\mathsf{C}|\beta)$. Consider any $\varphi \in L(\mathsf{C}) \cap \mathcal{L}_\beta$ and $F \in \mathsf{C}|\beta$. Then there is an $F' \in \mathsf{C}$ such that $F'|\beta = F$. Since $\varphi \in L(\mathsf{C})$, $F' \Vdash \varphi$, and as $\varphi \in \mathcal{L}_\beta$, also $F \Vdash \varphi$. Hence, $\varphi \in L(\mathsf{C}|\beta)$. Now consider any $\varphi \in L(\mathsf{C}|\beta)$. Then $\varphi \in \mathcal{L}_\beta$. Consider any $F \in \mathsf{C}$. Then $F|\beta \in \mathsf{C}|\beta$, so $F|\beta \Vdash \varphi$, and therefore $F \Vdash \varphi$. $\square$

## 2.4 Sahlqvist Theory

To be able to prove completeness results easily, I will use a well-known general result, called the Sahlqvist completeness theorem. To state this, a set of formulas of a certain syntactic structure is described, which are called *Sahlqvist formulas* (see, e.g., Blackburn et al. (2001, section 3.6) for a definition). One can prove the following:

**Theorem 2.6.** *For any set $\Gamma$ of Sahlqvist formulas, $\oplus\Gamma$ is strongly frame-complete and $+\Gamma$ is strongly FWDE-complete.*

For proofs, see Chagrov and Zakharyaschev (1997, section 10.3) or Kracht (1999, section 5.5). Besides this theorem, Sahlqvist formulas have another nice property: there is an algorithm, called the *Sahlqvist-van Benthem algorithm*, which can be used to calculate a condition for points of frames (which is expressible in first-order logic) for a given Sahlqvist formula, such that a frame validates the Sahlqvist formula at that point if and only if that point satisfies the condition.

However, not every logic we will be concerned with is the NML or QNML axiomatized by a set of Sahlqvist formulas. One logic that will be important is a quasi-normal join of such logics. So it will be useful to extend Sahlqvist's theorem to these joins. For joins of the form used in Proposition 2.4, we get the following completeness result:

**Theorem 2.7.** *If $\Gamma, \Delta_1, \ldots, \Delta_n$ (as in Proposition 2.4) are sets of Sahlqvist formulas, then $+_\alpha(\Gamma, \oplus_{\beta_1}\Delta_1, \ldots, \oplus_{\beta_n}\Delta_n)$ is strongly* FWDE*-complete.*

As far as I know, this result is new with this thesis, and I will prove it in the rest of this chapter. In fact, this result can be used to give a quite general completeness result to the effect that any logic that can be constructed from finitely many sets of Sahlqvist formulas and NMLs and QNMLs axiomatized by Sahlqvist formulas using finitely many applications of normal and quasi-normal joins is strongly FWDE-complete. This is the case since it can be proven that any such construction can be written in the form $+_\alpha(\Gamma, \oplus_{\beta_1}\Delta_1, \ldots, \oplus_{\beta_n}\Delta_n)$. However, as there will be no need for this general result in the following, I will not prove it, but only Theorem 2.7.

For this proof, a number of definitions are needed that I have not introduced, such as the notion of a descriptive general frame. They can be found in Blackburn et al. (2001, chapter 5) and other textbooks. Also, three definitions are needed that are not standard: a *descriptive general* FWDE is a descriptive general frame to which a subset of the points is added as the distinguished elements; note that this doesn't have to be an admissible proposition/element of the algebra. A logic $\Lambda$ is *quasi-d-persistent* if for every descriptive general FWDE in which $\Lambda$ is valid, $\Lambda$ is also valid in the underlying FWDE. Further, for a set of formulas $\Gamma$, $Gen\Gamma$ is the closure of $\Gamma$ under *Gen*.

**Lemma 2.8.** *If $\Gamma$ is a set of Sahlqvist formulas, then $\Gamma$ and $+\Gamma$ are quasi-d-persistent.*

*Proof.* If $\Gamma$ is a set of Sahlqvist formulas, then the elements of $\Gamma$ are locally d-persistent (see, e.g., Kracht (1999, Theorem 5.5.5)). Let $f$ be a descriptive general FWDE. If $f \Vdash \Gamma$, then $f, w \Vdash \Gamma$ for all distinguished elements of $f$. So by the local d-persistence of the elements of $\Gamma$, $f_\sharp, w \Vdash \Gamma$ (where $f_\sharp$ is the FWDE underlying $f$). Therefore $f_\sharp \Vdash \Gamma$.

Further, if $f \Vdash +\Gamma$, then $f \Vdash \Gamma$, and so as just seen, $f_\sharp \Vdash \Gamma$. Since $f_\sharp$ is an FWDE, $L(\{f_\sharp\})$ is a QNML, and therefore contains $\mathbf{K}$ and is closed under *MP* and *US*. So by a straightforward induction, $f_\sharp \Vdash +\Gamma$. □

**Lemma 2.9.** *For any set of formulas $\Gamma$, $\oplus\Gamma =+ Gen\Gamma$.*

*Proof.* We first prove $\oplus\Gamma \subseteq +Gen\Gamma$ by induction on the proof system $N\Gamma$ for $\oplus\Gamma$. It is immediate that $\Gamma \subseteq +Gen\Gamma$. Also, since $+Gen\Gamma$ is a QNML, $\mathbf{K} \subseteq +Gen\Gamma$, and therefore $+Gen\Gamma$ contains all $K$ axioms and propositional tautologies. As a QNML, $+Gen\Gamma$ is also closed under *MP* and *US*. That $+Gen\Gamma$ is closed under *Gen* can by shown by a straightforward induction on its proof system $QN\,Gen\Gamma$.

Now we prove $+Gen\Gamma \subseteq \oplus\Gamma$ by induction on the proof system $QN\,Gen\Gamma$ for $+Gen\Gamma$. Since $\oplus\Gamma$ is an NML, $\mathbf{K} \subseteq \oplus\Gamma$. Also, $\oplus\Gamma$ is closed under *Gen*, so $Gen\Gamma \subseteq \oplus\Gamma$. As $\oplus\Gamma$ is closed under *MP* and *US*, the claim follows. □

**Lemma 2.10.** *If $\Lambda_1, \ldots, \Lambda_n$ are quasi-d-persistent logics, then $+(\Lambda_1, \ldots, \Lambda_n)$ is quasi-d-persistent.*

*Proof.* Consider any descriptive general FWDE $f$ such that $f \Vdash +(\Lambda_1, \ldots, \Lambda_n)$. Let $i \leq n$. Then $f \Vdash \Lambda_i$, and since $\Lambda_i$ is quasi-d-persistent, it follows that $f_\sharp \Vdash \Lambda_i$. Therefore $f_\sharp \Vdash \bigcup_{i \leq n} \Lambda_i$, and since $f_\sharp$ is an FWDE, it follows as in Lemma 2.8 that $f_\sharp \Vdash +(\Lambda_1, \ldots, \Lambda_n)$. □

*Proof of Theorem 2.7.* Assume that $\Gamma$, $\Delta_1$, ..., and $\Delta_n$ are sets of Sahlqvist formulas. Since the set of Sahlqvist formulas is closed under *Gen*, for any $i \leq n$, $Gen_{\beta_i}\Delta_i$ is a set of Sahlqvist formulas. By Lemma 2.8, it follows that $\Gamma$ and $+_{\beta_i} Gen_{\beta_i}\Delta_i$ for all $i \leq n$ are quasi-d-persistent. For any $i \leq n$, by Lemma 2.9, $+_{\beta_i} Gen_{\beta_i}\Delta_i = \oplus_{\beta_i}\Delta_i$, and so $\oplus_{\beta_i}\Delta_i$ is quasi-d-persistent. Finally, it follows from Lemma 2.10 that $+_\alpha(\Gamma, \oplus_{\beta_1}\Delta_1, \ldots, \oplus_{\beta_n}\Delta_n)$ is quasi-d-persistent, and so along familiar lines that it is valid in its canonical FWDE, and therefore strongly FWDE-complete. (The canonical FWDE of a QNML is the canonical frame of its kernel with its MCSs as the distinguished elements. This construction is standard in the literature on QNMLs, and can already be found in Segerberg (1971, section 3.2).) □

# Chapter 3

# A Logic for Two-Dimensional Semantics

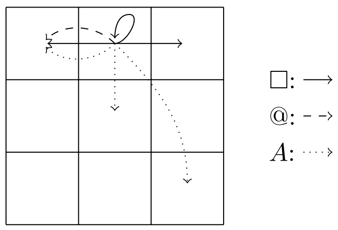
In this section, I will present a logic for two-dimensional semantics. As explained in chapter 1, this will be a propositional modal logic using operators for necessity, actuality and apriority. I will use the symbols $\Box$, @ and $A$ for them. As usual, I will use $\Diamond$ as the dual of $\Box$ (that is, $\neg\Box\neg$). As something is *conceivable* if it is not a priori that it is not the case, I will use $C$ for the dual of $A$ ($\neg A \neg$).

First, I am going to derive a formal semantics for this language from two-dimensional semantics. This will show that we can systematize the relevant aspects without running into inconsistencies, and thereby provide a general answer to the charges of incoherence. I will then give a first application of the formal semantics, by applying it to nesting problem. This will tell us how two-dimensionalists should answer the nesting problem according to the formal system. The problem will be discussed in much greater detail in chapter 6. After this, I will present a complete axiomatization of the logic. For technical reasons, I will first axiomatize the logic of general validity and then derive an axiomatization of real-world validity. The axiomatization can be seen as a way of showing that the logic is well-behaved, as it turns out to be axiomatizable in a standard fashion. Furthermore, it will be applied in chapter 5 to determine the logical commitments of two-dimensional semantics. In the final section of this chapter, I will highlight some interesting properties of the logic.

## 3.1 Formal Semantics

As described in chapter 1, the fundamental semantic value of an expression in two-dimensional semantics is a matrix that assigns extensions to scenario/world pairs. In a propositional logic, only sentential expressions are represented, so it is natural to build a formal model in which the scenarios and the worlds are represented by arbitrary sets of elements (as usual in modal logic), and proposition letters are assigned truth-values at tuples of such elements. Here, I make the simplification that in formal models, the same set represents both worlds and scenarios. This is a simplification, since it is plausible that for many worlds, there is more than one scenario based on it. But I will show in section 3.5.1 that this simplification makes no differences for the logic considered.

So we can base a model on a set (representing worlds/scenarios), and assign truth values to proposition letters relative to tuples of these elements. Further, if the Cartesian square of this set is written as a table analogously to how matrices were displayed in chapter 1, necessity is interpreted as truth on the horizontal, apriority as truth on the diagonal, and actuality as truth at the intersection of the horizontal and the diagonal. Using the interpretation of modalities by accessibility relations in frames presented in section 2.2, this suggests that we can use models based on frames in which the set of points is the Cartesian square of some set, and the accessibility relations are given as illustrated in the following frame for the upper middle element:




More formally, we can define the following class of frames to capture the logic of two-dimensional semantics:

**Definition 3.1.** *A* matrix frame *is a frame* $\mathfrak{F} = \langle W, R_\Box, R_@, R_A \rangle$*, where* $W = S \times S$ *for some set* $S$*, and the relations are given by the following conditions:*

- $\langle x, y\rangle R_\Box \langle x', y'\rangle$ *iff* $y = y'$
- $\langle x, y\rangle R_@ \langle x', y'\rangle$ *iff* $y = y'$ *and* $x' = y'$
- $\langle x, y\rangle R_A \langle x', y'\rangle$ *iff* $x' = y'$

*We say that* $\mathfrak{F}$ *is* based on $S$*. Let* $\mathsf{M}$ *be the class of matrix frames.*

Note that in such a frame, the elements of $S$ represent possible worlds or scenarios, and so the points $W$ of the frame represent tuples of such elements, and not possible worlds or scenarios themselves. Further, it should be noted that this logical interpretation of two-dimensional semantics is not strictly implied by the writings of its proponents, although it is arguably the most natural one.

To evaluate whether a formula is a consequence of some formulas on $\mathsf{M}$, we have to evaluate these formulas in all points in the models. Some of these points are tuples containing two different worlds. This class of frames therefore captures the general consequence relation, which was described in chapter 1. To capture the philosophically more plausible real-world consequence relation, the points on which consequence operates must be restricted to those on the diagonal. (See the appendix for arguments that this is the philosophically more plausible definition.) For this, we can use FWDEs, since we can choose the points on the diagonal as the distinguished elements in each frame.

**Definition 3.2.** *A* matrix FWDE *is an* FWDE $\mathcal{F} = \langle W, R_\Box, R_@, R_A, D\rangle$*, such that* $\langle W, R_\Box, R_@, R_A\rangle$ *is a matrix frame based on a set* $S$ *and* $D = \{\langle x, x\rangle : x \in S\}$*. Let* $\mathsf{MD}$ *be the class of matrix* FWDE*s.*

MD is a straightforward implementation of the ideas of two-dimensional semantics about necessity and apriority in a formal language. It thereby gives us a way of systematizing these ideas in a coherent way, which to some degree answers the challenge of those who worry that the two semantic values of primary and secondary intension provide an incoherent story.

At various points in this thesis, we will need an example of a matrix FWDE, e.g. to prove that some formula is not valid on MD. Mostly, it will suffice to take such an FWDE that is based on a two-element set. Therefore, I make the following definition:

**Definition 3.3.** *Let* $\mathcal{F}^2 = \langle W^2, R^2_{\Box}, R^2_{@}, R^2_A, D^2 \rangle$ *be the matrix* FWDE *based on the set* $\{0, 1\}$.

Note that $W^2 = \{\langle 0,0\rangle, \langle 0,1\rangle, \langle 1,0\rangle, \langle 1,1\rangle\}$ and $D^2 = \{\langle 0,0\rangle, \langle 1,1\rangle\}$. $\mathcal{F}^2$ can be illustrated using a table similar to the way two-dimensional matrices were presented in chapter 1:

| | 0 | 1 |
|---|---|---|
| 0 | $\langle 0,0\rangle$ | $\langle 1,0\rangle$ |
| 1 | $\langle 0,1\rangle$ | $\langle 1,1\rangle$ |

I will use such tables, in which the cells are filled with T for true and F for false, to illustrate models based on $\mathcal{F}^2$. E.g., a model with a valuation $V$ that maps a proposition letter $p$ to $V(p) = \{\langle 0,0\rangle, \langle 1,0\rangle\}$ is drawn as follows:

| $p$ | 0 | 1 |
|---|---|---|
| 0 | T | T |
| 1 | F | F |

## 3.2 The Nesting Problem in the Formal Semantics

To substantiate the claim that the formalization helps to deal with charges of incoherence, I apply the formal semantics to the nesting problem described in section 1.4.2. The crucial argument of the nesting problem can be written as follows in the formal language, where $N1$ and $N2$ represent the premises, and $N3$ the conclusion:

$$\begin{array}{ll} N1 & Ap \to \Box Ap \\ N2 & \Box(Ap \to p) \\ N3 & Ap \to \Box p \end{array}$$

We can prove that according to the formal semantics, this argument is valid, and the first premise is a logical truth. But we can also prove that the second premise and the conclusion are not logically true.

**Proposition 3.4.** *The following hold:*

(a) $\{N1, N2\} \vDash_{\mathsf{MD}} N3$

(b) $\mathsf{MD} \Vdash N1$

(c) $\mathsf{MD} \nVdash N2$ *and* $\mathsf{MD} \nVdash N3$

*Proof.* (a): Since $\mathsf{MD}$ is a class of FWDEs, its logic is a QNML. Therefore $\mathbf{K} \subseteq L(\mathsf{MD})$. $\vdash_{\mathbf{K}} (N1 \wedge N2) \to N3$, and so $\mathsf{MD} \Vdash (N1 \wedge N2) \to N3$. By the deduction theorem, $\{N1, N2\} \vDash_{\mathsf{MD}} N3$.

(b): Consider any model $\mathcal{M}$ based on an FWDE $\mathcal{F}$ in $\mathsf{MD}$, and distinguished point $\langle x, x \rangle$ in $\mathcal{F}$ such that $\mathcal{M}, \langle x, x \rangle \Vdash Ap$. Then $\mathcal{M}, \langle y, y \rangle \Vdash p$ for all distinguished points $\langle y, y \rangle$ of $\mathcal{F}$. So for any point $\langle z, x \rangle$ in $\mathcal{F}$, $\mathcal{M}, \langle z, x \rangle \Vdash Ap$, and therefore $\mathcal{M}, \langle x, x \rangle \Vdash \Box Ap$.

(c): Consider a model $\mathcal{M}$ based on $\mathcal{F}^2$ such that its valuation $V$ maps $p$ to $V(p) = \{\langle 0, 0 \rangle, \langle 1, 1 \rangle\}$:

| $p$ | 0 | 1 |
|---|---|---|
| 0 | T | F |
| 1 | F | T |

Then $\mathcal{M}, \langle 1, 0 \rangle \Vdash Ap \wedge \neg p$, so $\mathcal{M}, \langle 0, 0 \rangle \nVdash \Box(Ap \to p)$. Furthermore, $\mathcal{M}, \langle 0, 0 \rangle \nVdash Ap \to \Box p$. So $\mathcal{M}, \langle 0, 0 \rangle$ falsifies both $N2$ and $N3$. $\Box$

So according to the formalization, two-dimensionalists should answer the nesting problem by denying that all instances of the second premise are true. In particular, they should claim that if $p$ is a priori and contingent, then it is not necessarily the case that if $p$ is a priori then $p$ is the case. This may be surprising, and it may even seem that it runs counter to our intuitions about apriority. So the present answer is best seen as an argument that shows that the nesting problem doesn't show that two-dimensional semantics is incoherent. It may still show that two-dimensional semantics is implausible. In chapter 6, I will argue independently of two-dimensional semantics that the answer to the nesting problem just given is in fact the philosophically correct one.

## 3.3 The Logic of General Consequence

I now turn to axiomatizing the logic of two-dimensional semantics. Although $\mathsf{MD}$ is the class of structures that characterizes the philosophically interesting logic, I will first give an axiomatization for $\mathsf{M}$, and derive an proof system for $\mathsf{MD}$ from this. This is a common strategy in the logic of indexicals, see, e.g., Vlach (1973) or Crossley and Humberstone (1977). This section will only be concerned with defining the logic **2Dg** and proving that it is the logic of $\mathsf{M}$. The succeeding section will derive a logic **2D** from this, and prove that it is the logic of $\mathsf{MD}$. ("**2D**" stands for *two-dimensional*. So whereas in $\mathsf{MD}$, "D" stands for *distinguished (elements)*, in **2D**, "**D**" stands for *dimension*.) **2Dg**, the axiomatization of the logic of $\mathsf{M}$, is given by the following definition as the NML axiomatized by a number of formulas:

**Definition 3.5.** **2Dg** $= \oplus\{T_\Box, 5_\Box, D_@, D_{c@}, I1, I2, 4_A, 5_A, I3, I4\}$, *where these are the following axioms:*

| | | | |
|---|---|---|---|
| $T_\Box$ | $\Box p \to p$ | $I2$ | $@p \to \Box @p$ |
| $5_\Box$ | $\Diamond p \to \Box\Diamond p$ | $4_A$ | $Ap \to AAp$ |
| $D_@$ | $@p \to \neg @\neg p$ | $5_A$ | $Cp \to ACp$ |
| $D_{c@}$ | $\neg @\neg p \to @p$ | $I3$ | $Ap \to @p$ |
| $I1$ | $\Box p \to @p$ | $I4$ | $A(@p \to p)$ |

At first glance, this axiomatization consisting of ten formulas might look unwieldy. But the first six axioms are just the axioms of necessity and actuality as found in Davies and Humberstone (1980). (Crossley and Humberstone (1977) contains the same, plus an additional axiom, which can be shown to be redundant). So we only have to add the last four to express the logic of apriority and its interaction with necessity and actuality.

The philosophical plausibility of the first six axioms is motivated in the texts just mentioned, so I will only say something about the others. $4_A$ and $5_A$ can be understood as saying that apriority has the positive and negative introspection property. That is, whether something is a priori or not, it is a priori whether it is. $I3$ expresses that what is a priori is actually the case, and $I4$ says that it is a priori that if something is actually the case, then it is the case. The philosophical plausibility of some of these principles is discussed in more detail in chapter 5.

All these axioms are Sahlqvist formulas, therefore the first-order conditions they express can be calculated by the Sahlqvist-van Benthem algorithm. The following table lists the axioms of **2D**, local frame correspondents for them (a first-order formula expressing a condition that is satisfied by a frame at a point if and only if the axiom is valid in that frame at that point), and properties they define (a property which a frame has if and only if the axiom is valid in it). Here, $\mathrm{im}(R)$ is the image of a relation $R$, and $\circ$ composes relations.

| | | | |
|---|---|---|---|
| $T_\Box$ | $\Box p \to p$ | $wR_\Box w$ | $R_\Box$ is reflexive |
| $5_\Box$ | $\Diamond p \to \Box\Diamond p$ | $\forall vu((wR_\Box v \wedge wR_\Box u) \to vR_\Box u)$ | $R_\Box$ is euclidean |
| $D_@$ | $@p \to \neg @\neg p$ | $\exists v(wR_@ v)$ | $R_@$ is serial |
| $D_{c@}$ | $\neg @\neg p \to @p$ | $\forall vu((wR_@ v \wedge wR_@ u) \to v = u)$ | $R_@$ is functional |
| $I1$ | $\Box p \to @p$ | $\forall v(wR_@ v \to wR_\Box v)$ | $R_@ \subseteq R_\Box$ |
| $I2$ | $@p \to \Box @p$ | $\forall vu((wR_\Box v \wedge vR_@ u) \to wR_@ u)$ | $R_@ \circ R_\Box \subseteq R_@$ |
| $4_A$ | $Ap \to AAp$ | $\forall vu((wR_A v \wedge vR_A u) \to wR_A u)$ | $R_A$ is transitive |
| $5_A$ | $Cp \to ACp$ | $\forall vu((wR_A v \wedge wR_A u) \to vR_A u)$ | $R_A$ is euclidean |
| $I3$ | $Ap \to @p$ | $\forall v(wR_@ v \to wR_A v)$ | $R_@ \subseteq R_A$ |
| $I4$ | $A(@p \to p)$ | $\forall v(wR_A v \to vR_@ v)$ | $R_@$ is reflexive on $\mathrm{im}(R_A)$ |

Furthermore, the fact that the axioms are Sahlqvist formulas immediately gives us a completeness result:

**Theorem 3.6.** **2Dg** *is strongly frame-complete.*

*Proof.* By Theorem 2.6 (the Sahlqvist completeness theorem). □

In particular, **2Dg** is sound and strongly complete with respect to the class of frames it defines:

**Corollary 3.7.** **2Dg** *is sound and strongly complete with respect to* $\mathsf{Fr}(\mathbf{2Dg})$.

To prove that **2Dg** is sound and strongly complete with respect to $\mathsf{M}$, I will use the following strategy: in a first step, I will show that **2Dg** is sound and strongly complete with respect to a class of frames $\mathsf{R}$, which is contained in $\mathsf{Fr}(\mathbf{2Dg})$ and contains $\mathsf{M}$. In a second step, I will use this to show that **2Dg** is sound and strongly complete with respect to $\mathsf{M}$. The first step is done by proving that a set of formulas is satisfiable on $\mathsf{Fr}(\mathbf{2Dg})$ if and only if it is satisfiable on $\mathsf{R}$, and the second by proving that a set of formulas is satisfiable on $\mathsf{R}$ if and only if it is satisfiable on $\mathsf{M}$. More specifically, I will establish this by proving that $\mathsf{R}$ is the class of point-generated subframes of **2Dg**-frames, as well as the class of bounded morphic images of frames in $\mathsf{M}$. The desired claims about satisfiability follow from these structural connections by well-known preservation results.

In pursuing this proof strategy, we are doing a bit more work than is needed to establish the soundness and strong completeness of **2Dg** with respect to $\mathsf{M}$. It would suffice to show for soundness that all axioms of **2Dg** are valid on $\mathsf{M}$, which would be relatively simple, and to show for strong completeness that any point-generated subframe of a **2Dg**-frame is in $\mathsf{R}$, and that every frame in $\mathsf{R}$ is a bounded morphic image of some frame in $\mathsf{M}$. However, I think it is nice to see the structural connections between the three classes of frames more fully developed, which is why I follow the proof strategy outlined earlier.

In a slightly reformulated version, the intermediate class of frames $\mathsf{R}$ is used in Restall (2010), so I will call them *Restall frames*. Calling a relation a *function* if it is serial and functional, they can be defined as follows:

**Definition 3.8.** *A* Restall frame *is a frame* $\mathfrak{F} = \langle W, R_\Box, R_@, R_A\rangle$ *such that*

- $R_\Box$ *is an equivalence relation,*
- $R_@$ *is a function that maps any two* $R_\Box$*-related worlds to the same world, which is* $R_\Box$*-related to both of them, and*
- $wR_Av$ *iff* $v \in \mathrm{im}(R_@)$ *for all* $w, v \in W$.

*Let* $\mathsf{R}$ *be the class of Restall frames.*

### 3.3.1 Completeness of 2Dg with Respect to $\mathsf{R}$

I will now show that Restall frames are exactly the point-generated subframes of **2Dg**-frames. From this, it follows that a set of formulas is satisfiable on $\mathsf{R}$ if and only if it is satisfiable on $\mathsf{Fr}(\mathbf{2Dg})$, which gives us the soundness and strong completeness of **2Dg** with respect to $\mathsf{R}$.

**Lemma 3.9.** *Every Restall frame is a point-generated subframe of a* **2Dg**-*frame.*

*Proof.* Consider any Restall frame $\mathfrak{F} = \langle W, R_\Box, R_@, R_A\rangle$. We first show that for any $w \in W$, $\mathfrak{F}_w$ (the subframe of $\mathfrak{F}$ generated by $w$) is $\mathfrak{F}$ itself. Consider any $v \in W$. Since $R_@$ is serial, there is a $u \in W$ such that $vR_@u$. So also $vR_\Box u$, and by symmetry of $R_\Box$, $uR_\Box v$. It is also the case that $u \in \mathrm{im}(R_@)$, so $wR_Au$. Therefore $w(R_\Box \circ R_A)v$, so $v$ is in $\mathfrak{F}_w$. As $v$ was chosen arbitrarily, $\mathfrak{F}_w = \mathfrak{F}$.

To show that $\mathfrak{F}$ is itself a **2Dg**-frame, it suffices to go through the axioms of **2Dg** and verify that the properties defined by them are satisfied by Restall frames. This is straightforward for all axioms except $I4$. For this, we can reason as follows: let $w \in \mathrm{im}(R_A)$. Then $w \in \mathrm{im}(R_@)$, so there is a $v$ such that $vR_@w$. It follows that $vR_\Box w$, and therefore that $R_@$ must map $v$ and $w$ to the same point. So $wR_@w$, which means that $R_@$ is reflexive on $\mathrm{im}(R_A)$. $\Box$

The next result will make use of the fact that $N1$, the formula representing the first premise in the nesting argument, is a theorem of **2Dg**. This is derived in the following lemma:

**Lemma 3.10.** $\vdash_{\mathbf{2Dg}} Ap \to \Box Ap$.

*Proof.* By the following derivation:

| | | |
|---|---|---|
| (1) | $Cp \to ACp$ | $5_A$ |
| (2) | $ACp \to @Cp$ | $I3$ |
| (3) | $Cp \to @Cp$ | (1), (2) |
| (4) | $\neg@\neg A\neg p \to A\neg p$ | (3) |
| (5) | $@Ap \to Ap$ | $D_@$, (4) |
| (6) | $\Box@Ap \to \Box Ap$ | $K_\Box$, Gen$_\Box$, (5) |
| (7) | $@Ap \to \Box@Ap$ | $I2$ |
| (8) | $@Ap \to \Box Ap$ | (6), (7) |
| (9) | $Ap \to AAp$ | $4_A$ |
| (10) | $AAp \to @Ap$ | $I3$ |
| (11) | $Ap \to @Ap$ | (9), (10) |
| (12) | $Ap \to \Box Ap$ | (8), (11) |

□

Note that $Ap \to \Box Ap$ is also a Sahlqvist formula, and that it is therefore straightforward to calculate that it has the following local frame correspondent: $\forall vu((wR_\Box v \wedge vR_A u) \to wR_A u)$.

**Lemma 3.11.** *Every point-generated subframe of a* **2Dg**-*frame is a Restall frame.*

*Proof.* Consider any **2Dg**-frame $\mathfrak{F} = \langle W, R_\Box, R_@, R_A\rangle$ and $w \in W$. Let $\mathfrak{F}_w = \langle W', R'_\Box, R'_@, R'_A\rangle$ be the subframe generated by $w$. Since validity is preserved under taking generated subframes, all of the axioms of **2Dg** are valid in $\mathfrak{F}_w$.

Using $T_\Box$ and $5_\Box$, it is routine to show that $R'_\Box$ is an equivalence relation. From $D_@$ and $D_{c@}$, it follows that $R'_@$ is a function. Likewise, it follows from $I1$ and $I2$ that $R'_@$ maps $R'_\Box$-related points to the same point, to which both are $R'_\Box$-related.

To show that $vR'_A u$ iff $u \in \mathrm{im}(R'_@)$, assume first that $vR'_A u$. Then by $I4$, $uR'_@ u$, and so $u \in \mathrm{im}(R'_@)$. It only remains to show that if $u \in \mathrm{im}(R'_@)$, then $vR'_A u$. We will do this in the rest of this proof. To do so, I adopt the notation to write $R[Y]$ for the image of the set $Y$ under the relation $R$. Let $w'$ be the element of $W'$ such that $wR'_@ w'$. The existence and uniqueness of this point are guaranteed by the fact that $R'_@$ is a function. We first prove a preliminary claim:

*Claim 1:* $W' = X$, where $X = R_\Box[R_A[\{w\}]]$. Clearly $X \subseteq W'$. We first show that $w \in X$, and then that $X$ is closed under each of the relations, that is, that $R_\nabla[X] \subseteq X$ for every modality $\nabla$.

$wR_@ w'$, so both $wR_A w'$ and $wR_\Box w'$. Since $R_\Box$ is symmetric, $w'R_\Box w$, and therefore $w \in X$. Assume that $v \in R_\Box[X]$. Then there is a $u \in X$ such that $uR_\Box v$. Since $u \in X$, there is a $u' \in W$ such that $wR_A u'$ and $u'R_\Box u$. By transitivity of $R_\Box$, $u'R_\Box v$, and so $v \in X$. Assume that $v \in R_@[X]$. Then there is a $u \in X$ such that $uR_@ v$. By I1, $R_@ \subseteq R_\Box$, so $uR_\Box v$. That $v \in X$ follows by transitivity of $R_\Box$ as before. Assume that $v \in R_A[X]$. Then there is a $u \in X$ such that $uR_A v$, and therefore a $u' \in W$ such that $wR_A u'$ and $u'R_\Box u$. By Lemma 3.10, it follows that $u'R_A v$, and so by transitivity of $R_A$ that $wR_A v$. Since $R_\Box$ is reflexive, $v \in X$. This concludes the proof of claim 1.

Now consider any $u \in \text{im}(R'_@)$ and $v \in W'$. We have to prove that $vR'_A u$. We do this by first proving that $vR'_A w'$ and then that $w'R'_A u$.

*Claim 2:* $vR'_A w'$. Since $v \in W'$, it follows from claim 1 that there is a $v' \in W'$ such that $wR'_A v'$ and $v'R'_\Box v$. By symmetry of $R'_\Box$, $vR'_\Box v'$. Since $wR'_@ w'$, by $I3$ also $wR'_A w'$. So since $R'_A$ is euclidean, $v'R'_A w'$. By Lemma 3.10, it follows that $vR'_A w'$.

*Claim 3:* $w'R'_A u$. Since $u \in W'$, there is a $u' \in W'$ such that $wR'_A u'$ and $u'R'_\Box u$. As we've seen before, $wR'_A w'$, so since $R'_A$ is euclidean, $w'R'_A u'$. Also $u \in \text{im}(R'_@)$, so there is a $u'' \in W'$ such that $u''R'_@ u$. By I1 also $u''R'_\Box u$, and with the fact that $R'_\Box$ is an equivalence relation, $u'R'_\Box u''$. So by I2, it follows that $u'R'_@ u$. So by I3, $u'R'_A u$. So by transitivity for $R'_A$, $w'R'_A u$.

By transitivity of $R'_A$, it follows from claims 2 and 3 that $vR'_A u$. □

**Theorem 3.12.** **2Dg** *is sound and strongly complete with respect to* $\mathsf{R}$.

*Proof.* By Lemmas 3.9 and 3.11, Restall frames are exactly the point-generated subframes of **2Dg**-frames. Since truth is invariant under taking generated submodels (see Blackburn et al. (2001, Proposition 2.6)), a set of formulas is satisfiable on $\mathsf{R}$ if and only if it is satisfiable on $\mathsf{Fr}(\mathbf{2Dg})$. From this, the theorem follows with Corollary 3.7. □

### 3.3.2 Completeness of 2Dg with Respect to M

To show the completeness of **2Dg** with respect to $\mathsf{M}$, it now suffices to show that a set of formulas is satisfiable on $\mathsf{R}$ if and only if it is satisfiable on $\mathsf{M}$. For this, I will show that Restall frames are the bounded morphic images of matrix frames. Essentially, the fact that $\mathsf{R}$ and $\mathsf{M}$ have the same logic is already contained in Restall (2010, Theorem 8), although without bringing out the structural connections between the classes of frames in the way I do here.

**Lemma 3.13.** *Every bounded morphic image of a matrix frame is a Restall frame.*

*Proof.* By checking the conditions on Restall frames, one can verify that matrix frames are Restall frames. With this, the claim follows from the fact that $\mathsf{R}$ is closed under taking bounded morphic images, which is routine to prove. □

**Lemma 3.14.** *Every Restall frame is a bounded morphic image of a matrix frame.*

*Proof.* Let $\mathfrak{F} = \langle W, R_\Box, R_@, R_A \rangle$ be a Restall frame. We proceed by constructing a matrix frame $\mathfrak{F}'$ and a surjective bounded morphism $f$ from $\mathfrak{F}'$ to $\mathfrak{F}$. I will use the following notation: $[x]_E$ is the equivalence class of $x$ under the equivalence relation $E$. For a relation $R$ that is a function, $R(x)$ is the unique $y$ such that $xRy$.

Let $I$ be a set of cardinality $|W|$. Let $\alpha : I \to W$ be a surjection, and for every $i \in I$, let $\beta_i : I \to [\alpha(i)]_{R_\Box}$ be a surjection such that $\beta_i(i) = R_@(\alpha(i))$. Such surjections exist for cardinality reasons, and the fact that $R_@$ is a function for which $R_@ \subseteq R_\Box$ holds. We define a function $f : I \times I \to W$ by $f(\langle i, j \rangle) = \beta_j(i)$. We prove that $f$ is a surjective bounded morphism from $\mathfrak{F}'$ to $\mathfrak{F}$, where $\mathfrak{F}'$ is the matrix frame $\langle W', R'_\Box, R'_@, R'_A \rangle$ based on $I$ (that is, $W' = I \times I$). We start by showing that $f$ is a bounded morphism, by going through the modalities, and checking the forth and back conditions for each of them.

$\Box$: Assume that $\langle i,j\rangle R'_\Box \langle i',j'\rangle$. Then $j = j'$. Since $f(\langle i,j\rangle) \in [\alpha(j)]_{R_\Box}$ and $f(\langle i',j'\rangle) \in [\alpha(j')]_{R_\Box}$, it follows that $f(\langle i,j\rangle)R_\Box f(\langle i',j'\rangle)$. Now assume that $f(\langle i,j\rangle)R_\Box w$. Then $w \in [\alpha(j)]_{R_\Box}$, so there is an $i' \in I$ such that $f(\langle i',j\rangle) = w$. Further, $\langle i,j\rangle R'_\Box \langle i',j\rangle$.

@: Assume that $\langle i,j\rangle R'_@ \langle i',j'\rangle$. Then $j = i' = j'$. So $f(\langle i',j'\rangle) = R_@(\alpha(j))$. Also $f(\langle i,j\rangle) \in [\alpha(j)]_{R_\Box}$. Therefore $f(\langle i,j\rangle)R_@ f(\langle i',j'\rangle)$. Now assume that $f(\langle i,j\rangle)R_@ w$. Note that $\langle i,j\rangle R'_@ \langle j,j\rangle$. Also $f(\langle j,j\rangle) = R_@(\alpha(j))$. Further, $f(\langle i,j\rangle)$ is in $[\alpha(j)]_{R_\Box}$, so $f(\langle i,j\rangle)R'_@ f(\langle j,j\rangle)$. So $f(\langle j,j\rangle) = w$.

$A$: Assume that $\langle i,j\rangle R'_A \langle i',j'\rangle$. Then $i' = j'$. So $f(\langle i',j'\rangle) = R_@(\alpha(i'))$. Therefore $f(\langle i',j'\rangle) \in \mathrm{im}(R_@)$, and so $f(\langle i,j\rangle)R_A f(\langle i',j'\rangle)$. Now assume that $f(\langle i,j\rangle)R_A w$. Then $w \in \mathrm{im}(R_@)$. So there is a $v$ such that $R_@(v) = w$. Let $i' \in I$ be such that $\alpha(i') = v$. Then $f(\langle i',i'\rangle) = R_@(\alpha(i')) = R_@(v) = w$. Furthermore, $\langle i,j\rangle R'_A \langle i',i'\rangle$.

So $f$ is a bounded morphism. For surjectivity, consider any $w \in W$. For some $i \in I$, $\alpha(i) = w$. So $\beta_i : I \to [w]_{R_\Box}$ is a surjective function. Therefore, there is a $j \in I$ such that $\beta_i(j) = w$. So $f(\langle j,i\rangle) = w$. $\Box$

From this, the central result of this section follows:

**Theorem 3.15.** **2Dg** *is sound and strongly complete with respect to* $\mathsf{M}$.

*Proof.* The theorem follows with Theorem 3.12 from the claim that a set of formulas $\Gamma$ is satisfiable on $\mathsf{M}$ if and only if it is satisfiable on $\mathsf{R}$, which we now prove. As noted in the proof of Lemma 3.13, matrix frames are Restall frames, therefore if $\Gamma$ is satisfiable on $\mathsf{M}$, it is also satisfiable on $\mathsf{R}$.

So assume that $\Gamma$ is satisfiable on $\mathsf{R}$. Then $\Gamma$ is true at some point $w$ in some model $\langle \mathfrak{F}, V\rangle$ based on a Restall frame $\mathfrak{F}$. By Lemma 3.14, $\mathfrak{F}$ is the bounded morphic image of some matrix frame $\mathfrak{F}'$. Let $f$ be a surjective bounded morphism from $\mathfrak{F}'$ to $\mathfrak{F}$. Define a valuation $V'$ for $\mathfrak{F}'$ by letting $V'(p) = \{w \in W' : f(w) \in V(p)\}$ for any proposition letter $p$, where $W'$ is the set of points in $\mathfrak{F}'$. Then $\langle \mathfrak{F}, V\rangle$ is a bounded morphic image of $\langle \mathfrak{F}', V'\rangle$, and as truth at a point is invariant under bounded morphisms between models (see Blackburn et al. (2001, Proposition 2.14)), it follows that $\Gamma$ is true at $w$ in $\langle \mathfrak{F}', V'\rangle$, and therefore satisfiable on $\mathsf{M}$. $\Box$

## 3.4 The Logic of Real-World Consequence

We can now use **2Dg** to define a logic **2D**, and derive from Theorem 3.15 that it is complete with respect to $\mathsf{MD}$. Since $\mathsf{MD}$ is the model theory that captures two-dimensional semantics, this completeness result means that the definition of **2D** below gives us a syntactic characterization of the logic of two-dimensional semantics. Similar to the derivation of the logic of real-world validity from the logic of general validity in Crossley and Humberstone (1977), we can define **2D** from **2Dg** as follows:

**Definition 3.16.** $\vdash_{\mathbf{2D}} \varphi$ *iff* $\vdash_{\mathbf{2Dg}} @\varphi$.

To prove completeness, I first prove a slightly more general lemma, which can also be used for a second class of structures.

**Lemma 3.17.** *Let $\Lambda$ be an* NML *that is sound and strongly complete with respect to a class of frames* $\mathsf{C}$ *and that contains both* $D_@$ *and* $D_{c@}$*. Let* $\Lambda'$ *be defined by* $\vdash_{\Lambda'} \varphi$ *iff* $\vdash_\Lambda @\varphi$*. Then* $\Lambda'$ *is sound and strongly complete with respect to the class* $\mathsf{CD}$ *of* FWDE*s based on a frame in* $\mathsf{C}$ *with* $\mathrm{im}(R_@)$ *as the set of distinguished elements.*

*Proof.* It is straightforward to check that $\Lambda'$ contains **K** and is closed under *MP* and *US*, and therefore is a QNML. We now prove the claim by showing that a set of formulas is $\Lambda'$-consistent iff it is satisfiable on $\mathsf{CD}$.

First, let $\Gamma$ be a $\Lambda'$-inconsistent set. Then there are $\varphi_1, \ldots, \varphi_n \in \Gamma$ such that $\vdash_{\Lambda'} \neg \bigwedge_{i \leq n} \varphi_i$. So by definition of $\Lambda'$, $\vdash_\Lambda @\neg \bigwedge_{i \leq n} \varphi_i$. Consider any FWDE $\mathcal{F}$ with set of points $W$, relation $R_@$, and distinguished points $D$ in $\mathsf{CD}$, and let $w \in D$. Then $w \in \mathrm{im}(R_@)$, so there is a $v \in W$ such that $vR_@w$. By soundness of $\Lambda$, $\mathcal{F}, v \Vdash @\neg \bigwedge_{i \leq n} \varphi_i$. Therefore $\mathcal{F}, w \Vdash \neg \bigwedge_{i \leq n} \varphi_i$. Hence $\Gamma$ is not satisfiable on $\mathsf{CD}$.

Now, let $\Gamma$ be a set that is not satisfiable on $\mathsf{CD}$. Assume for contradiction that $\Gamma^@ = \{@\varphi : \varphi \in \Gamma\}$ is satisfiable on $\mathsf{C}$. Then there is a frame $\mathfrak{F}$ with set of points $W$ and relation $R_@$ in $\mathsf{C}$ and $w \in W$ such that $\Gamma^@$ is satisfiable in $\mathfrak{F}$ at $w$. Since $\vdash_\Lambda D_@$, it follows from the soundness of $\Lambda$ that there is a $v \in W$ such that $wR_@v$, so $\Gamma$ is satisfiable in $\mathfrak{F}$ at $v$. But $v \in \mathrm{im}(R_@)$, so $\Gamma$ is satisfiable in $\langle \mathfrak{F}, \mathrm{im}(R_@) \rangle$ and therefore on $\mathsf{CD}$. $\lightning$, so $\Gamma^@$ is not satisfiable on $\mathsf{C}$.

By strong completeness of $\Lambda$ it follows that $\Gamma^@ \vdash_\Lambda \bot$, and so that there are $\varphi_1, \ldots, \varphi_n \in \Gamma$ such that $\vdash_\Lambda \neg \bigwedge_{i \leq n} @\varphi_i$. Since $D_@$ and $D_{c@}$ are theorems of $\Lambda$, $@$ distributes over Boolean connectives in $\Lambda$. Therefore $\vdash_\Lambda @\neg \bigwedge_{i \leq n} \varphi_i$, and so by definition of $\Lambda'$, $\vdash_{\Lambda'} \neg \bigwedge_{i \leq n} \varphi_i$. Hence $\Gamma$ is not $\Lambda'$-consistent. □

### 3.4.1 Completeness of 2D

Using the previous lemma, we can derive completeness for **2D** from the completeness result for **2Dg**. This is the central result of this chapter, which shows that **2D** is in fact the logic of two-dimensional semantics as given by the class of FWDEs $\mathsf{MD}$.

**Theorem 3.18.** **2D** *is sound and strongly complete with respect to* $\mathsf{MD}$.

*Proof.* Since $\mathsf{MD}$ is the class of FWDEs based on matrix frames with $\mathrm{im}(R_@)$ as the distinguished elements, this follows from Theorem 3.15 and Lemma 3.17. □

The advantage of having proved Lemma 3.17 first is that we can now easily define a class of FWDEs based on Restall frames with respect to which **2D** is sound and strongly complete as well.

**Definition 3.19.** *A* Restall FWDE *is an* FWDE $\mathcal{F} = \langle W, R_\Box, R_@, R_A, D \rangle$*, such that* $\langle W, R_\Box, R_@, R_A \rangle$ *is a Restall frame and* $D = \mathrm{im}(R_@)$*. Let* $\mathsf{RD}$ *be the set of Restall* FWDE*s.*

**Theorem 3.20.** **2D** *is sound and strongly complete with respect to* $\mathsf{RD}$.

*Proof.* By definition, $\mathsf{RD}$ is the class of FWDEs based on Restall frames with $\mathrm{im}(R_@)$ as the distinguished elements, so this follows immediately from Theorem 3.12 and Lemma 3.17. □

### 3.4.2 2D as a Quasi-Normal Join

There is another way of deriving **2D** from **2Dg**, besides the one used in Definition 3.16. We can also define **2D** as the quasi-normal join of **2Dg** with the formula $T_@ = @p \to p$. This provides another syntactic perspective on **2D**, and we will also use it in the proof of Theorem 4.5. Again, I first prove a slightly more general lemma, which will be useful later.

**Lemma 3.21.** *Let* $\Lambda$ *be an* NML*, from which* $\Lambda'$ *is defined by* $\vdash_{\Lambda'} \varphi$ *iff* $\vdash_\Lambda @\varphi$*. If* $\vdash_\Lambda @(@p \to p)$*, then* $\Lambda' = \Lambda + \{T_@\}$*.*

*Proof.* $\subseteq$: Consider any $\varphi \in \Lambda'$. Then $\vdash_\Lambda @\varphi$, hence $@\varphi \in \Lambda + \{T_@\}$. Since $@\varphi \to \varphi \in \Lambda + \{T_@\}$, it follows that $\varphi \in \Lambda + \{T_@\}$.

$\supseteq$: Let $\varphi \in \Lambda + \{T_@\}$. Note that $\Lambda + \{T_@\} = L(QN(\Lambda \cup \{T_@\}))$, hence we can proceed by induction on derivations in the proof system $QN(\Lambda \cup \{T_@\})$. If $\varphi \in \Lambda$, then by generalization for @, $\vdash_\Lambda @\varphi$, so $\vdash_{\Lambda'} \varphi$. Since $\mathbf{K} \subseteq \Lambda$, we do not have to consider $\varphi \in \mathbf{K}$ separately. By assumption $\vdash_\Lambda @(@p \to p)$, and so $\vdash_{\Lambda'} T_@$. As noted in the proof of Lemma 3.17, it can be shown that $\Lambda'$ is a QNML, so the cases of modus ponens and uniform substitution follow immediately by induction. $\square$

With this, it is easy to prove that **2D** can be defined as a join:

**Theorem 3.22.** $\mathbf{2D} = \mathbf{2Dg} + \{T_@\}$.

*Proof.* That $\vdash_{\mathbf{2Dg}} @(@p \to p)$ follows from axioms $I3$ and $I4$, so the claim holds by Lemma 3.21. $\square$

A historical note: since **2Dg** is the NML given by some axioms, we can think of this alternative characterization of **2D** as giving us a proof system in which proofs have two stages. In the first stage, the rule of generalization is admissible, and only the axioms of **2Dg** are available. In the second stage, we get the additional axiom of $T_@$, but lose the rule of generalization. A proof system of this kind has already been used in one of the earliest works on the logic of indexicals, namely Kamp (1971, see pp. 243–245).

## 3.5 Some Properties of 2D

In the previous sections of this chapter, I have derived the semantics MD from two-dimensional semantics, applied it to the nesting problem, and syntactically characterized its logic **2D**. In this section, I will look at some philosophically relevant properties of **2D**. In particular, it will be interesting to see what **2D**, and thereby two-dimensional semantics as represented by MD, says about the logics of necessity, actuality and apriority, and the interactions of these modalities.

I will start by showing that the simplification made in the formalization of not using different sets for worlds and scenarios does not make a difference, which is needed to justify the adequacy of the formalization. I will then show that **2D** is not normal, as it is not closed under the rule of generalization for $\Box$. Since the logic includes @, this is not surprising. In the third part, I will show that if we look at formulas containing only $\Box$ in **2D**, we get the familiar logic **S5**, and similarly for $A$. I will then consider some specific interaction principles in **2D** that are philosophically interesting. This will be quite unsystematic; the interactions will be investigated more fully in chapter 5. Finally, I will show that @ is redundant in **2D** in the sense that for every formula, there is a **2D**-equivalent formula not containing @. This last item might not be so interesting philosophically, but it will also be needed for a result in section 4.1.

### 3.5.1 Rectangular Frames

At the beginning of this chapter, I said that matrix frames make the philosophically implausible simplification of identifying scenarios and worlds. It would be much more adequate if matrix models were generalized to tables, in which there are at least as many rows as columns, and a function is given that maps every row (representing a scenario) to a column (representing its possible world). More formally, we can define:

**Definition 3.23.** *A* rectangular frame *is a frame* $\mathfrak{F} = \langle W, R_\Box, R_@, R_A \rangle$*, where* $W = P \times S$ *for some sets* $P$ *(representing possible worlds) and* $S$ *(representing scenarios), there is a surjective function* $d : S \to P$ *(representing the function that maps scenarios to their possible worlds), and the relations are given by the following conditions:*

- $\langle x, y \rangle R_\Box \langle x', y' \rangle$ *iff* $y = y'$
- $\langle x, y \rangle R_@ \langle x', y' \rangle$ *iff* $y = y'$ *and* $x' = d(y')$
- $\langle x, y \rangle R_A \langle x', y' \rangle$ *iff* $x' = d(y')$

*A* rectangular FWDE *is an* FWDE *based on a rectangular frame such that the set of distinguished elements* $D = \{\langle d(x), x \rangle : x \in S\}$*. Let* Rec *be the class of rectangular frames and* RecD *the class of rectangular* FWDE*s.*

It is not difficult to see that every matrix frame is a rectangular frame, that every rectangular frame is a Restall frame, and that the analogous observations hold for the respective classes of FWDEs. Since a set of formulas is satisfiable on M if and only if it is satisfiable on R, and similarly for MD and RD, it follows from the soundness and strong completeness of **2Dg** and **2D** with respect to these classes that **2Dg** is sound and strongly complete with respect to Rec and **2D** is sound and strongly complete with respect to RecD.

So on the level of the propositional logic of $\Box$, @ and $A$, the simplification of not distinguishing between scenarios and possible worlds made in matrix models is not problematic. It should be noted that this might be different if the logic contained additional operators, e.g., propositional quantifiers. See also Humberstone (2004, p. 20) and Restall (2010, p. 22) for similar observations.

One might also wonder whether the space of possible worlds could vary with the scenario considered as actual. For this, we could also define a class of frames and FWDEs, which would properly include Rec and RecD. But as with Rec and RecD, all of these frames and FWDEs would be contained in R and RD, and therefore by the same argument, it would follow that **2Dg** and **2D** are sound and strongly complete with respect to them.

### 3.5.2 Quasi-Normality

At the beginning of chapter 2, I noted that the logic of necessity and actuality is not closed under the generalization rule for $\Box$, which motivated the introduction of QNMLs. It therefore does not come as a surprise that this applies to **2D** as well:

**Theorem 3.24.** ***2D** is not closed under $Gen_\Box$, and so is not an NML.*

*Proof.* By Theorem 3.22, $\vdash_{\mathbf{2D}} @p \to p$. To show that $\nvdash_{\mathbf{2D}} \Box(@p \to p)$, we use the soundness of **2D** with respect to MD (Theorem 3.18), and specify a model based on a matrix FWDE falsifying $\Box(@p \to p)$. This is done by a model $\mathcal{M}$ based on $\mathcal{F}^2$ such that its valuation $V$ maps $p$ to $V(p) = \{\langle 0, 0\rangle\}$:

| $p$ | 0 | 1 |
|---|---|---|
| 0 | T | F |
| 1 | F | F |

$\mathcal{M}, \langle 1, 0\rangle \Vdash @p \wedge \neg p$, so $\mathcal{M}, \langle 0, 0\rangle \Vdash \neg\Box(@p \to p)$. Hence MD $\nVdash \Box(@p \to p)$, and therefore $\nvdash_{\mathbf{2D}} \Box(@p \to p)$. So **2D** is not closed under $Gen_\Box$. □

Note that the proof of this result also shows that **2D** is not closed under the rule of replacement of equivalents: although $\vdash_{\mathbf{2D}} p \leftrightarrow @p$ (by $T_@$ and $D_{c@}$) and $\vdash_{\mathbf{2D}} \Box(p \to p)$, we have $\nvdash_{\mathbf{2D}} \Box(@p \to p)$. However, as noted in section 2.1.3, QNMLs are closed under replacing formulas that are equivalent in their kernel, so as $\mathbf{2Dg} \subseteq \ker(\mathbf{2D})$, we have at least that **2D** is closed under replacing **2Dg**-equivalents.

Although **2D** is not closed under generalization for □, it is closed under the generalization rules for @ and $A$, as the following proposition shows:

**Proposition 3.25.** *If $\vdash_{\mathbf{2D}} \varphi$, then $\vdash_{\mathbf{2D}} @\varphi$ and $\vdash_{\mathbf{2D}} A\varphi$.*

*Proof.* Assume that $\vdash_{\mathbf{2D}} \varphi$. Then by definition, $\vdash_{\mathbf{2Dg}} @\varphi$. So by $Gen$, $\vdash_{\mathbf{2Dg}} @@\varphi$, and therefore $\vdash_{\mathbf{2D}} @\varphi$. Further, by $Gen$, $\vdash_{\mathbf{2Dg}} A@\varphi$. From $I4$ and $K_A$, it follows that $\vdash_{\mathbf{2Dg}} A@\varphi \to A\varphi$, so $\vdash_{\mathbf{2Dg}} A\varphi$. By $Gen$, $\vdash_{\mathbf{2Dg}} @A\varphi$, and therefore $\vdash_{\mathbf{2D}} A\varphi$. □

### 3.5.3 Reductions

I now turn to more philosophical questions about the logic of necessity, actuality and apriority according to two-dimensional semantics, as captured by **2D**. The first natural question is: what does two-dimensional semantics say about the logic of the individual modalities? @ does not have an interesting logic on its own, but □ and $A$ do.

I will show that if we look at formulas in **2D** that contain only □, the widely accepted logic of necessity $\mathbf{S5}_\Box$ will result, and that the analogous fact holds for $A$. (I indicate the operator used in a logic that has a standard name, like **S5**, by indexing this name with the operator.) In terms of the notion of conservativity defined in section 2.3.3, **2D** is a conservative extension of both $\mathbf{S5}_\Box$ and $\mathbf{S5}_A$. This is easily shown using Proposition 2.5 and the well-known fact that **S5** is sound and strongly complete with respect to universal frames (frames with a universal relation):

**Theorem 3.26.** **2D** *is a conservative extension of* $\mathbf{S5}_\Box$.

*Proof.* For any $\mathcal{F} \in \mathsf{RD}$, the subFWDE generated by any distinguished element of $\mathcal{F}|\Box$ is an FWDE based on a universal frame containing one point as a distinguished element. Further, any FWDE based on a universal frame containing one point as a distinguished element is isomorphic to an FWDE that can be derived in this way from an FWDE in RD. So $L(\mathsf{RD}|\Box)$ is the logic of the class of FWDEs based on a universal frame containing one point as a distinguished element. It is easy to see that this is the same as the logic of the class of universal frames, so the claim follows from Proposition 2.5. □

**Theorem 3.27.** **2D** *is a conservative extension of* $\mathbf{S5}_A$.

*Proof.* For any $\mathcal{F} \in \mathsf{RD}$, the subFWDE generated by any distinguished element of $\mathcal{F}|A$ is based on a universal frame, and contains all points as distinguished elements. Therefore, it validates the same $\mathcal{L}_A$-formulas as the frame obtained by removing the non-distinguished elements. Further, any universal frame is isomorphic to a frame that can be derived in this way from an FWDE in RD. Therefore $L(\mathsf{RD}|A)$ is the logic of the class of universal frames, and so the claim follows from Proposition 2.5. □

So **2D** is a conservative extension of both $\mathbf{S5}_\Box$ and $\mathbf{S5}_A$. In section 5.1.1, I will show that **2D** is also a conservative extension of the logic of necessity and actuality given by real-world validity as defined in Crossley and Humberstone (1977).

### 3.5.4 Interactions

Besides the logics for individual modalities that **2D** contains, I also want to sketch some aspects of the interaction of the modalities in **2D**, which will tell us something about the relations of the modalities in two-dimensional semantics. We have seen a bit about this in section 3.2, where it was shown that $N1 = Ap \to \Box Ap$ is a theorem of **2D**, but $N2 = \Box(Ap \to p)$ and $N3 = Ap \to \Box p$ are not. A good further test case are the following four principles, which are discussed in Anderson (1993, p. 4):

$$\begin{array}{ll} TP & Ap \to \Box p \\ HP & \Box p \to Ap \\ KP^+ & \Box p \to A\Box p \\ CP^+ & (\Box p \lor \Box\neg p) \to A(\Box p \lor \Box\neg p) \end{array}$$

*TP* is just $N3$, and I have just noted that this is not a theorem of **2D**. It can also be shown that none of the other principles hold in **2D**. This is done by a model $\mathcal{M}$ based on $\mathcal{F}^2$ such that its valuation $V$ maps $p$ to $V(p) = \{\langle 0,0\rangle, \langle 0,1\rangle, \langle 1,0\rangle\}$:

| $p$ | 0 | 1 |
|---|---|---|
| 0 | T | T |
| 1 | T | F |

*HP*, $KP^+$, and $CP^+$ are all false in $\mathcal{M}$ at $\langle 0,0\rangle$.

As with *TP*, that *HP* fails is just what was to be expected, since Kripke's examples of a posteriori necessities and a priori contingencies are one of the starting points for two-dimensional semantics. Similarly, that $KP^+$ fails is not surprising. As Anderson (1993, p. 5) mentions, $KP^+$ plausibly implies $HP$. In fact, it is not difficult to verify that $\vdash_{\mathbf{2D}} KP^+ \to HP$.

The principle $CP^+$ is more interesting. It says that if $p$ is not contingent, this fact is a priori. This is an interesting principle, since one may agree that Kripke's examples show that the truth-value of a necessary truth is not always knowable a priori, while still hoping that it is always a priori whether a statement is contingent or not. This thesis is partly captured by $CP^+$. A similar principle is mentioned in Edgington (2004, p. 11), which we can formalize as follows:

$$CP_1^+ \quad \Box p \to A(p \to \Box p)$$

According to this principle, although a necessary truth need not be a priori, it must at least be a priori that if it is true then it is necessary. It is plausible that $CP^+$ entails $CP_1^+$, and in fact, $\vdash_{\mathbf{2D}} CP^+ \to CP_1^+$, which is straightforward to verify semantically. Furthermore, $CP_1^+$ is also not a theorem of **2D**. For this, consider a model $\mathcal{M}$ based on $\mathcal{F}^2$ such that its valuation $V$ maps $p$ to $V(p) = \{\langle 0,0\rangle, \langle 1,0\rangle, \langle 1,1\rangle\}$:

| $p$ | 0 | 1 |
|---|---|---|
| 0 | T | T |
| 1 | F | T |

$CP_1^+$ is false in $\mathcal{M}$ at $\langle 0,0\rangle$, so according to **2D**, neither $CP^+$ nor $CP_1^+$ are theorems. Is this philosophically correct?

We can show that it is using an example mentioned by Edgington (which she attributes to Timothy Williamson, in conversation): let $p$ be a true identity statement like "Hesperus is Phosphorus" and $q$ an a posteriori contingent true predication like "John is happy". Then $p \vee q$ is necessary since $p$ is, but it is conceivable that $p$ is false while $q$ is true. In such a case, $p \vee q$ is the case, but since it is still contingent whether $q$ is true and $p$ is necessarily false, it is contingent whether $p \vee q$ is true. So although it is necessary that $p \vee q$, it is conceivable that $p \vee q$ is contingently true. As this example shows, it is philosophically correct that **2D** does not contain $CP_1^+$ or $CP^+$ as theorems.

### 3.5.5 Redundancy of @

As observed in Crossley and Humberstone (1977, section 3) and Hazen (1978), in the logic of necessity and actuality on which **2D** builds, the actuality operator is redundant in the following sense: for every formula, there is an equivalent formula not containing @. To conclude this chapter, I will now show that the analogous fact holds for **2D**. This may be of limited interest itself, but it will be useful for a completeness result in section 4.1. We should not take it to indicate that indexicals like "actually" are redundant in natural language as well; as shown in Hazen (1976), the redundancy of @ does not carry over to quantified modal logic.

The proof strategy for the result is this: I will first state some theorems which will allow us to give an inductive argument that shows that every formula can **2Dg**-equivalently be written in a normal form in which @ occurs unnestedly. Since **2Dg**-equivalence implies **2D**-equivalence and unnested occurrences of @ can be removed in **2D**, the redundancy of @ follows.

**Lemma 3.28.** *The following hold:*

(a) $\vdash_{\mathbf{2Dg}} p \leftrightarrow (p \vee @\bot)$

(b) $\vdash_{\mathbf{2Dg}} \Box(p \vee @q) \leftrightarrow (\Box p \vee @q)$

(c) $\vdash_{\mathbf{2Dg}} @@p \leftrightarrow @p$

(d) $\vdash_{\mathbf{2Dg}} A(p \vee @q) \leftrightarrow A(p \vee q)$

(e) $\vdash_{\mathbf{2D}} \bigwedge_{i\leq n}(p_i \vee @q_i) \leftrightarrow \bigwedge_{i\leq n}(p_i \vee q_i)$

*Proof.* These can be verified by semantic arguments using the completeness results of **2Dg** with respect to M and **2D** with respect to MD. □

**Lemma 3.29.** *Any $\varphi \in \mathcal{L}_{\Box @ A}$ is* **2Dg***-equivalent to $\bigwedge_{i\leq n}(\psi_i \vee @\chi_i)$ for some $n \in \mathbb{N}$ and $\psi_i, \chi_i \in \mathcal{L}_{\Box A}$ for all $i \leq n$.*

*Proof.* By induction on the complexity of $\varphi$. Note that **2Dg** is normal, and therefore the rule of replacement of equivalents holds.

- $\varphi = p$. Then $\varphi$ is **2Dg**-equivalent to $p \vee @\bot$ by Lemma 3.28 (a).
- $\varphi = \neg\varphi'$. By induction hypothesis, $\varphi'$ is **2Dg**-equivalent to some $\bigwedge_{i\leq n}(\psi_i \vee @\chi_i)$. So $\neg\varphi'$ is **2Dg**-equivalent to some $\bigvee_{i\leq n}(\neg\psi_i \wedge \neg@\chi_i)$. Switching from disjunctive to conjunctive normal form, this is **2Dg**-equivalent to a conjunction of disjunctions of formulas $\neg\psi_i$, $\neg@\chi_i$ and their negations. Since @ distributes over Boolean operators in **2Dg** (by $D_@$ and $D_{c@}$), $\neg\varphi'$ is **2Dg**-equivalent to a formula of the required form.
- $\varphi = \varphi' \wedge \varphi''$. Immediate by induction hypothesis.
- $\varphi = \Box\varphi'$. By induction hypothesis, $\varphi'$ is **2Dg**-equivalent to some $\bigwedge_{i\leq n}(\psi_i \vee @\chi_i)$. Normal modalities distribute over conjunction, so $\Box\varphi'$ is **2Dg**-equivalent to $\bigwedge_{i\leq n}\Box(\psi_i \vee @\chi_i)$. By Lemma 3.28 (b), this is is **2Dg**-equivalent to $\bigwedge_{i\leq n}(\Box\psi_i \vee @\chi_i)$.
- $\varphi = @\varphi'$. By induction hypothesis, $\varphi'$ is **2Dg**-equivalent to some $\bigwedge_{i\leq n}(\psi_i \vee @\chi_i)$. @ distributes over Boolean operators in **2Dg**, so $@\varphi'$ is **2Dg**-equivalent to $\bigwedge_{i\leq n}(@\psi_i \vee @@\chi_i)$. By Lemma 3.28 (c), this is **2Dg**-equivalent to $\bigwedge_{i\leq n}(@\psi_i \vee @\chi_i)$. Again by distribution of @ over Boolean operators, this is **2Dg**-equivalent to $@\bigwedge_{i\leq n}(\psi_i \vee \chi_i)$, and therefore to $\bot \vee @\bigwedge_{i\leq n}(\psi_i \vee \chi_i)$.
- $\varphi = A\varphi'$. By induction hypothesis, $\varphi'$ is **2Dg**-equivalent to some $\bigwedge_{i\leq n}(\psi_i \vee @\chi_i)$. Normal modalities distribute over conjunction, so $A\varphi'$ is **2Dg**-equivalent to $\bigwedge_{i\leq n} A(\psi_i \vee @\chi_i)$. By Lemma 3.28 (d), this is is **2Dg**-equivalent to $\bigwedge_{i\leq n} A(\psi_i \vee \chi_i)$. By Lemma 3.28 (a) this is is **2Dg**-equivalent to $\bigwedge_{i\leq n} A(\psi_i \vee \chi_i) \vee @\bot$.

□

**Theorem 3.30.** *Every formula is* **2D***-equivalent to one not containing @.*

*Proof.* Let $\varphi \in \mathcal{L}_{\Box @ A}$. By Lemma 3.29, there are $\psi_i, \chi_i \in \mathcal{L}_{\Box A}$ such that $\varphi$ is **2Dg**-equivalent to $\bigwedge_{i\leq n}(\psi_i \vee @\chi_i)$. Since $\mathbf{2Dg} \subseteq \mathbf{2D}$, the formulas are also **2D**-equivalent. By Lemma 3.28 (e), it follows that $\varphi$ is **2D**-equivalent to $\bigwedge_{i\leq n}(\psi_i \vee \chi_i)$, and so to a formula not containing @. □

# Chapter 4

# Comparisons

As stated in section 1.3.3, there are two texts that are concerned with constructing a logic similar to the one I have presented in the last chapter, namely Davies and Humberstone (1980) and Restall (2010). In this chapter, I will discuss them and their relation to the logic developed in the last chapter, to show why the latter is important for a formal understanding of two-dimensional semantics in addition to the former. Throughout, I will freely adapt notation to make the presentation more uniform. I will consider the papers in turn.

## 4.1 Davies and Humberstone (1980)

Davies and Humberstone (1980) is an influential article in the development of two-dimensional semantics. It can be seen as an attempt to formalize observations made in Evans (1979) on Kripke's discussion of necessity and apriority using a formal apparatus that is an extension of the one introduced in Crossley and Humberstone (1977). I will first present Evans' observations, and then show how Davies and Humberstone propose to formally capture them. After this, I will compare Davies and Humberstone's logic to the one presented in chapter 3, and show how a variant of it can be defined that lies between the two, which can be used to formalize the logic of an interesting fragment of natural language according to two-dimensional semantics.

### 4.1.1 Deep and Superficial Necessity

Evans (1979) is concerned with Kripke's claim that there are contingent a priori truths. To argue that this is less surprising than it is often taken to be, Evans distinguishes two notions of necessity, which he calls *deep* and *superficial* necessity. He introduces these as follows, via the corresponding notions of contingency:

> Whether a statement is deeply contingent depends upon what makes it true; whether a statement is superficially contingent depends upon how it embeds inside the scope of modal operators. (Evans (1979, p. 161))

He does not elaborate much on this, but rather illustrates the distinction with the following example: let us stipulate to use the name "Julius" to refer to whoever invented the zip. Evans claims that then, "necessarily, Julius invented the zip" is false, since whoever invented the zip could also have done something else. Therefore it is not superficially necessary that Julius invented the zip. However, he claims that it is deeply necessary that Julius invented the zip, since the statement that Julius invented the zip is made true by every world. In this particular example, the difference comes down to whether we seek the referent of "Julius" via the description with which we introduced the name in the actual world (superficial necessity), or in any counterfactual world considered (deep necessity).

Evans applies these observations to Kripke's claim that some truths are a priori but contingent as follows: he states that in virtue of the way the name "Julius" was introduced above, it is a priori that Julius invented the zip. So this truth is an example for one that is a priori but contingent – if contingency is understood as superficial contingency. This is supposed to show that Kripke's examples of a priori contingencies are not as surprising or problematic as they might seem. Evans claims that if it seems counter-intuitive that some truths are contingent and a priori, then this is because we are thinking of deep contingency. But Kripke did not show that there are deeply contingent a priori truths, only that there are superficially contingent a priori truths, and these, according to Evans, are unproblematic.

Evans is not explicit about the relation between apriority and deep necessity besides claiming that "it would be intolerable for there to be a statement which is both knowable *a priori* and deeply contingent" (Evans (1979, p. 161)). In Davies and Humberstone (1980, p. 10), Davies (1981, pp. 240–241), and Davies (2004, sections 4 and 5), it is suggested that the notions may not coincide.

### 4.1.2 "Fixedly" and the Two Notions of Necessity

Davies and Humberstone (1980) develop a formal logic to capture Evans' distinction between deep and superficial necessity. To do so, they start with the logic of necessity and actuality presented in Crossley and Humberstone (1977).

The semantics of the latter logic is roughly as follows: models are Kripke models without an accessibility relation but with an added single world called the *designated actual world*. (In this section, I will call the points of models "worlds", as they represent possible worlds, in contrast to the semantics used in chapter 3.) $\Box$ is interpreted as truth in all worlds, and @ as truth in the designated actual world. This is evidently a notational variant of the class of $\Box$@-frames in which $R_\Box$ is the universal relation, and there is some point $w$ such that every point is $R_@$-related to $w$ and only $w$. Crossley and Humberstone (1977) also present an axiomatization for general validity, which will be discussed in section 5.1.1.

Davies and Humberstone take $\Box$ to model superficial necessity. To model deep necessity, they first introduce a new operator, $\mathscr{F}$, read "fixedly". Its semantics is as follows: $\mathscr{F}\varphi$ is true in a model and a world if it is true in that world in each model that can be obtained from the original one by changing the designated actual world to another world. This means that $\mathscr{F}@\varphi$ is true in a model and a world if it is true in the designated actual world in each model obtained from the original one by changing the designated actual world to another world. Davies and Humberstone take Evans' notion of being made true by a world (as in the explanation of deep necessity) to be captured by the formal notion of being true in a world in the model in which that world is set to be the designated actual world. Therefore, they suggest that $\mathscr{F}@$ ("fixedly actually") captures Evans' notion of deep necessity.

An axiomatization of the logic of necessity, actuality, and fixedly given by this semantics is presented in Davies and Humberstone (1980), but for the completeness proof, Davies and Humberstone refer to a forthcoming paper called 'The logic of "fixedly"'. This paper never appeared, although the material was printed as Appendix 10 of Davies (1981).

One unusual feature of this logic is the fact that it is not closed under uniform substitution. This is surprising since normally, proposition letters in modal logics represent arbitrary sentential expressions, from which it follows that uniform substitution should be valid. One can try to explain the failure of uniform substitution in Davies and Humberstone's logic by stating that in it, proposition letters can only represent sentential expressions that do not contain "actually". Since some principles may hold for all such expressions, but not for ones containing "actually", the corresponding formula may be valid without some substitution instances introducing "actually" being valid.

This explanation can also be used to give an account of another surprising feature of the logic, which is discussed in Davies and Humberstone (1980, p. 11), namely that Evans' example for a contingent a priori truth cannot be represented in the logic. The problem is this: ignoring the temporal component, the sentence "Julius invented the zip" is a predication, so it should be represented by a proposition letter, say $p$. However, it is a theorem that $\mathscr{F}@p \leftrightarrow \Box p$, which conflicts with Evans' claim that "Julius invented the zip" is deeply necessary but superficially contingent. It can now be claimed that the descriptive name "Julius" is an abbreviation for the rigidified description "the person who actually invented the zip". Then $p$ may not stand for "Julius invented the zip" since this covertly contains "actually". Hence to model this sentence, it would have to be further analyzed, but we cannot do this here, as the logic does not contain quantificational resources.

### 4.1.3 The Logics Compared

Davies and Humberstone's logic and the logic presented in chapter 3 were meant to capture different things. The former is meant to capture the difference between superficial necessity and deep necessity according to Evans, whereas the latter is meant to capture the difference between necessity and apriority according to two-dimensional semantics. So it is immediately clear that we can't just use Davies and Humberstone's logic to formalize two-dimensional semantics, although it might turn out that their technical apparatus can be re-interpreted for such a purpose. I will start comparing the logics on a semantic level by comparing their definitions of models.

The central observation is that every model in the sense used by Davies and Humberstone corresponds to a model based on a matrix frame in which the valuation is the same for every row, and vice versa. The evaluation clauses for $\Box$ and $@$ in such corresponding models are exactly analogous, and the evaluation condition for $\mathscr{F}@$ corresponds exactly to that of $A$. This equivalence is best illustrated graphically. We can display a model of Davies and Humberstone's semantics by listing the worlds horizontally, and marking the distinguished actual world by enclosing it in parentheses. E.g., a model $\mathfrak{M}$ based on the set

$\{0, 1, 2, 3\}$ in which 0 is distinguished has the following structure:

$$(0) \quad 1 \quad 2 \quad 3$$

Of course, the valuation function $V$, which maps every proposition letter to a set of worlds, is not captured in the picture. Since $\mathscr{F}$ varies the distinguished actual world of the model, its semantic function can be illustrated by listing the variants of a given model vertically. For the above example, this looks like this:

$$\begin{array}{cccc} (0) & 1 & 2 & 3 \\ 0 & (1) & 2 & 3 \\ 0 & 1 & (2) & 3 \\ 0 & 1 & 2 & (3) \end{array}$$

Just like we can think of $\Box$ as quantifying over the worlds in one row, we can think of $\mathscr{F}$ as quantifying over the worlds in one column. Now take the matrix frame based on $\{0, 1, 2, 3\}$, and the valuation that maps any proposition letter to the set of points $\langle x, y\rangle$ such that $x \in V(p)$. It is not difficult to see that the evaluation of formulas proceeds exactly analogous, if we start from a formula in $\mathcal{L}_{\Box @ A}$ and translate $A$ as $\mathscr{F}@$ when evaluating it in $\mathfrak{M}$. In general, we can find a corresponding model based on a matrix frame for every model in Davies and Humberstone's sense. Also, as long as the valuation of a model based on a matrix frame makes a proposition letter true at a point if and only if makes it true at every point in the same column, we can find a corresponding model in Davies and Humberstone's sense. This correspondence could be made formally precise, but I hope the example has made the matter sufficiently clear.

The correspondence of the models of Davies and Humberstone's logic to some models based on matrix frames shows that essentially, their logic differs from the logic used in chapter 3 in three respects: firstly, their model theory is a restriction on the valuations on matrix frames; secondly, they use a different set of primitive modal operators than the one used in chapter 3 (instead of $A$, they have $\mathscr{F}$); and thirdly, Davies and Humberstone use general validity, whereas I use real-world validity.

The second and third of these aspects are not essential in Davies and Humberstone's formalization. Concerning the second, it is remarked in Davies (2004, pp. 92–93) that since Davies and Humberstone (1980) was mainly concerned with formalizing deep necessity, they could also have introduced a primitive operator with the semantics of $\mathscr{F}@$ instead of $\mathscr{F}$. This can even be seen as the more natural option, since there seems to be no natural language expression whose semantics is captured by $\mathscr{F}$ – "fixedly" certainly does not have this meaning. Concerning the third aspect, as Humberstone (2004, pp. 22–23) notes, there are compelling reasons for real-world validity. These will also be discussed in greater detail in the appendix.

Therefore, Davies and Humberstone's system can be changed in two of the three aspects in which it differs from the system presented in chapter 3 without distorting the philosophical picture. What remains is just a restriction in the models used for the semantics. It should be noted that it is not clear whether this restriction should really be in place for a model of deep and superficial necessity. E.g., it is a quite substantial assumption that every deeply contingent but superficially necessary truth involves "actually" in some way. Therefore, Davies and Humberstone's logic may also have to be changed in the first aspect, which would just give us the logic **2D** for deep and superficial necessity (where $A$ represents deep necessity). See also Restall (2010, especially footnote 1), for a discussion of this.

Whether or not the restriction on models is justified, it is interesting to ask what consequences it has for the logic. Furthermore, I will argue below that a variant of Davies and Humberstone's logic in which only the second and third aspect are adapted can also be used to give a model of two-dimensional semantics for an interesting fragment of natural language. Therefore, I will investigate the logic that results from restricting the models on frames in $\mathsf{MD}$ in the way indicated in the next section.

### 4.1.4 A Variant of Davies and Humberstone's Logic

If we adopt real-world validity and interpret $A$ analogously to $\mathscr{F}@$, then the following semantics corresponds to the one in Davies and Humberstone (1980):

**Definition 4.1.** *A* Davies & Humberstone model *is a model based on a matrix* FWDE *based on a set $S$ such that its valuation $V$ satisfies the following condition:*

(†) *For all $p$ and $x, y, z \in S$, $\langle x, y\rangle \in V(p)$ if and only if $\langle x, z\rangle \in V(p)$.*

*Let* $\mathsf{DHD}$ *be the class of Davies & Humberstone models.*

Condition (†) expresses the condition that the valuation function must not vary the interpretation of proposition letters across worlds considered as actual if the world considered as counterfactual is kept constant, which was motivated and related to Davies and Humberstone's semantics above. Note that $\mathsf{DHD}$ is a subclass of the models based on matrix FWDEs. Therefore, the logic of this class of models is a superset of **2D**.

An important observation is that in the absence of the actuality operator, the notions of $\Box$ and $A$ coincide in the logic of $\mathsf{DHD}$ in the following sense:

**Proposition 4.2.** *If $\varphi \in \mathcal{L}_{\Box A}$, then $\mathsf{DHD} \Vdash \Box\varphi \leftrightarrow A\varphi$.*

*Proof.* Let $\mathcal{M} = \langle W, R_\Box, R_@, R_A, D, V\rangle \in \mathsf{DHD}$ such that $W = S \times S$. We first prove the following claim for all $\varphi \in \mathcal{L}_{\Box A}$ by induction on the complexity of $\varphi$: for all $x, y, z \in S$, $\mathcal{M}, \langle x, y\rangle \Vdash \varphi$ iff $\mathcal{M}, \langle x, z\rangle \Vdash \varphi$. The induction case for proposition letters is just condition (†) on Davies & Humberstone models. The cases for Boolean operators are immediate by induction.

Let $\varphi = \Box\psi$. $\mathcal{M}, \langle x, y\rangle \Vdash \Box\psi$ iff $\mathcal{M}, \langle x', y\rangle \Vdash \psi$ for all $x' \in S$. By induction hypothesis, this is the case iff $\mathcal{M}, \langle x', z\rangle \Vdash \psi$ for all $x' \in S$, which is in turn the case iff $\mathcal{M}, \langle x, z\rangle \Vdash \Box\psi$.

Finally, consider the case where $\varphi = A\psi$. $\mathcal{M}, \langle x, y\rangle \Vdash A\psi$ iff $\mathcal{M}, \langle x', x'\rangle \Vdash \psi$ for all $x' \in S$ iff $\mathcal{M}, \langle x, z\rangle \Vdash A\psi$.

Now let $x \in S$ and $\varphi \in \mathcal{L}_{\Box A}$. $\mathcal{M}, \langle x, x\rangle \Vdash \Box\varphi$ iff $\mathcal{M}, \langle y, x\rangle \Vdash \varphi$ for all $y \in S$. By the claim just proven, this is the case iff $\mathcal{M}, \langle y, y\rangle \Vdash \varphi$ for all $y \in S$, which is the case iff $\mathcal{M}, \langle x, x\rangle \Vdash A\varphi$. □

With this proposition, it is easy to show that the logic of $\mathsf{DHD}$ is not closed under uniform substitution, and therefore not a QNML:

**Proposition 4.3.** *The logic of* $\mathsf{DHD}$ *is not closed under uniform substitution.*

*Proof.* By Proposition 4.2, $\mathsf{DHD} \Vdash \Box p \leftrightarrow Ap$. We show that $\mathsf{DHD} \nVdash \Box @p \leftrightarrow A@p$. Consider a model $\mathcal{M}$ based on $\mathcal{F}^2$ such that its valuation $V$ maps $p$ to $V(p) = \{\langle 0,0\rangle, \langle 0,1\rangle\}$:

| $p$ | 0 | 1 |
|---|---|---|
| 0 | T | F |
| 1 | T | F |

Note that $\mathcal{M} \in \mathsf{DHD}$. $\mathcal{M}, \langle 0,0\rangle \Vdash p$, therefore $\mathcal{M}, \langle 0,0\rangle \Vdash \Box @p$, but $\mathcal{M}, \langle 1,1\rangle \nVdash p$, and therefore $\mathcal{M}, \langle 0,0\rangle \nVdash A@p$. So $\mathsf{DHD} \nVdash \Box @p \leftrightarrow A@p$. □

Furthermore, we can axiomatize the logic of DHD by taking the $A$-generalizations of the validities of Proposition 4.2 and the theorems of **2D** as axioms and modus ponens as the single rule:

**Definition 4.4.** *Let $P$ be the proof system that contains as axioms all theorems of* **2D** *and the instances of $A(\Box\varphi \leftrightarrow A\varphi)$ for all $\varphi \in \mathcal{L}_{\Box A}$, and MP as the single rule. Let* $\mathbf{DH} = L(P)$.

**Theorem 4.5.** **DH** *is sound and strongly complete with respect to* DHD.

*Proof. Soundness:* By induction on the proof system in Definition 4.4. Since every model in DHD is based on a matrix FWDE, every such model validates **2D**, as proven in Theorem 3.18. By Proposition 4.2, the instances of $\Box\varphi \leftrightarrow A\varphi$ for all $\varphi \in \mathcal{L}_{\Box A}$ are valid on DHD, and since only distinguished elements are $A$-accessible from distinguished elements in matrix FWDEs, the instances of $A(\Box\varphi \leftrightarrow A\varphi)$ for all $\varphi \in \mathcal{L}_{\Box A}$ are also valid on DHD. Finally, the set of formulas that are true at a point in a model is closed under modus ponens. So $\mathsf{DHD} \Vdash \mathbf{DH}$.

*Strong Completeness:* As usual, we make use of the fact that **DH** is strongly complete with respect to DHD if every **DH**-consistent set is satisfiable on DHD. (See Blackburn et al. (2001, Theorem 4.12).) Assume that $\Gamma$ is a **DH**-consistent set. Then by Lindenbaum's lemma, which is straightforward to prove, there is a **DH**-MCS $\Delta \supseteq \Gamma$.

To prove completeness, we will use the canonical model for **2D**. The strategy is to take its submodel generated by $\Delta$, and use it to define a DHD-model that contains a point corresponding to $\Delta$ that makes the same formulas true, which shows that $\Gamma$ is satisfiable on DHD. So let $\mathcal{M}^{\mathbf{2D}}$ be the canonical model for **2D**. This is constructed from the canonical frame for ker(**2D**) by adding the **2D**-MCS as distinguished elements and the canonical valuation as usual. Let $\mathcal{M} = \langle W, R_\Box, R_@, R_A, D, V\rangle$ be $(\mathcal{M}^{\mathbf{2D}})_\Delta$, that is, the point-generated submodel of $\mathcal{M}^{\mathbf{2D}}$ generated by $\Delta$. By the truth lemma for canonical models and the preservation of truth under generated submodels, $\mathcal{M}, \Delta \Vdash \Delta$.

To define the DHD-model, let $X = R_A[\{\Delta\}]$, and $\mathcal{M}' = \langle W', R'_\Box, R'_@, R'_A, D', V'\rangle$ be the model based on the matrix FWDE based on the set $X$ (so $W' = X \times X$) where $\langle w, v\rangle \in V'(p)$ iff $w \in V(p)$, for all proposition letters $p$. That $\mathcal{M}' \in \mathsf{DHD}$ is immediate by construction. We now prove for all $\varphi \in \mathcal{L}_{\Box A}$ the following claim: for all $w \in X$, $\mathcal{M}, w \Vdash \varphi$ iff $\mathcal{M}', \langle w, w\rangle \Vdash \varphi$. We do so by induction on the complexity of $\varphi$. The case of proposition letters follows by the definition of $V'$. The cases for Boolean operators are immediate by induction.

Let $\varphi = A\psi$. Consider any $w \in X$. We start by showing that $R_A[\{w\}] = X$. Using the characterization of **2D** as a join (Theorem 3.22) it follows by the proof of Theorem 2.7 that **2D** is valid in its canonical FWDE. As $\Delta$ is a **2D**-MCS, it is a distinguished element of this FWDE, so since $\vdash_{\mathbf{2D}} T_A$, it follows that $\Delta R_A \Delta$, and therefore $\Delta \in X$. First, assume that $v \in R_A[\{w\}]$. Then since $\vdash_{\mathbf{2D}} 4_A$, also $v \in X$. Now, assume that $v \in X$. Then since $\vdash_{\mathbf{2D}} 5_A$, also $v \in R_A[\{w\}]$.

With this, it follows that $\mathcal{M}, w \Vdash A\varphi$ iff $\mathcal{M}, v \Vdash \varphi$ for all $v \in X$. By induction hypothesis, this is the case iff $\mathcal{M}', \langle v, v\rangle \Vdash \varphi$ for all $v \in X$. This is the case iff $\mathcal{M}', \langle w, w\rangle \Vdash A\varphi$.

Finally, let $\varphi = \Box\psi$. Consider any $w \in X$. Since $\Delta$ is a **DH**-MCS, $A(\Box\psi \leftrightarrow A\psi) \in \Delta$, and so $\mathcal{M}, \Delta \Vdash A(\Box\psi \leftrightarrow A\psi)$. Therefore $\mathcal{M}, v \Vdash \Box\psi \leftrightarrow A\psi$ for all $v \in X$, in particular $\mathcal{M}, w \Vdash \Box\psi \leftrightarrow A\psi$. So $\mathcal{M}, w \Vdash \Box\psi$ iff $\mathcal{M}, w \Vdash A\psi$. As the previous induction clause has shown this is the case iff $\mathcal{M}', \langle w, w\rangle \Vdash A\psi$. Since $\mathcal{M}' \in \mathsf{DHD}$, it follows from Proposition 4.2 that $\mathcal{M}', \langle w, w\rangle \Vdash \Box\psi \leftrightarrow A\psi$. So $\mathcal{M}', \langle w, w\rangle \Vdash A\psi$ iff $\mathcal{M}', \langle w, w\rangle \Vdash \Box\psi$. Together, these equivalences show that $\mathcal{M}, w \Vdash \Box\psi$ iff $\mathcal{M}', \langle w, w\rangle \Vdash \Box\psi$, as required. This concludes the induction.

As shown above, $\Delta \in X$. So for all $\varphi \in \mathcal{L}_{\Box A}$, $\mathcal{M}, \Delta \Vdash \varphi$ iff $\mathcal{M}', \langle \Delta, \Delta\rangle \Vdash \varphi$. Now consider any $\varphi \in \mathcal{L}_{\Box @ A}$. By Theorem 3.30, there is a $\varphi' \in \mathcal{L}_{\Box A}$ such that $\vdash_{\mathbf{2D}} \varphi \leftrightarrow \varphi'$. Since $\mathcal{M}, \Delta \Vdash \Delta$ and $\mathbf{2D} \subseteq \Delta$, $\mathcal{M}, \Delta \Vdash \varphi \leftrightarrow \varphi'$. Since $\mathcal{M}'$ is a matrix model, $\mathcal{M}' \Vdash \mathbf{2D}$, and therefore $\mathcal{M}', \langle \Delta, \Delta\rangle \Vdash \varphi \leftrightarrow \varphi'$. It follows that $\mathcal{M}, \Delta \Vdash \varphi$ iff $\mathcal{M}, \Delta \Vdash \varphi'$ iff $\mathcal{M}', \langle \Delta, \Delta\rangle \Vdash \varphi'$ (as proven above) iff $\mathcal{M}', \langle \Delta, \Delta\rangle \Vdash \varphi$. So as $\mathcal{M}, \Delta \Vdash \Gamma$, also $\mathcal{M}', \langle \Delta, \Delta\rangle \Vdash \Gamma$. □

It may be interesting to note that the development of the logic in chapter 3 and the variant obtained by taking a subclass of models in this section is in some ways analogous to the work in Segerberg (1973). There, Segerberg defines a logic based on frames which can be seen as a generalization of matrix frames, and presents a complete axiomatization. Referring to philosophical motivation from Åqvist (1973), he considers the class of models based on these frames that satisfy condition (†). He then axiomatizes the logic of these models by adding as axiom schema similar to the one used in Definition 4.4 to the axiomatization of the logic of his class of frames.

### 4.1.5 Re-Interpreting the Variant

**DH** is a natural variant of Davies and Humberstone's logic of deep and superficial necessity. But we can also go back to the original interpretation of $A$ as apriority and ask whether there is some way of understanding **DH** on which it captures the ideas of two-dimensional semantics. As noted earlier, since **DH** is not closed under uniform substitution, we cannot understand proposition letters as representing arbitrary sentential expressions. But the restriction on the valuations used in the definition of $\mathsf{DHD}$ suggests that we can interpret **DH** as the logic of two-dimensional semantics on the assumption that proposition letters may only stand for sentences in which the truth value in a possible world is independent of the scenario considered as actual. Following Chalmers (2004, pp. 191–193), we can call these the *neutral* sentences.

This interpretation of **DH** is interesting since it highlights a connection between necessity and apriority in two-dimensional semantics that is obscured in a logic like the one discussed in chapter 3. Considering only the fragment of natural language modeled by the formal syntax on the assumption that propositional letters represent neutral sentences, this connection consists in the fact that the only difference between "necessity" and "apriority" is how "actually" embeds in them. Of course, this restriction rules out most interesting cases, like "Hesperus is Phosphorus". But it seems likely that the connection can be brought out more clearly and with weaker restrictions in a quantified logic.

## 4.2 Restall (2010)

Restall (2010) discusses a logical system very close to the one presented in chapter 3. He also considers a propositional modal logic containing three connectives representing necessity, actuality, and apriority. As we will see, Restall's logic of this language is essentially the same as the one given by the matrix semantics used in chapter 3, although he approaches it from a very different angle than I have above. He starts with a proof system for the language $\mathcal{L}_{\Box @ A}$ formulated in a generalization of a Gentzen-style sequent calculus that may be called a "two-dimensional hypersequent system". This is a generalization of the hypersequent calculus for the unimodal logic **S5** presented in Restall (2008). Applying a strategy from Restall (2009), he then goes on to extract a model-theoretic semantics from his calculus, for which he shows that the calculus is sound and complete (this will be made precise below). I will not describe this strategy in detail, but only outline the kind of proof system he deals with, and the resulting semantics, as far as it is necessary to compare it to the work in chapter 3.

### 4.2.1 Restall's Logic

Standard sequent systems are proof systems in which the individual lines of proofs are *sequents*, instead of formulas. Restall understands sequents to be of the form $\Gamma \Rightarrow \Delta$, where $\Gamma$ and $\Delta$ are finite multisets of formulas. The sequent $\Gamma \Rightarrow \Delta$ can be interpreted as expressing that the disjunction of $\Delta$ is a consequence of (the conjunction of) $\Gamma$. Restall suggests another way of interpreting it: we can also understand it as precluding asserting all formulas in $\Gamma$ while at the same time denying all formulas in $\Delta$.

Some sequents, such as $\Gamma \Rightarrow \Gamma$, are taken as axioms in the proof system. Further, a number of rules are specified, which state which sequents may be deduced from which sequents. E.g., one might have a rule that says that from the sequent $\Gamma, \varphi, \psi \Rightarrow \Delta$, the sequent $\Gamma, \varphi \wedge \psi \Rightarrow \Delta$ can be deduced, where $\Theta, \chi$ is the the result of adding $\chi$ (once more) to $\Theta$. This is usually written as follows:

$$\frac{\Gamma, \varphi, \psi \Rightarrow \Delta}{\Gamma, \varphi \wedge \psi \Rightarrow \Delta}$$

On Restall's interpretation, this rule says that if asserting $\varphi$, $\psi$, and the formulas in $\Gamma$ precludes denying the formulas in $\Delta$, then asserting $\varphi \wedge \psi$ and the formulas in $\Gamma$ also precludes denying the formulas in $\Delta$.

To give a proof system of such a kind for **S5**, Restall (2008) generalizes the notion of a sequent to that of a hypersequent, which is a multiset of sequents. He writes $\Gamma_0 \Rightarrow \Delta_0 \mid \ldots \mid \Gamma_n \Rightarrow \Delta_n$ for the hypersequent consisting of the sequents $\Gamma_0 \Rightarrow \Delta_0, \ldots,$ and $\Gamma_n \Rightarrow \Delta_n$. Restall intends a hypersequent to be understood by assuming that each sequent is evaluated under a different counterfactual supposition. So a hypersequent of the form $\Gamma_0 \Rightarrow \Delta_0 \mid \Gamma_1 \Rightarrow \Delta_1$ precludes asserting the formulas in $\Gamma_0$ while denying the formulas in $\Delta_0$ under one counterfactual supposition, while at the same time asserting the formulas in $\Gamma_1$ while denying the formulas in $\Delta_1$ under another counterfactual supposition.

To give a sequent system for $\mathcal{L}_{\Box @ A}$, Restall generalizes hypersequents further to two-dimensional hypersequents, which are multisets of hypersequents, where each hypersequent contains a unique sequent called the *actual sequent*, which is marked as $\Rightarrow_@$. Separating individual hypersequents in a two-dimensional hypersequent by $||$, he writes two-dimensional hypersequents as follows:

$$\begin{array}{c} \Gamma_0^0 \Rightarrow_@ \Delta_0^0 \mid \Gamma_1^0 \Rightarrow \Delta_1^0 \mid \ldots \mid \Gamma_{n_0}^0 \Rightarrow \Delta_{n_0}^0 \mid\mid \\ \vdots \\ \Gamma_0^m \Rightarrow_@ \Delta_0^m \mid \Gamma_1^m \Rightarrow \Delta_1^m \mid \ldots \mid \Gamma_{n_m}^m \Rightarrow \Delta_{n_m}^m \end{array}$$

The idea is that the different hypersequents in such a two-dimensional hypersequent are evaluated under different indicative suppositions, where in every one of them, the actual sequent precludes assertions and denials under the indicative supposition, and the other sequents preclude assertions and denials under this indicative supposition and different counterfactual suppositions. So two-dimensional hypersequents can still be understood as precluding assertions and denials, although now under potentially complicated clauses of supposition.

To be able to state the rules of the system succintly, Restall introduces the following notation: for any two-dimensional hypersequent $\mathcal{H}[\Gamma \Rightarrow \Delta]$, $\mathcal{H}[\Gamma' \Rightarrow \Delta']$ is $\mathcal{H}$, with $\Gamma \Rightarrow \Delta$ replaced by $\Gamma' \Rightarrow \Delta'$. Besides structural rules and rules for the Boolean operators, Restall's system contains the following rules for the modal operators:

$$\frac{\mathcal{H}[\Gamma \Rightarrow \Delta \mid \Gamma', \varphi \Rightarrow \Delta']}{\mathcal{H}[\Gamma, \Box\varphi \Rightarrow \Delta \mid \Gamma' \Rightarrow \Delta']}\,[\Box L] \qquad \frac{\mathcal{H}[\Rightarrow \varphi \mid \Gamma \Rightarrow \Delta]}{\mathcal{H}[\Gamma \Rightarrow \Box\varphi, \Delta]}\,[\Box R]$$

$$\frac{\mathcal{H}[\Gamma \Rightarrow \Delta \mid \Gamma', \varphi \Rightarrow_@ \Delta']}{\mathcal{H}[\Gamma, @\varphi \Rightarrow \Delta \mid \Gamma' \Rightarrow_@ \Delta']}\,[@L] \qquad \frac{\mathcal{H}[\Gamma \Rightarrow \Delta \mid \Gamma' \Rightarrow_@ \varphi, \Delta']}{\mathcal{H}[\Gamma \Rightarrow @\varphi, \Delta \mid \Gamma' \Rightarrow_@ \Delta']}\,[@R]$$

$$\frac{\mathcal{H}[\Gamma \Rightarrow \Delta \mid\mid \Gamma', \varphi \Rightarrow_@ \Delta']}{\mathcal{H}[\Gamma, A\varphi \Rightarrow \Delta \mid\mid \Gamma' \Rightarrow_@ \Delta']}\,[AL] \qquad \frac{\mathcal{H}[\Rightarrow_@ \varphi \mid\mid \Gamma \Rightarrow \Delta]}{\mathcal{H}[\Gamma \Rightarrow A\varphi, \Delta]}\,[AR]$$

Since this is the only sequent calculus I will consider here, I will just say that a two-dimensional hypersequent is *derivable* if it is derivable in this system.

Generalizing a limit construction from Restall (2009), Restall goes on to show that a model-theoretic semantics can be derived from these two-dimensional hypersequents. This semantics is a notational variant of the class of models based on Restall frames as defined in chapter 3 – hence their name. I will ignore this notational difference in the following and work with models as defined above.

Restall defines a notion of two-dimensional hypersequents failing and holding in such a model: a two-dimensional hypersequent *fails* in a model if for every sequent $\Gamma \Rightarrow \Delta$ contained, there is a point in the model in which each member of $\Gamma$ is true and each member of $\Delta$ is false, such that for every hypersequent, these points for its sequents are related by $R_\Box$, and the point for the actual sequent of this hypersequent is in the image of $R_@$. A two-dimensional hypersequent *holds* in a model if it doesn't fail there. Restall proves soundness and completeness in the sense that a two-dimensional hypersequent is derivable iff it is holds in all models. If we restrict ourselves to two-dimensional hypersequents of the forms $\Gamma \Rightarrow \varphi$ and $\Gamma \Rightarrow_@ \varphi$, it follows from this results that:

**Theorem 4.6** (Restall (2010))**.** *For any* $\Gamma$ *and* $\varphi$*:*

- $\Gamma \Rightarrow \varphi$ *is derivable if and only if* $\Gamma \vDash_{\mathsf{R}} \varphi$.
- $\Gamma \Rightarrow_{@} \varphi$ *is derivable if and only if* $\Gamma \vDash_{\mathsf{RD}} \varphi$.

### 4.2.2 Equivalence of the Logics

Given that **2Dg** and **2D** are sound and strongly complete with respect to R and RD, we can connect the two completeness theorems to show that Restall's sequent calculus and the Hilbert calculus of chapter 3 describe the same logic:

**Theorem 4.7.** *For any* $\Gamma$ *and* $\varphi$*:*

- $\Gamma \Rightarrow \varphi$ *is derivable if and only if* $\Gamma \vdash_{\mathbf{2Dg}} \varphi$.
- $\Gamma \Rightarrow_{@} \varphi$ *is derivable if and only if* $\Gamma \vdash_{\mathbf{2D}} \varphi$.

*Proof.* From Theorems 3.12, 3.20, and 4.6. □

Restall's paper therefore describes the same logic as I have presented above, using one way of specifying its semantics described there, but a completely different proof system. As usual when comparing Hilbert and Gentzen calculi, there are no obvious connections between specific axioms of the first and specific derivation rules of the second. Rather, it is not hard to see that the two-dimensional hypersequents correspond structurally to certain models, which is exploited in the derivation of the semantics from the proof theory.

Besides providing a proof system and semantics for $\mathcal{L}_{\Box @ A}$, Restall follows a more ambitious program in his paper. By starting with a proof system for which he gives independent motivation, and deriving a semantics from it that has strong structural connections to two-dimensional semantics (as described in Lemmas 3.13 and 3.14), he intends to argue for two-dimensional semantics. This chapter is not the place to discuss this undertaking. To do it justice, much would have to be said about proof theory and semantics that is not my concern in this text. I only want to note that a similar argument for two-dimensional semantics could be constructed using the completeness theorem of chapter 3: by arguing that all of the theorems of **2D** should be accepted, one may be able to argue for the semantic structure of $\mathsf{FrD}(\mathbf{2D})$.

Since Restall's logic and the logic presented in chapter 3 use essentially the same semantics, they can both be used to formalize two-dimensional semantics. But this does not mean that the work done in chapter 3 was superfluous. Having a Hilbert calculus in addition to the Gentzen calculus provided by Restall is very useful: as far as I can see, the Hilbert calculus is far more flexible. E.g. the completeness result in the last section on the variant of Davies and Humberstone's logic would probably be very difficult to obtain in Restall's setting. Likewise, I don't see how the work on logical commitments in the next chapter could be carried out using a Gentzen calculus. These remarks are not meant to show that the Hilbert calculus is better than Restall's Gentzen calculus. E.g., Restall's way of arguing *for* matrix frames using two-dimensional hypersequents has no direct equivalent in the Hilbert calculus. So the two calculi should be seen as different ways of describing the same logic, which are useful for different purposes.

# Chapter 5

# Logical Commitments of Two-Dimensional Semantics

In chapter 3, I specified the logic **2D** of necessity, actuality and apriority according to two-dimensional semantics, and described some of its properties. Some of its theorems are commonly accepted, such as $\Box p \to p$, while others seem more controversial, e.g. $Ap \to \Box Ap$. In this chapter and the next, I will try to draw some general conclusions about the logical commitments of two-dimensional semantics as encoded in **2D**. In particular, I will try to argue that all theorems of **2D** can be shown to be plausible independently of two-dimensional semantics.

Since **2D** is not defined as the QNML axiomatized by some set of axioms, this cannot be done by just going through its axioms and arguing for them. Instead, I will use the following strategy: first, I will describe a logic of the three modalities that should be relatively uncontroversial, in the sense that all of its theorems should be commonly accepted. I will call this the "minimal logic". Then, I will prove that adding one rule and a few axioms to this logic produces the logic of two-dimensional semantics **2D**. Finally, I will argue that this rule and all of these axioms except one should also be commonly accepted. This leaves only the remaining axiom as a logical commitment of two-dimensional semantics that has not yet been shown to be independently plausible. It turns out that this is $N1$, the formula representing the first premise of the nesting argument discussed in section 1.4.2. In chapter 6, I will argue independently of two-dimensional semantics that accepting this formula is the correct answer to the nesting problem. Together with the arguments in this chapter, this will show that all theorems of **2D** are plausible independently of two-dimensional semantics.

I will start by defining the minimal logic **Min**. This will be done by taking the quasi-normal join of a logic for necessity and actuality with a logic of apriority. As such, it will be specified in a way that is very different from the way **2D** was specified in chapter 3, which will make it difficult to compare the two logics. To allow this comparison, I will then specify alternative proof systems for both **2D** and **Min** which have a common structure. With this, the axioms and the rule that have to be added to **Min** to produce **2D** can be specified easily, and I will then go on to argue for their plausibility.

## 5.1 A Minimal Logic

The first step, which will be taken in this section, is to define a logic of the three modalities that is acceptable to most philosophers, in the sense that it contains only theorems that are plausible independently of two-dimensional semantics. As there has not been much discussion on the correct logic of necessity, actuality and apriority in philosophy, I will define the logic by combining the commonly accepted logic for necessity and actuality with a plausible logic for apriority in a minimal way. To indicate which operators are being used, I index the relevant notation with the operators, as explained in chapter 2.

### 5.1.1 Necessity and Actuality

The commonly accepted logic of necessity and actuality is described in Crossley and Humberstone (1977). This is the logic given by frames without an accessibility relation and a distinguished actual world (which was mentioned in section 4.1.2), using real-world validity. It is sometimes called **S5A** (see Crossley and Humberstone (1977) and Hazen (1978)), but I will not follow this usage, since I use $A$ for apriority, and **S5A** would be too easily confused with $\mathbf{S5}_A$. Instead, I will use **Act**. As it is done in Crossley and Humberstone (1977), I first define a logic **Actg** for general consequence and then derive the intended logic **Act** for real-world consequence.

**Definition 5.1.** *Let* $\mathbf{Actg} = \oplus_{\Box @}\{T_\Box, 5_\Box, D_@, D_{c@}, I1, I2\}$. *From this, define* **Act** *by* $\vdash_{\mathbf{Act}} \varphi$ *iff* $\vdash_{\mathbf{Actg}} @\varphi$.

As with **2D**, we can also describe **Act** as a quasi-normal join:

**Proposition 5.2.** $\mathbf{Act} = \mathbf{Actg} +_{\Box @} \{T_@\}$.

*Proof.* As noted in Davies and Humberstone (1980, footnote 2), $\vdash_{\mathbf{Actg}} @(@p \to p)$, so the claim follows by Lemma 3.21. $\Box$

Since it will be more convenient later, I will use a different semantics for this than the one described above. The difference is similar to the difference between using frames with a universal relation and frames with an equivalence relation for **S5**.

**Definition 5.3.** $\mathsf{A}$ *is the class of frames* $\mathfrak{F} = \langle W, R_\Box, R_@\rangle$ *such that*

- $R_\Box$ *is an equivalence relation and*
- $R_@$ *is a function that maps any two* $R_\Box$*-related worlds to the same world, which is* $R_\Box$*-related to both of them.*

$\mathsf{AD}$ *is the class of* FWDE*s* $\mathcal{F} = \langle \mathfrak{F}, D\rangle$ *such that* $\mathfrak{F} = \langle W, R_\Box, R_@\rangle \in \mathsf{A}$ *and* $D = \mathrm{im}(R_@)$.

**Theorem 5.4.** **Actg** *is strongly frame-complete and defines* $\mathsf{A}$. **Act** *is sound and strongly complete with respect to* $\mathsf{AD}$.

*Proof.* As noted in chapter 3, all the axioms used in the definition of **Actg** are Sahlqvist formulas, so it follows from Theorem 2.6 that **Actg** is strongly frame-complete. The first-order conditions expressed by the axioms where also noted in chapter 3, and it is not hard to see that they describe $\mathsf{A}$. (This is essentially a special case of the proof in Blackburn and Marx (2002) of the completeness results of Gregory (2001).) By Lemma 3.17, it follows that **Act** is sound and strongly complete with respect to $\mathsf{AD}$. ☐

It is now time to catch up on a result I promised in section 3.5.3, namely that the set of **2D**-theorems that only contain the modalities $\Box$ and @ is **Act**:

**Theorem 5.5.** **2D** *is a conservative extension of* **Act**.

*Proof.* Using Proposition 2.5 and Theorem 3.20, it suffices to show that $L(\mathsf{AD}) = L(\mathsf{RD}|\Box@)$. This follows immediately from the fact that $\mathsf{AD} = \mathsf{RD}|\Box@$. ☐

It should be noted that although **Act** is commonly accepted, it is not universally agreed upon. E.g., it builds on the logic $\mathbf{S5}_\Box$ for necessity, which is sometimes called into question (e.g., in Salmon (1989) or the references in Gregory (2011, pp. 1–2)). I will not consider such objections here. Also, as noted in section 1.2.3, there is some debate about which definition of consequence is correct. As before, I will assume that the correct definition is real-world consequence, for which I argue (independently of two-dimensional semantics) in the appendix.

### 5.1.2 Apriority

What is the correct logic of apriority? In contrast to the logic of necessity and actuality, there is no commonly accepted one, so I will present some considerations that the correct logic is $\mathbf{S5}_A$.

$\mathbf{S5}_A$ is the logic $\oplus_A\{T_A, 5_A\}$. It is well-known that this is strongly frame-complete and defines the class of frames whose relation is an equivalence relation, which would also be easy to establish using the Sahlqvist completeness theorem for NMLs. More formally:

**Definition 5.6.** *Let* $\mathsf{E}_A$ *be the class of frames* $\mathfrak{F} = \langle W, R_A\rangle$ *such that* $R_A$ *is an equivalence relation on* $W$.

**Proposition 5.7.** $\mathbf{S5}_A$ *is strongly frame-complete and defines* $\mathsf{E}_A$.

To argue convincingly that $\mathbf{S5}_A$ is the correct logic of apriority, it is crucial to note that apriority is a theoretical notion that admits of a number of different explications. E.g., one might call a truth a priori if it is metaphysically possible that there is some agent who has knowledge of that fact without empirical justification. As Chalmers (2004, section 3.9, especially footnote 15) notes, this is not the notion that two-dimensional semantics is supposed to capture. I will therefore only consider the notion of apriority that is relevant for two-dimensional semantics, and argue that its logic is $\mathbf{S5}_A$.

This observation also shows that by doing so, I do not disagree with Anderson (1993), who proposes a very different logic for apriority. The reason for this is that Anderson's logic is intended to model the explication of apriority via metaphysically possible knowledge, and not the one given by two-dimensionalists. So since Anderson and I formalize different notions, the fact that we propose different logics for apriority does not mean that we disagree in any way.

Ignoring some details about the bearers of apriority, two-dimensional semantics uses something like the following account: a truth is a priori "when it can be conclusively non-experientially justified on ideal rational reflection" (Chalmers (2004, p. 208)). I will now argue that $\mathbf{S5}_A$ is the correct logic of this notion by going through the proof system $N_A\{T_A, 5_A\}$ which generates $\mathbf{S5}_A$.

I will take the correctness of classical propositional logic and the rule of modus ponens for granted. In contrast to some logics discussed in section 4.1, the logic of apriority used here should capture all sentential expressions of natural language, therefore proposition letters may stand for any such expression. This justifies the rule of uniform substitution. I will now consider the remaining three axioms and the remaining rule of the proof system:

| | |
|---|---|
| $K_A$ | $A(p \to q) \to (Ap \to Aq)$ |
| $T_A$ | $Ap \to p$ |
| $5_A$ | $Cp \to ACp$ |
| $Gen_A$ | From $\varphi$, derive $A\varphi$ |

$K_A$ is propositionally equivalent to $(A(p \to q) \wedge Ap) \to Aq$. We can therefore understand it as saying that a priori truths are closed under modus ponens. This should be the case: assume that $p \to q$ and $p$ can be conclusively non-experientially justified on ideal rational reflection. Surely modus ponens is available in ideal rational reflection, so $q$ should also be conclusively non-experientially justifiable on ideal rational reflection, and therefore a priori.

$T_A$ says that if $p$ is a priori, then $p$. This should be uncontroversial – in fact, above, I already assumed this, as I only explained what it is to be a priori for *truths*.

Given the other axioms and rules, $5_A$ can equivalently be written as $\neg Ap \to A\neg Ap$. In this form, $5_A$ says that if $p$ is not a priori, then it is a priori that $p$ is not a priori. To borrow terminology from epistemic logic, this means that apriority obeys *negative introspection*. It is not completely clear whether this holds on the given account of apriority. But it is at least plausible that on a truly ideal conception of ideal rational reflection, the reach of ideal rational reflection is itself accessible by ideal rational reflection, and the principle therefore holds.

Finally, the rule of necessitation for $A$ has to be justified. Assume that $\varphi$ represents a logical truth. A logical truth should be conclusively non-experientially justifiable on ideal rational reflection, so it should be a priori. This behaviour of apriority is something that we want to capture with our logic, so it should be a logical truth that it is a priori. Hence for any theorem $\varphi$, $A\varphi$ should be a theorem, which is the rule of generalization.

Therefore, the logic of apriority should include the theorems of $\mathbf{S5}_A$. In fact, we could also argue that the logic should not contain any additional theorems, using a result from Scroggs (1951) (see also Segerberg (1971, pp. 122–128) and Gärdenfors (1973)). However, this will not be needed in the following.

There is one potential worry I want to address before moving on. If apriority is understood as a sentential predicate rather than as a sentential operator, then the arguments in Montague (1963) show that apriority cannot obey $\mathbf{S5}_A$, on account of Gödelian problems with self-reference. (See also Humberstone (2004, p. 28) for a related skeptical attitude towards using $\mathbf{S5}_A$ as a logic of apriority.) But the following considerations indicate that apriority should not be understood in this way: it seems that "It is a priori that three is prime" and the German translation "Es ist a priori dass drei prim ist" say the same thing, but if the sentences would be sloppy ways of stating "'Three is prime' is a priori" and "'Drei ist prim' ist a priori", they would make statements about different sentences, and therefore would say different things. Therefore, apriority should not be understood as a sentential predicate, but rather as a sentential operator. It might be the case that the problems with self-reference persist even if apriority is understood as a sentential operator, but this question is beyond the scope of this thesis. If such worries turn out to be substantial, we can still understand $\mathbf{S5}_A$ as the correct logic of apriority of some sufficiently simple fragment of English. The present argument would then still show that **2D** is the correct logic of the modalities in the simple fragment. Note especially that all of the examples of contingent a priori or necessary a posteriori truths discussed here are clearly simple in the sense required.

### 5.1.3 Min

There is very little discussion on the systematic interaction between actuality and apriority, as well as between necessity and apriority, so there is certainly no consensus view that could be adopted. In contrast to the logic of apriority, the interaction between these modalities poses some difficult questions, and it is not likely that an uncontroversial logic for them can be devised. Therefore, I will only assume that the correct logic of necessity and actuality is **Act**, and that the correct logic of apriority is $\mathbf{S5}_A$, and use these to define the minimal logic.

Just assuming that **Act** and $\mathbf{S5}_A$ hold produces a very weak and implausible logic for the three modalities – e.g., it will not even contain all propositional tautologies of $\mathcal{L}_{\Box @ A}$. To allow for the use of a relational semantics, I will assume that the combined logic is a QNML. Since the logic **K** and the rules of uniform substitution and modus ponens have not caused any trouble before, this should not constitute a strong assumption. Therefore, it is natural to suggest the smallest QNML containing both **Act** and $\mathbf{S5}_A$ as a minimal logic of necessity, actuality, and apriority that should be acceptable to most philosophers. That is, we can define:

**Definition 5.8.** $\mathbf{Min} = \mathbf{Act} +_{\Box @ A} \mathbf{S5}_A$.

By saying that this logic should be a relatively uncontroversial minimal logic, I only claim that all of its theorems should be accepted as expressing logical truths by most philosophers. There may be commonly accepted logical truths it does not classify as theorems.

**Min** would not be a good minimal logic if it wasn't contained in **2D**, but this is easily shown to be the case:

**Lemma 5.9.** $\mathbf{Min} \subseteq \mathbf{2D}$.

*Proof.* By Theorems 3.27 and 5.5, **2D** contains both $\mathbf{S5}_A$ and **Act**. Since it is a QNML, it contains $\mathbf{K}_{\Box @ A}$, and is closed under *MP* and *US*. With these facts, we can show the claim by induction on the proof system $QN_{\Box @ A}(\mathbf{Act} \cup \mathbf{S5}_A)$ for **Min**. □

### 5.1.4 A Semantics for Min

For technical reasons, it will later be useful to have a semantics with respect to which **Min** is sound and complete. Since **Min** is constructed from logics axiomatized by Sahlqvist formulas, it can comfortably be shown to be FWDE-complete, which immediately gives us the class of FWDEs $\mathsf{FrD}(\mathbf{Min})$ as a semantics with respect to which **Min** is sound and strongly complete.

**Lemma 5.10.** $\mathbf{Min} = +(\{T_@\}, \oplus_{\Box @}\{T_\Box, 5_\Box, D_@, D_{c@}, I1, I2\}, \oplus_A\{T_A, 5_A\})$.

*Proof.* By Propositions 2.2 and 5.2. ☐

**Proposition 5.11.** **Min** *is strongly* FWDE*-complete.*

*Proof.* By Lemma 5.10 and Theorem 2.7. ☐

**Proposition 5.12.** $\mathsf{FrD}(\mathbf{Min})$ *is the class of* FWDE*s* $\mathcal{F} = \langle \mathfrak{F}, D\rangle$ *such that*

- $wR_@w$ *for all* $w \in D$,
- $(\mathfrak{F}|\Box @)_D \in \mathsf{A}$*, and*
- $(\mathfrak{F}|A)_D \in \mathsf{E}_A$.

*Proof.* By Proposition 2.4, $\mathsf{FrD}(\mathbf{Min})$ is the class of FWDEs $\mathcal{F} = \langle \mathfrak{F}, D\rangle$ such that $\mathcal{F} \Vdash T_@$, $(\mathfrak{F}|\Box @)_D \Vdash \{T_\Box, 5_\Box, D_@, D_{c@}, I1, I2\}$, and $(\mathfrak{F}|A)_D \Vdash \{T_A, 5_A\}$. The claim follows by Theorem 5.4 and Proposition 5.7. ☐

## 5.2 Alternative Proof Systems

I have specified two logics for necessity, actuality and apriority: **2D** and **Min**. We can assume that **Min** is widely accepted, so the question now is: what has to be added to **Min** to produce **2D**? Although I have given syntactic characterizations of both logics, the immediate obstacle to answering this question is that the two logics are not just specified as the QNMLs axiomatized by some set of axioms, but constructed in a more complicated manner as joins. To remove this obstacle, I specify alternative proof systems for **2D** and **Min** in this section, which are easier to compare.

The alternative proof systems are constructed so as to only contain a set of axioms and rules that apply to all theorems, which contrasts with the multi-step constructions of **2D** and **Min** given above. The key idea of these systems is to restrict the rules of generalization. Let me define $RGen_\Box$ to be the following rule:

- $RGen_\Box$: If $\varphi$ is a theorem and in $\mathcal{L}_\Box$, then $\Box\varphi$ is a theorem.

$RGen_A$ is the obvious analog to $RGen_\Box$. I will first provide an alternative proof system for **Act**; this will be useful for constructing similar proof systems for **2D** and **Min**.

### 5.2.1 An Alternative Proof System for Act

I start by formally specifying the alternative proof system for **Act**.

**Definition 5.13.** $P_{\mathbf{Act}}$ *is the proof system containing the rules MP, US and* $RGen_\Box$*, and as axioms the propositional tautologies and the following:*

| | |
|---|---|
| $K_\Box$ | $\Box(p \to q) \to (\Box p \to \Box q)$ |
| $T_\Box$ | $\Box p \to p$ |
| $5_\Box$ | $\Diamond p \to \Box\Diamond p$ |
| $\Box K_@$ | $\Box(@(p \to q) \to (@p \to @q))$ |
| $\Box D_@$ | $\Box(@p \to \neg @\neg p)$ |
| $\Box D_{c@}$ | $\Box(\neg @\neg p \to @p)$ |
| $\Box I1$ | $\Box(\Box p \to @p)$ |
| $\Box I2$ | $\Box(@p \to \Box @p)$ |
| $T_@$ | $@p \to p$ |

A historical note: a similar proof system was used in Prior (1968a, p. 113) as a proposed axiomatization of a temporal logic with the indexical operator "now" (the paper does not contain a completeness proof). A simplification of the proof system is described in Prior (1968b), which might also be applicable to $P_{\mathbf{Act}}$. However, for present purposes, such a simplification would make the proof system less useful.

To prove that $P_{\mathbf{Act}}$ in fact produces **Act**, a number of lemmas are needed. I will first prove that $\mathbf{Act} \subseteq L(P_{\mathbf{Act}})$ (Lemma 5.16) and then that $L(P_{\mathbf{Act}}) \subseteq \mathbf{Act}$ (Lemma 5.18), from which the claim follows. For each of these two results, additional lemmas are needed.

**Lemma 5.14.** $\mathbf{S5}_\Box \subseteq L(P_{\mathbf{Act}})$.

*Proof.* Immediate by induction on $\mathbf{S5}_\Box$'s normal proof system $N_\Box\{T_\Box, 5_\Box\}$. □

**Lemma 5.15.** *If* $\vdash_{\mathbf{Actg}} \varphi$ *then* $\vdash_{P_{\mathbf{Act}}} \Box\varphi$.

*Proof.* By induction on **Actg**'s proof system $N_{\Box @}\{T_\Box, 5_\Box, D_@, D_{c@}, I1, I2\}$. If $\varphi$ is a propositional tautology or $K_\Box$, then $\vdash_{P_{\mathbf{Act}}} \varphi$, and by $RGen_\Box$, $\vdash_{P_{\mathbf{Act}}} \Box\varphi$. For $K_@$, note that $\Box K_@$ is an axiom of $P_{\mathbf{Act}}$. Like with $K_\Box$, it follows by $RGen_\Box$ that $\vdash_{P_{\mathbf{Act}}} \Box T_\Box$ and $\vdash_{P_{\mathbf{Act}}} \Box 5_\Box$. Also, like $K_@$, the cases of $D_@$, $D_{c@}$, $I1$, and $I2$ are immediate.

*MP*: Assume that $\psi$ is proven by *MP* from $\varphi$ and $\varphi \to \psi$. Then by induction hypothesis, $\vdash_{P_{\mathbf{Act}}} \Box\varphi$ and $\vdash_{P_{\mathbf{Act}}} \Box(\varphi \to \psi)$. By $K_\Box$, $\vdash_{P_{\mathbf{Act}}} \Box\varphi \to \Box\psi$, so $\vdash_{P_{\mathbf{Act}}} \Box\psi$.

*US*: Assume that $\varphi[\psi/p]$ is proven by *US* from $\varphi$. (Here and in the following, $\varphi[\psi/p]$ is the result of uniformly substituting $\psi$ for $p$ in $\varphi$.) By induction hypothesis, $\vdash_{P_{\mathbf{Act}}} \Box\varphi$, and so by uniform substitution, $\vdash_{P_{\mathbf{Act}}} (\Box\varphi)[\psi/p]$. Therefore $\vdash_{P_{\mathbf{Act}}} \Box(\varphi[\psi/p])$, as needed.

$Gen_@$: Assume that $@\varphi$ is proven by $Gen_@$ from $\varphi$. Then by induction hypothesis, $\vdash_{P_{\mathbf{Act}}} \Box\varphi$. So by $T_\Box$ and $\Box I1$: $\vdash_{P_{\mathbf{Act}}} @\varphi$. Finally, by $T_\Box$ and $\Box I2$: $\vdash_{P_{\mathbf{Act}}} \Box @\varphi$.

$Gen_\Box$: Assume that $\Box\varphi$ is proven by $Gen_\Box$ from $\varphi$. Then by induction hypothesis, $\vdash_{P_{\mathbf{Act}}} \Box\varphi$. By Lemma 5.14, $\vdash_{P_{\mathbf{Act}}} 4_\Box$, so $\vdash_{P_{\mathbf{Act}}} \Box\Box\varphi$. □

**Lemma 5.16.** $\mathbf{Act} \subseteq L(P_{\mathbf{Act}})$.

*Proof.* Consider any $\varphi \in \mathbf{Act}$. By definition, $\vdash_{\mathbf{Actg}} @\varphi$, so by Lemma 5.15, $\vdash_{P_{\mathbf{Act}}} \Box @\varphi$. Therefore, by $T_\Box$, $\vdash_{P_{\mathbf{Act}}} @\varphi$, and then by $T_@$, $\vdash_{P_{\mathbf{Act}}} \varphi$. □

**Lemma 5.17.** **Act** *is a conservative extension of* $\mathbf{S5}_\Box$.

*Proof.* This follows from the fact that **2D** is a conservative extension of $\mathbf{S5}_\Box$ as well as **Act** (Theorems 3.26 and 5.5). □

**Lemma 5.18.** $L(P_{\mathbf{Act}}) \subseteq \mathbf{Act}$.

*Proof.* By induction on $P_{\mathbf{Act}}$. If $\varphi$ is an axiom of $P_{\mathbf{Act}}$, it is easy to establish that $\vdash_{\mathbf{Act}} \varphi$ – this is either immediate by Proposition 5.2, or can be shown by an application of $Gen_\Box$ to an axiom of **Actg**. The rules of modus ponens and uniform substitution follow straightforward by induction, since **Act** is closed under them. Since $\mathbf{S5}_\Box$ is normal, it follows by Lemma 5.17 that **Act** is closed under $RGen_\Box$. □

**Theorem 5.19.** $L(P_{\mathbf{Act}}) = \mathbf{Act}$.

*Proof.* From Lemmas 5.16 and 5.18. □

### 5.2.2 An Alternative Proof System for 2D

By adding to $P_{\mathbf{Act}}$, an alternative proof system for **2D** can be obtained:

**Definition 5.20.** $P_{\mathbf{2D}}$ *is the proof system obtained by adding the rule* $Gen_A$ *and the following axioms to* $P_{\mathbf{Act}}$*:*

| | |
|---|---|
| $\Box K_A$ | $\Box(A(p \to q) \to (Ap \to Aq))$ |
| $\Box 4_A$ | $\Box(Ap \to AAp)$ |
| $\Box 5_A$ | $\Box(Cp \to ACp)$ |
| $\Box D_A$ | $\Box(Ap \to Cp)$ |
| $T_A$ | $Ap \to p$ |
| $N1$ | $Ap \to \Box Ap$ |

It should be noted that these axioms were chosen with the application of comparing it to **Min** in mind. Also, the definition is slightly imprecise: $P_{\mathbf{2D}}$ is a proof system in $\mathcal{L}_{\Box @ A}$, whereas $P_{\mathbf{Act}}$ is a proof system in $\mathcal{L}_{\Box @}$. E.g. the rule of *US* in $P_{\mathbf{Act}}$ applies only to formulas in $\mathcal{L}_{\Box @}$, whereas in $P_{\mathbf{2D}}$ it applies to all formulas in $\mathcal{L}_{\Box @ A}$. I assume that these issues are clear enough from the context. As before, to prove that the alternative proof system produces the right logic, a number of lemmas are needed. I will first show that $\mathbf{2D} \subseteq L(P_{\mathbf{2D}})$ (Lemma 5.26), and then that $L(P_{\mathbf{2D}}) \subseteq \mathbf{2D}$ (Lemma 5.27). To prove that $L(P_{\mathbf{2D}})$ contains **2D**, it is useful to start by proving that it contains $\mathbf{K}_{\Box @ A}$ (Lemma 5.22), which will simplify some necessary deductions.

**Lemma 5.21.** *If* $\vdash_{\mathbf{K}_{\Box @ A}} \varphi$ *then* $\vdash_{P_{\mathbf{2D}}} \Box\varphi$.

*Proof.* By induction on the proof system $N_{\Box @ A}\emptyset$ for $\mathbf{K}_{\Box @ A}$, analogously to the proof of of Lemma 5.15. □

**Lemma 5.22.** $\mathbf{K}_{\Box @ A} \subseteq L(P_{\mathbf{2D}})$.

*Proof.* If $\vdash_{\mathbf{K}_{\Box @ A}} \varphi$, then by Lemma 5.21, $\vdash_{P_{\mathbf{2D}}} \Box\varphi$, so by $T_\Box$, $\vdash_{P_{\mathbf{2D}}} \varphi$. □

**Lemma 5.23.** $\vdash_{P_{\mathbf{2D}}} Cp \to \Box Cp$.

*Proof.* By the following derivation:

| | | |
|---|---|---|
| (1) | $\Box(Ap \to Cp)$ | $\Box D_A$ |
| (2) | $\Box(ACp \to CCp)$ | (1), $US$ |
| (3) | $\Box(Ap \to AAp)$ | $\Box 4_A$ |
| (4) | $\Box(Ap \to AAp) \to \Box(CC\neg p \to C\neg p)$ | $\mathbf{K}_{\Box@A}$ |
| (5) | $\Box(CCp \to Cp)$ | (3), (4) |
| (6) | $\Box(ACp \to Cp)$ | (2), (5), $\mathbf{K}_{\Box@A}$ |
| (7) | $\Box ACp \to \Box Cp$ | (6), $\mathbf{K}_{\Box@A}$ |
| (8) | $ACp \to \Box ACp$ | $N1$, $US$ |
| (9) | $Cp \to ACp$ | $\Box 5_A$, $T_\Box$ |
| (10) | $Cp \to \Box Cp$ | (9), (8), (7) |

□

**Lemma 5.24.** $\vdash_{P_{\mathbf{2D}}} \Box(Ap \to @p)$.

*Proof.* By the following derivation:

| | | |
|---|---|---|
| (1) | $Ap \to p$ | $T_A$ |
| (2) | $p \to Cp$ | (1) |
| (3) | $Cp \to \Box Cp$ | Lemma 5.23 |
| (4) | $p \to \Box Cp$ | (2), (3) |
| (5) | $\Diamond Ap \to p$ | (4) |
| (6) | $p \to \neg @\neg p$ | $T_@$ |
| (7) | $\neg @\neg p \to @p$ | $\Box D_{c@}$, $T_\Box$ |
| (8) | $p \to @p$ | (6), (7) |
| (9) | $@p \to \Box @p$ | $\Box I2$, $T_\Box$ |
| (10) | $p \to \Box @p$ | (8), (9) |
| (11) | $\Diamond Ap \to \Box @p$ | (5), (10) |
| (12) | $(\Diamond Ap \to \Box @p) \to \Box(Ap \to @p)$ | $\mathbf{K}_{\Box@A}$ |
| (13) | $\Box(Ap \to @p)$ | (11), (12) |

□

**Lemma 5.25.** *If* $\vdash_{\mathbf{2Dg}} \varphi$ *then* $\vdash_{P_{\mathbf{2D}}} \Box\varphi$.

*Proof.* By induction on **2Dg**'s normal proof system. If $\varphi$ is a propositional tautology, $K_\Box$, $T_\Box$, or $5_\Box$, then $\vdash_{P_{\mathbf{2D}}} \varphi$. Since any such $\varphi$ is in $\mathcal{L}_\Box$, it follows by $RGen_\Box$ that $\vdash_{P_{\mathbf{2D}}} \Box\varphi$. The cases of $K_@$, $K_A$, $D_@$, $D_{c@}$, $I1$, $I2$, $4_A$, and $5_A$ are immediate. The case of $I3$ is shown in Lemma 5.24. For $I4$, note that by $T_@$ and $Gen_A$, $\vdash_{P_{\mathbf{2D}}} A(@p \to p)$, and so with $N1$, $\vdash_{P_{\mathbf{2D}}} \Box A(@p \to p)$.

The rules $MP$, $US$, $Gen_@$, and $Gen_\Box$ can be dealt with as in the proof of Lemma 5.15. For $Gen_A$, assume that $A\varphi$ is proven from $\varphi$. Then by induction hypothesis, $\vdash_{P_{\mathbf{2D}}} \Box\varphi$. So by $T_\Box$, $\vdash_{P_{\mathbf{2D}}} \varphi$, and by $Gen_A$, $\vdash_{P_{\mathbf{2D}}} A\varphi$. By $N1$, $\vdash_{P_{\mathbf{2D}}} \Box A\varphi$. □

**Lemma 5.26.** $\mathbf{2D} \subseteq L(P_{\mathbf{2D}})$.

*Proof.* Consider any $\varphi \in \mathbf{2D}$. By definition, $\vdash_{\mathbf{2Dg}} @\varphi$, so by Lemma 5.25, $\vdash_{P_{\mathbf{2D}}} \Box @\varphi$. Therefore, by $T_\Box$, $\vdash_{P_{\mathbf{2D}}} @\varphi$, and then by $T_@$, $\vdash_{P_{\mathbf{2D}}} \varphi$. □

**Lemma 5.27.** $L(P_{\mathbf{2D}}) \subseteq \mathbf{2D}$.

*Proof.* By induction on $P_{\mathbf{2D}}$. If $\varphi$ is an axiom of $P_{\mathbf{Act}}$, $\Box K_A$, $\Box 4_A$, or $\Box 5_A$, it is easy to establish that $\vdash_{\mathbf{2D}} \varphi$ – this is either immediate, or can be shown by an application of Gen$_\Box$ to an axiom of **2Dg**. $\vdash_{\mathbf{2D}} T_A$ since **2D** is an extension of $\mathbf{S5}_A$ (Theorem 3.27), and $\vdash_{\mathbf{2D}} N1$ was shown in Lemma 3.10. The only remaining axiom is $\Box D_A$, which can easily be seen to be a **2D**-theorem by the completeness of **2D** with respect to $\mathsf{MD}$ (Theorem 3.18).

The rules of modus ponens and uniform substitution follow straightforwardly by induction, since **2D** is closed under them. By Theorem 3.26, **2D** is a conservative extension of $\mathbf{S5}_\Box$, so **2D** is closed under $RGen_\Box$. It was proven in Proposition 3.25 that **2D** is closed under $Gen_A$. $\Box$

**Theorem 5.28.** $L(P_{\mathbf{2D}}) = \mathbf{2D}$.

*Proof.* From Lemmas 5.26 and 5.27. $\Box$

### 5.2.3 An Alternative Proof System for Min

Finally, I also specify an alternative proof system for **Min**. Here, $\{\Box, @, A\}^*$ is the set of finite sequences on $\{\Box, @, A\}$.

**Definition 5.29.** $P_{\mathbf{Min}}$ *is the proof system containing the rules MP, US, $RGen_\Box$, and $RGen_A$, and as axioms the following:*

| | |
|---|---|
| $*Prop$ | $\heartsuit\varphi$ *for all propositional tautologies* $\varphi$ *and* $\heartsuit \in \{\Box, @, A\}^*$ |
| $*K_?$ | $\heartsuit(\nabla(p \to q) \to (\nabla p \to \nabla q))$ *for all* $\heartsuit \in \{\Box, @, A\}^*$, $\nabla \in \{\Box, @, A\}$ |
| $T_\Box$ | $\Box p \to p$ |
| $5_\Box$ | $\Diamond p \to \Box\Diamond p$ |
| $\Box D_@$ | $\Box(@p \to \neg@\neg p)$ |
| $\Box D_{c@}$ | $\Box(\neg@\neg p \to @p)$ |
| $\Box I1$ | $\Box(\Box p \to @p)$ |
| $\Box I2$ | $\Box(@p \to \Box@p)$ |
| $T_@$ | $@p \to p$ |
| $T_A$ | $Ap \to p$ |
| $5_A$ | $Cp \to ACp$ |

Again, a number of lemmas are required to prove that the logic of $P_{\mathbf{Min}}$ is in fact **Min**. I will first prove that $\mathbf{Min} \subseteq L(P_{\mathbf{Min}})$ (Lemma 5.35), by showing that $L(P_{\mathbf{Min}})$ contains $\mathbf{K}_{\Box@A}$, $\mathbf{S5}_A$ and **Act**. Then I will show that $L(P_{\mathbf{Min}}) \subseteq \mathbf{Min}$ (Lemma 5.37).

**Lemma 5.30.** *For any formulas $\varphi, \psi$ and $\heartsuit \in \{\Box, @, A\}^*$, if $\vdash_{P_{\mathbf{Min}}} \heartsuit(\varphi \to \psi)$ then $\vdash_{P_{\mathbf{Min}}} \heartsuit\varphi \to \heartsuit\psi$.*

*Proof.* By induction on the length of $\heartsuit$. If $\heartsuit$ is of length 0, the claim is immediate. Assume it holds for length $n$, and consider any formulas $\varphi, \psi$ and $\spadesuit \in \{\Box, @, A\}^{n+1}$. Then there are $\heartsuit \in \{\Box, @, A\}^n$ and $\nabla \in \{\Box, @, A\}$ such that $\spadesuit = \heartsuit\nabla$. Assume that $\vdash_{P_{\mathbf{Min}}} \heartsuit\nabla(\varphi \to \psi)$. By $*K_?$ and *US*, $\vdash_{P_{\mathbf{Min}}} \heartsuit(\nabla(\varphi \to \psi) \to (\nabla\varphi \to \nabla\psi))$. So by induction hypothesis, $\vdash_{P_{\mathbf{Min}}} \heartsuit\nabla(\varphi \to \psi) \to \heartsuit(\nabla\varphi \to \nabla\psi)$. Using *MP* and the assumption, it follows that $\vdash_{P_{\mathbf{Min}}} \heartsuit(\nabla\varphi \to \nabla\psi)$. By induction hypothesis, we obtain $\vdash_{P_{\mathbf{Min}}} \heartsuit\nabla\varphi \to \heartsuit\nabla\psi$. $\Box$

**Lemma 5.31.** $\mathbf{K}_{\Box @ A} \subseteq \ker(L(P_{\mathbf{Min}}))$.

*Proof.* By induction on the normal proof system $N_{\Box @ A}\emptyset$ for $\mathbf{K}_{\Box @ A}$. If $\varphi$ is a propositional tautology, $K_\Box$, $K_@$, or $K_A$, then $\heartsuit\varphi$ is an axiom of $P_{\mathbf{Min}}$ for all $\heartsuit \in \{\Box, @, A\}^*$, so it follows directly that $\varphi \in \ker(L(P_{\mathbf{Min}}))$.

If $\varphi$ is obtained by *MP* from $\varphi$ and $\varphi \to \psi$, it follows by induction hypothesis that $\varphi, \varphi \to \psi \in \ker(L(P_{\mathbf{Min}}))$. Consider any $\heartsuit \in \{\Box, @, A\}^*$. Then $\heartsuit\varphi, \heartsuit(\varphi \to \psi) \in L(P_{\mathbf{Min}})$. By Lemma 5.30, $\vdash_{P_{\mathbf{Min}}} \heartsuit\varphi \to \heartsuit\psi$. So by *MP*, $\vdash_{P_{\mathbf{Min}}} \heartsuit\psi$. So $\psi \in \ker(L(P_{\mathbf{Min}}))$.

For uniform substitution, assume that $\varphi[\psi/p]$ is derived from $\varphi$ by *US*. By induction hypothesis, $\varphi \in \ker(L(P_{\mathbf{Min}}))$. Let $\heartsuit \in \{\Box, @, A\}^*$. Then $\vdash_{P_{\mathbf{Min}}} \heartsuit\varphi$. By *US*, $\vdash_{P_{\mathbf{Min}}} (\heartsuit\varphi)[\psi/p]$, so $\vdash_{P_{\mathbf{Min}}} \heartsuit(\varphi[\psi/p])$. Therefore $\varphi[\psi/p] \in \ker(L(P_{\mathbf{Min}}))$.

Finally, let $\nabla \in \{\Box, @, A\}$ and $\nabla\varphi$ be obtained from $\varphi$ by *Gen*. By induction hypothesis, $\varphi \in \ker(L(P_{\mathbf{Min}}))$. Let $\heartsuit \in \{\Box, @, A\}^*$. Then $\vdash_{P_{\mathbf{Min}}} \heartsuit\nabla\varphi$. So $\nabla\varphi \in \ker(L(P_{\mathbf{Min}}))$. □

**Lemma 5.32.** $\mathbf{K}_{\Box @ A} \subseteq L(P_{\mathbf{Min}})$.

*Proof.* Immediate from Lemma 5.31, since in general $\ker(X) \subseteq X$. □

**Lemma 5.33.** $\mathbf{S5}_A \subseteq L(P_{\mathbf{Min}})$.

*Proof.* Immediate by induction on $\mathbf{S5}_A$'s normal proof system $N_A\{T_A, 5_A\}$. □

**Lemma 5.34.** $\mathbf{Act} \subseteq L(P_{\mathbf{Min}})$.

*Proof.* Immediate, since $\mathbf{Act} = L(P_{\mathbf{Act}})$ (Theorem 5.19) and all axioms and rules of $P_{\mathbf{Act}}$ are in $P_{\mathbf{Min}}$. □

**Lemma 5.35.** $\mathbf{Min} \subseteq L(P_{\mathbf{Min}})$.

*Proof.* By definition, **Min** is the smallest set of formulas containing $\mathbf{K}_{\Box @ A}$, $\mathbf{S5}_A$, and **Act** that is closed under *MP* and *US*, so the claim follows from Lemmas 5.32, 5.33, and 5.34, and the fact that $P_{\mathbf{Min}}$ contains *MP* and *US*. □

**Lemma 5.36.** **Min** *is a conservative extension of both* $\mathbf{S5}_\Box$ *and* $\mathbf{S5}_A$.

*Proof.* Analogous to the proofs of Theorems 3.26 and 3.27, using Propositions 5.7, 5.11 and 5.12. □

**Lemma 5.37.** $L(P_{\mathbf{Min}}) \subseteq \mathbf{Min}$.

*Proof.* By induction on $P_{\mathbf{Min}}$. Note that all instances of the axiom schemas $*Prop$ and $*K_?$ are theorems of $\mathbf{K}_{\Box @ A}$. $T_\Box$, $5_\Box$, $\Box D_@$, $\Box D_{c@}$, $\Box I1$, $\Box I2$, and $T_@$ are theorems of **Act**, since they are all axioms of $P_{\mathbf{Act}}$. $T_A$ and $5_A$ are theorems of $\mathbf{S5}_A$. Therefore all axioms of $P_{\mathbf{Min}}$ are **Min**-theorems. Further, by definition, **Min** is closed under *MP* and *US*. Finally, Lemma 5.36 shows that **Min** is closed under $RGen_\Box$ and $RGen_A$. □

**Theorem 5.38.** $\mathbf{Min} = L(P_{\mathbf{Min}})$.

*Proof.* From Lemmas 5.35 and 5.37. □

## 5.3 Comparing Min and 2D

So far in this chapter, I have specified an uncontroversial logic **Min** of necessity, actuality and apriority, and provided alternative proof systems for both **2D** and **Min**. With these proof systems in place, the logics can now be compared. The first task is to specify a way of enriching **Min** to produce **2D**.

### 5.3.1 Commitments Beyond Min

One way of getting **2D** from **Min** is by adding the axioms listed below, and the rule $Gen_A$. Of course, there are other axioms that would to the same job.

**Definition 5.39.** *Let $P^+_{\mathbf{Min}}$ be the proof system containing the theorems of* **Min***, $\Box 4_A$, $\Box 5_A$, $\Box D_A$, and $N1$ as axioms, and MP, US, and $Gen_A$ as rules (see below for a list of the axioms).*

**Theorem 5.40.** $L(P^+_{\mathbf{Min}}) = \mathbf{2D}$.

*Proof.* $L(P^+_{\mathbf{Min}}) \subseteq \mathbf{2D}$: As shown in Lemma 5.9, **2D** contains all theorems of **Min**. Further, $P_{\mathbf{2D}}$ contains all additional axioms and rules of proof of $P^+_{\mathbf{Min}}$, so this part follows from the fact that $L(P_{\mathbf{2D}}) = \mathbf{2D}$ (Theorem 5.28).

$\mathbf{2D} \subseteq L(P^+_{\mathbf{Min}})$: Using Theorem 5.28, by induction on $P_{\mathbf{2D}}$. It is straightforward to verify that all axioms of $P_{\mathbf{2D}}$ are theorems of $P^+_{\mathbf{Min}}$. For the rules of $P_{\mathbf{2D}}$, note that $P^+_{\mathbf{Min}}$ also contains *MP*, *US*, and $Gen_A$. For $RGen_\Box$, assume that $\varphi \in \mathcal{L}_\Box$ such that $\vdash_{\mathbf{2D}} \varphi$. Then since **2D** is a conservative extension of $\mathbf{S5}_\Box$ (Theorem 3.26), $\varphi \in \mathbf{S5}_\Box$, so $\Box\varphi \in \mathbf{S5}_\Box$, and by Lemma 5.36, $\Box\varphi \in \mathbf{Min}$. □

As **Min** is already closed under *MP* and *US*, and these rules are philosophically unproblematic, the logical commitments of two-dimensional semantics for the propositional logic of necessity, actuality and apriority that go beyond **Min** are contained in the following:

| | |
|---|---|
| $\Box 4_A$ | $\Box(Ap \to AAp)$ |
| $\Box 5_A$ | $\Box(Cp \to ACp)$ |
| $\Box D_A$ | $\Box(Ap \to Cp)$ |
| $N1$ | $Ap \to \Box Ap$ |
| $Gen_A$ | From $\varphi$, derive $A\varphi$ |

### 5.3.2 Independently Plausible Commitments

I will now argue that the rule $Gen_A$ and the axioms $\Box 4_A$, $\Box 5_A$, and $\Box D_A$ are plausible independently of two-dimensional semantics. I don't want to claim that they are philosophically completely uncontroversial, but only that *prima facie*, we have good reasons to accept them, and that these reasons are not dependent on the acceptance of two-dimensional semantics. The plausibility of $Gen_A$ has already been motivated in section 5.1.2, so I will only consider the axioms.

$\Box 4_A$ and $\Box 5_A$ are necessitated versions of what can be described as the positive and negative introspection properties of $A$. I will first argue for the unnecessitated variants $4_A$ and $5_A$. I have argued in section 5.1.2 that $5_A$ is a plausible principle, and one can argue similarly for $4_A$: as I said there, on truly ideal rational reflection, the reach of ideal rational reflection is itself accessible by ideal rational reflection, so if it is a priori that $p$, this fact should be a priori as well, which is just what $4_A$ says. $\Box D_A$ is the necessitated version of the principle $Ap \rightarrow Cp$, which is propositionally equivalent to $\neg(Ap \wedge A\neg p)$. Surely, $p$ and $\neg p$ can not both be conclusively non-experientially justified on ideal rational reflection, so $D_A$ must hold as well. So $4_A$, $5_A$ and $D_A$ are plausible, but what about their necessitations?

The arguments given for the principles do not turn on any contingent feature of the world, so there is no reason to assume that they should not hold of necessity. So at least *prima facie*, the necessitated principles are plausible as well. But there is one potential argument that one could level against the necessitated principles: as we have seen above, it is also very plausible that $T_A$ holds, that is, that apriority implies truth. Again, *prima facie*, there is no reason to assume that this principle should not also hold of necessity; that is, that $\Box T_A = \Box(Ap \rightarrow p)$ should hold. But this is just the premise $N2$ of the nesting problem, and as we have seen in section 3.2, this is not a theorem of **2D**. The problem with $\Box T_A$ can be illustrated with the formula $T_@ = @p \rightarrow p$ where $p$ stands for a contingent truth. In this case, $T_@$ represents a truth which is contingent and a priori. As we will see in the next chapter, there are reasons for taking the following instance of $N2$ to be false in this case:

$N2(T_@)$ $\Box(A(@p \rightarrow p) \rightarrow (@p \rightarrow p))$

Therefore, we have to make sure that examples like this do not cast doubt on the validity of the principles $\Box 4_A$, $\Box 5_A$ and $\Box D_A$. As I will argue in the next chapter, the reason why some instances of $N2(T_@)$ are false is that the indexical operator @ behaves differently in the scope of $A$ than outside of it. More specifically, it refers (in a way) to the actual world in its second occurrence, but not in its first. Clearly, this makes a difference if @ is embedded in $\Box$, and this can be used to explain the failure of $N2(T_@)$. However, the situation is different for $\Box 4_A$, $\Box 5_A$, and $\Box D_A$. If we substitute $T_@$ for $p$ in these formulas, @ occurs always inside the scope of an occurrence of an $A$, and therefore, the mismatch in reference for @ does not occur. Therefore, there is no reason to assume that the counterexamples to $\Box T_A$ are also counterexamples to $\Box 4_A$, $\Box 5_A$, and $\Box D_A$.

### 5.3.3 $N1$

So to get from **Min** to **2D**, we need a rule and some axioms that are plausible independently of two-dimensional semantics, as well as $N1 = Ap \rightarrow \Box Ap$. Not assuming two-dimensional semantics, is it plausible that if it is a priori that $p$, it is necessarily a priori that $p$? Using the earlier explication of apriority: if $p$ can be conclusively non-experientially justified on ideal rational reflection, does it follow that $p$ can necessarily be conclusively non-experientially justified on ideal rational reflection? Indeed, this seems quite plausible – whether something is in the reach of ideal rational reflection should not be contingent. However, $N2$, the other premise of the nesting argument, says that necessarily, if $p$ is a priori, then $p$. So it says that necessarily, if $p$ can be conclusively non-experientially justified on ideal rational reflection, then $p$. This is also very plausible. But as the nesting argument shows, they cannot both be accepted.

I take the nesting argument to be a genuine puzzle independent of two-dimensional semantics – the premises seem correct, but the conclusion is false.

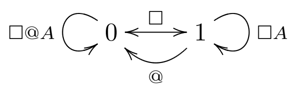
So although $N1$ is in fact *prima facie* plausible even independently of two-dimensional semantics, the nesting problem shows that we have to be careful here, as it requires us to give up the equally plausible principle $N2$. So $N1$ is a commitment of two-dimensional semantics that goes beyond what one can expect most philosophers to agree with. However, I will argue in the next chapter that accepting $N1$ is in fact the correct answer to the nesting problem, independently of two-dimensional semantics. If the arguments I will put forward there are convincing, they will show that $N1$ is a commitment which is plausible independently of two-dimensional semantics, even in the light of the nesting problem. Together with the arguments in this chapter, this establishes that all theorems of **2D** are plausible independently of two-dimensional semantics.

To conclude this chapter, I will show that $N1$ is essential for $P^+_{\mathbf{Min}}$ to produce **2D**. Of course, if this would not be the case, then the discussion of the nesting problem independently of two-dimensional semantics would not be needed, which would save us a lot of work. Therefore, I will now show that $N1$ is needed, in the sense that removing it from the proof system $P^+_{\mathbf{Min}}$ gives a proof system that does not produce all theorems of **2D**.

**Definition 5.41.** *Let $P^*_{\mathbf{Min}}$ be the proof system containing the same rules and the same axioms, except for $N1$, as $P^+_{\mathbf{Min}}$.*

**Proposition 5.42.** $\nvdash_{P^*_{\mathbf{Min}}} N1$.

*Proof.* Let $\mathcal{F} = \langle W, R_\Box, R_@, R_A, D\rangle$, where $W = \{0,1\}$, $D = \{0\}$, and the relations are given by the following diagram:




We first show that $\mathcal{F} \Vdash L(P^*_{\mathbf{Min}})$ by induction on $P^*_{\mathbf{Min}}$. We can verify that $\mathcal{F} \Vdash$ **Min** by checking the conditions in Proposition 5.12. Since $R_A$ is an equivalence relation, $\Box 4_A$, $\Box 5_A$, and $\Box D_A$ are valid in $\mathcal{F}$. Since only 0 is accessible by $R_A$ from 0, for any formula $\varphi$ such that $\mathcal{F} \Vdash \varphi$, also $\mathcal{F} \Vdash A\varphi$. So the validities in $\mathcal{F}$ are closed under $Gen_A$. Finally, since $\mathcal{F}$ is an FWDE, its validities are closed under *MP* and *US*.

We now show that $\mathcal{F} \nVdash N1$. Let $V$ be a valuation for $\mathcal{F}$ such that $V(p) = \{0\}$. Then since $\langle \mathcal{F}, V\rangle, 0 \Vdash Ap$ and $\langle \mathcal{F}, V\rangle, 0 \nVdash \Box Ap$, the claim follows. ∎

# Chapter 6

# The Nesting Problem

As we have seen in the last chapter, the logical commitments of two-dimensional semantics that cannot be assumed to be widely accepted among philosophers are completely captured by the formula $Ap \rightarrow \Box Ap$. Since this formula represents one of the premises of the nesting problem, I discuss this problem in this chapter, and argue independently of two-dimensional semantics that counting $Ap \rightarrow \Box Ap$ as valid is in fact the correct answer to the problem. Together, this shows that all of the logical commitments of two-dimensional semantics as captured by **2D** can be argued for independently of two-dimensional semantics.

I have already presented the nesting problem in section 1.4.2, and applied the logic of two-dimensional semantics of chapter 3 to it in section 3.2. In the first section of this chapter, I will describe the nesting problem again, and argue that it is more general than it may seem at first, by presenting a variant using different modalities. In the second section, I will propose a general solution to the problem by postulating an ambiguity, which allows us to preserve our intuitions in favor of both premises without having to accept the conclusion of the problematic argument.

## 6.1 A General Dilemma

The nesting problem derives from arguments against two-dimensional semantics. In this section, I present it in a general form, which shows that it is not just a problem for two-dimensionalists. I then go on to show that a similar problem can be devised for other modalities besides necessity and apriority.

### 6.1.1 Not Just for Two-Dimensionalists

A relatively complicated argument is described in Soames (2005, see *Argument 5* on pp. 278–279), which is intended to show that some of the claims implied by two-dimensional semantics are inconsistent. Crucially, this argument relies on nesting "actually" inside the scope of "a priori", which in turn is nested inside the scope of "necessary". A more abstract form of this argument is discussed in Dever (2007, see *Version 2* on p. 11), where it is presented as a challenge for, rather than an argument against, two-dimensional semantics. From this argument, the nesting problem was derived in Chalmers (2011, endnote 25) in the form in which I will discuss it here. I have already presented it in chapter 1 in the form of the following schematic argument, with premises (A1) and (A2), and conclusion (A3):

(A1) If it is a priori that $p$, then it is necessarily a priori that $p$.

(A2) Necessarily, if it is a priori that $p$, then $p$.

(A3) If it is a priori that $p$, then it is necessary that $p$.

As indicated in section 5.3.3, I believe that both premises are plausible independently of two-dimensional semantics. This is not difficult to motivate: the first premise seems to say that it is not contingent what is a priori. As we have seen earlier, what is a priori depends on what can be conclusively non-experientially justified on ideal rational reflection. What can be justified in this way seems not to depend on contingent features of the world, and so it should not be contingent what is a priori. The second premise seems to say that necessarily, what is a priori is the case. Again, using the explication of apriority, this means that necessarily, what can be conclusively non-experientially justified on ideal rational reflection is the case. It certainly seems true that what can be conclusively non-experientially justified on ideal rational reflection is the case. This also seems not to rely on any contingent feature of the world, and therefore should be necessarily true. So necessarily, what is a priori is the case.

Therefore, both premises can be motivated independently of two-dimensional semantics, and so – as Chalmers stresses as well – the nesting problem is a problem for everyone who accepts the existence of contingent a priori truths, and not just proponents of two-dimensional semantics.

As we have seen in section 3.2, we can represent the premises and the conclusion of the argument of the nesting problem in the formal language as follows:

$$
\begin{array}{ll}
N1 & Ap \rightarrow \Box Ap \\
N2 & \Box(Ap \rightarrow p) \\
N3 & Ap \rightarrow \Box p
\end{array}
$$

It follows from Proposition 3.4 and the completeness result of Theorem 3.18 that $\{N1, N2\} \vdash_{\mathbf{2D}} N3$, so the argument is valid according to **2D**. However, although $N1$ is a theorem of **2D**, $N2$ and $N3$ are not. Therefore, according to the logic **2D**, two-dimensionalists need not accept the conclusion of the nesting argument, but can answer it by denying that $N2$ expresses a logical truth. I will argue below that this is in fact the correct response to the nesting problem concerning the logic of necessity and apriority. But first, I will show that one natural argument for $N1$ is incorrect, and then argue that the nesting problem as just presented is an instance of a more general problem, by describing a similar problem using some other modalities.

As Chalmers (2011, endnote 25) observes, it is quite natural to come up with the following argument in favor of $N1$: besides analyzing apriority in terms of justification as it was done above, we could also analyze it on a different level, and say that $p$ is a priori iff $p$ is knowable apriorily. Further, $p$ is knowable apriorily just in case it is possible that there is someone who knows $p$ apriorily. If we use $A^*p$ to formalize that there is someone who knows $p$ apriorily, then $N1$ is just $\Diamond A^*p \rightarrow \Box\Diamond A^*p$. Since this is a substitution instance of the axiom $5_\Box$, it is valid.

There are two problems with this argument. Firstly, it assumes that the modality hidden in "knowable" is metaphysical possibility. Although this may be the case, it cannot be assumed without further argument. Secondly, even if metaphysical possibility is the correct modality, we do not only have to assume that $p$ is a priori iff it is possible that $p$ is known apriorily, which would be formalized by $Ap \leftrightarrow \Diamond A^*p$, but that this is the case necessarily, which would be formalized by $\Box(Ap \leftrightarrow \Diamond A^*p)$. This is a much stronger assumption, and it could be argued that it begs the question.

### 6.1.2 Not Just for Necessity and Apriority

To illustrate that the nesting problem is not specific to necessity and apriority, I will now give a similar problem for knowledge and temporal modality. Variants of the nesting problem involving doxastic and epistemic modalities were already proposed in Soames (2005), and one of them is formally analyzed in Michels (2011). The argument I want to present is more closely based on Chalmers' version of the nesting problem presented above. The basic idea is this: let Jane be a perfectly rational agent. Since it is a logical truth that if it is raining now, it is raining, Jane – being a rational agent – always knows this. Furthermore, it is always the case that what is known is true. So always, if Jane knows that if it is raining now, it is raining, then if it is raining now, it is raining. But then by ordinary logical assumptions, it follows that always, if it is raining now, it is raining, which means that if it is raining, it is always raining.

More formally, we can assume that we have a QNML with operators $\triangle$ (for "always"), $N$ (for "now") and $K$ (for "the agent knows that"). As premises of the argument, we assume that $\triangle(Kp \to p)$ is a theorem (knowledge always implies truth), and that the rule $Gen_{(\triangle K)}$ of generalization for $\triangle K$ holds (what is a theorem is always known by the agent). Besides these, we only need two assumptions about the logic of "always" and "now", which are just analogs of theorems for $\Box$ and @ in **Act**, namely $Np \to p$ and $p \to \triangle Np$. So we assume:

| | |
|---|---|
| $\triangle T_K$ | $\triangle(Kp \to p)$ |
| $Gen_{(\triangle K)}$ | From $\varphi$, derive $\triangle K\varphi$. |
| $T_N$ | $Np \to p$ |
| $I\triangle N$ | $p \to \triangle Np$ |

With this, we can formalize the argument as follows:

| | | |
|---|---|---|
| (1) | $\triangle(Kp \to p)$ | $\triangle T_K$ |
| (2) | $\triangle Kp \to \triangle p$ | $K_\triangle$, $US$, (1), $MP$ |
| (3) | $\triangle K(Np \to p) \to \triangle(Np \to p)$ | (2), $US$ |
| (4) | $Np \to p$ | $T_N$ |
| (5) | $\triangle K(Np \to p)$ | (4), $Gen_{(\triangle K)}$ |
| (6) | $\triangle(Np \to p)$ | (3), (5), $MP$ |
| (7) | $\triangle Np \to \triangle p$ | $K_\triangle$, $US$, (6), $MP$ |
| (8) | $p \to \triangle Np$ | $I\triangle N$ |
| (9) | $p \to \triangle p$ | (7), (8), propositional logic |

Clearly, the conclusion that if it is raining, it is always raining is not acceptable, so one of the assumptions must be wrong. I take it that the principles $T_N$ and $I\triangle N$ are not up for debate, as well as the use of a QNML. Therefore, the question is whether $\triangle T_K$ or $Gen_{(\triangle K)}$ is correct. Note that $\triangle T_K$ is analogous to $N2$, and that $Gen_{(\triangle K)}$ bears some connection to $N1$. In the following, I will first argue that $\triangle T_K$ should not be assumed to be valid, and then extend this argument to $N2$.

## 6.2 A Proposal

I will now argue independently of two-dimensional semantics that on the level of the propositional logics used, the correct response to the nesting problems described above is to deny that $N2$ and $\triangle T_K$ are theorems of the respective logics, but to accept that $N1$ is a theorem and $Gen_{(\triangle K)}$ is a valid rule. Note that these are claims about which formulas should be theorems of the respective logics. What is philosophically more significant are the natural language sentences that are represented with these formulas. I therefore start with the natural language sentences.

In the nesting problem for necessity and apriority, it may seem that our intuitions clearly favor (A2) over (A1): although we might not know well what is a priori in counterfactual circumstances, we have the firm opinion that also in those, apriority implies truth. This is suggested in Chalmers (2011, endnote 25), and a way of amending the analysis of apriority given by two-dimensional semantics is sketched that gives a different answer to the nesting problem. According to it, (A1) has false instances, whereas (A2) does not.

I am not happy with this proposal, as I think that we also have firm intuitions for (A1). Therefore, I would prefer an answer to the nesting problem which explains why we would like to accept both premises. A promising strategy to obtain such an answer is to postulate some kind of ambiguity, and this is what I will try to do in the remainder of this chapter. I will start with the example of knowledge and temporal operators. I want to suggest that the ambiguity is caused by two ways of reading the ascription of knowledge in the problematic argument. To do so, I will introduce some standard distinctions concerning such states.

### 6.2.1 Attitudes

Attitudes such as knowledge are often categorized in two groups: attitudes *de dicto* and *de re*. Roughly, a *de dicto* attitude is one that is had with respect to a proposition, whereas a *de re* attitude is one that is had with respect to an object. An example for the former is my knowledge that there are no unicorns, whereas an example of the latter is my knowledge of this pen (a pen that is lying in front of me) that it is blue. It should be noted that an analogous distinction can be made on the level of attitude *reports* (statements reporting attitudes of agents) rather than the attitudes themselves, and that although there are obviously connections between the two, they should be kept apart.

As Perry (1979) and Lewis (1979) have argued, some attitudes are special in that they essentially consist in the agent ascribing a property to themselves. E.g., if my pants are on fire, then under suitably unfavorable circumstances, I might read that Peter's pants are on fire as well as see myself in a mirror, which might give me *de dicto* as well as *de re* states of belief that my pants are on fire, without forming the belief of myself that my own pants are on fire which is necessary to quickly take action. Lewis calls these attitudes *de se*.

As Lewis argues, we can understand attitudes *de dicto* as special cases of attitudes *de se*. The idea is that a *de dicto* attitude has as its object a proposition, which can be understood as a class of possible worlds. Such an attitude can then also be understood as a *de se* attitude that the agent is located in one of these worlds.

### 6.2.2 Knowledge and Temporal Operators

I want to suggest that the distinction between *de se* and *de re* can explain our conflicting intuitions concerning the nesting problems. As this distinction is more commonly applied to knowledge than to apriority, let me start with the nesting problem using knowledge and temporal operators. To help our intuitive judgement, I will use the concrete example mentioned above, rather than the abstract argument. Given that we accept that it is a logical truth that if it is raining now, it is raining, we can ask whether we should accept the relevant rule of generalization, and so that it is therefore always known by Jane (our perfectly rational agent) that if it is raining now, it is raining. Given the earlier presentation, we therefore have to decide which one of the following two to reject (to disambiguate different possible readings, I specify formulas that represent the intended readings of the natural language sentences; here, $p$ stands for "if it is raining now, it is raining"):

(B1) Always, Jane knows that if it is raining now, it is raining.
$\triangle Kp$

(B2) Always, if Jane knows that if it is raining now, it is raining, then if it is raining now, it is raining.
$\triangle(Kp \to p)$

First of all, I have to motivate that there are two ways of reading the statement "Jane knows that if it is raining now, it is raining"; one reporting a *de se* and one reporting a *de re* state of knowledge of Jane. It is natural to give the following *de se* reading: Jane ascribes to herself being in a situation (a centered world, as Lewis would say) in which it is the case that if it is raining at the time of that situation, then it is raining in that situation. This is clearly a quite insubstantial state of knowledge. But there is also a *de re* reading of the report on which it says that Jane knows of the point in time of the utterance of the attitude *report* that if it is raining at it, then it is raining. This is a more substantial state of knowledge. If I am correct in claiming that the attitude report has these two readings, then the two readings are also available in (B1) and (B2). It is clear that no good argument can result if we read (B1) in one way and (B2) in another, so there are two sensible readings of the premises of the argument, one *de se* and one *de re*.

I start with the *de se* reading. As noted above, on this reading, the state of knowledge reported is quite trivial, so it is safe to assume that any perfectly rational agent (the kind of agents modeled by epistemic logics) is always in this state. So on this reading, (B1) is very plausible: Jane is always in the state of knowing that if it is raining at the time of the situation in which she is, then it is raining in that situation. In contrast, (B2) is not plausible on the *de se* reading: there may well be a time at which Jane is in the appropriate state of *de se* knowledge – in fact, as just argued, this is a trivial condition – although it is not raining despite it now raining. Moreover, we can give an explanation of why the usually uncontentious move from knowledge to truth fails under the "always" operator on the *de se* reading: on it, "now" does not refer to the time of utterance of (B2) in the antecedent, whereas it does so in the consequent. (So according to my account of the *de se* reading, "knows" can change the referent of the indexicals in its scope; in the terminology of Kaplan (1989b), it is a *monster*.)

On the *de re* reading, the plausibility of the premises is reversed. As noted above, on the *de re* reading, the state of knowledge reported in (B1) is quite substantial, and there is no reason why even a perfectly rational agent should always possess such knowledge. Further, (B2) sounds very plausible on this reading: at any time, if Jane knows of the time of utterance that it has a certain property, then it should also have this property. In contrast to the *de se* reading above, there is no reason to assume that the usual assumption that knowledge implies truth does not always hold.

If this ambiguity analysis is correct, then there are two ways to read the problematic argument, one on which (B1) fails, and one on which (B2) fails. For the correct logic of the three modalities used, the question is now simply which reading of the knowledge operator the logic is supposed to capture. We could just say that this is a matter of arbitrary choice, but I think there are good reasons to favor the *de se* reading. *De re* readings are best formalized in a quantified logic by appropriate scopings of variable-binding operators and modalities. Since the logic used above uses a propositional language, I think it is more natural to take it to model the *de se* reading. Even if specific natural language sentences like (B1) and (B2) are more plausibly read *de re*, I think it would be better to use the logic to model a somewhat artificial *de se* reading – the semantics of attitude ascriptions is notoriously difficult, and there is no reason why a propositional logic should not constitute a substantially simplified account of natural language semantics. Therefore, we should answer the nesting problem for knowledge and temporal operators by denying that $\triangle T_K$ is a theorem.

### 6.2.3 Necessity and Apriority

To argue that we can also distinguish between a *de se* and a *de re* reading of the nesting argument using necessity and apriority, I have to motivate that claims about apriority can also be read in these two ways. This might seem like a strange suggestion, but it becomes more reasonable if we analyse apriority using a priori knowledge. As I have noted above, $p$ is a priori iff $p$ is knowable apriorily, or we might say: iff a priori knowledge of $p$ is obtainable. On this explication, questions about apriority are concerned with obtainable states of a special kind of knowledge, and we have already applied the *de se*/*de re* distinction to these.

I will now consider the premises of the argument of the nesting problem for a standard example of a contingent a priori truth, using the analysis of apriority in terms of a priori knowledge. The example is the contingent truth that I am here. If "I am here" is not a priori because it apriorily entails "I exist and am spatially located" and the latter is not a priori (see Chalmers (2006, section 4.1)), we can use the conditional "If I exist and am spatially located, then I am here" instead. For present purposes, this only makes the example slightly more complicated.

So assuming that "I am here" is a priori, the question is whether we can use (A1) to derive from this that it is necessarily a priori that I am here. As we do not want to conclude that I am necessarily here, we have to decide which of the following two to reject:

(C1) Necessarily, a priori knowledge that I am here is obtainable.
$\Box Ap$

(C2) Necessarily, if a priori knowledge that I am here is obtainable, then I am here.
$\Box(Ap \rightarrow p)$

For the sentence "I know that I am here", it is not difficult to discern a *de se* and a *de re* reading. On the first, it states that I am in the relatively trivial state of knowing to be in a situation in which the agent of the situation is at the situation's location. On the second, it states that I and here are such that I know of the former that it is at the latter. I'm not sure whether this constitutes a substantial state of knowledge.

In (C1) and (C2), the state of knowledge is not attributed to the speaker of the utterance via the indexical "I"; rather, we can say that it is attributed to an arbitrary individual with the phrase "a priori knowledge that I am here is obtainable". But there is still a natural way of distinguishing a *de se* and a *de re* reading: on the *de se* reading, it says that the state of knowledge which I expressed by "I know that I am here" on the *de se* reading is obtainable in a way such that it constitutes a priori knowledge; and on the *de re* reading, it says the same of the state of knowledge which I expressed by the same statement on the *de re* reading. As before, this distinction gives us two readings of the premises of the argument. Again, I start with the *de se* reading.

As the relevant *de se* state of knowledge is quite insubstantial, it is plausible that it is necessarily obtainable in a way that constitutes a priori knowledge, no matter how exactly "obtainable" has to be understood. That is, it is plausible that the state of a priori knowledge to be in a situation in which the agent of the situation is at the situation's location is necessarily obtainable. So (C1) is plausible on the *de se* reading. But as in the case of the nesting problem using knowledge and temporal modalities, in the antecedent of (C2), "I" and "here" do not refer to the speaker of the premise and their location, whereas they do in the consequent. So there is no reason to assume that the move from knowability (obtainable knowledge) to truth is guaranteed. So (C2) is not plausible on the *de se* reading.

In contrast, there is no reason on the *de re* reading why we cannot necessarily move from knowability to truth, so (C2) is plausible: necessarily, if I and here are such that it is a priori knowable that the former is at the latter, then I am here. But since I am not necessarily here, (C1) must be false. Given the murky nature of the concepts of a priori knowability or obtainability, there is some room for discussion here. But for my current argument, this is not so important: just as in the case of knowledge and temporal operators, we are working here with a propositional logic, and in this, it is most natural that formulas represent the *de se* readings of the natural language sentences they are supposed to capture. On this reading, (C1) is plausible, but (C2) is not. But then it follows that the instance of (A1) where $p$ is replaced by "I am here" is plausible, whereas the corresponding instance of (A2) is not. So since we take propositional formulas to capture the *de se* reading, the example of "I am here" indicates that $N1$ should be counted as valid, rather than $N2$. That (C2) is plausible on a *de re* reading is only important insofar as it is supposed to explain our intuitions in favor of $N2$.

It might be objected that the *de se* readings of (C1) and (C2) are even more unnatural than the *de se* readings of (B1) and (B2). In fact, it could be argued that (C1) and (C2) do not allow *de se* readings at all. This would not be such a big problem for the proposed solution as it might seem: if the *de se* readings are not available, then the natural language sentences (C1) and (C2) can only be formalized in a quantified logic. Therefore, they say nothing about the validity of formulas in the propositional logic, and in particular don't contradict the claim that $N1$ should be counted as valid and $N2$ as invalid.

This concludes my argument for $N1$ in the formalization of the nesting problem. I have argued that there is a *de se* and a *de re* reading of the argument, that on the *de se* reading, (A1) should be accepted and (A2) rejected, and that the *de se* reading is the one that is most naturally formalized with a propositional language, which shows that $N1$ should be understood as being valid, and $N2$ as invalid. Note that this argument does not use any resources of two-dimensional semantics. Therefore, if it is successful, it shows that $N1$ is plausible independently of two-dimensional semantics. In combination with the considerations in chapter 5, it constitutes an argument that the logic of necessity, actuality, and apriority given by two-dimensional semantics as specified by **2D** does not contain any theorems that are not already plausible independently of two-dimensional semantics. Hence it could be said to show that two-dimensional semantics gets the logic of necessity, actuality and apriority right, and thereby constitute an indirect argument in favor of the theory.

I want to conclude the discussion of the nesting problem by considering two natural objections against my account.

### 6.2.4 Objection 1: Contingent A Priori Without Indexicals

The first objection is this: the strategy described above may work for all cases of contingent a priori truths that involve indexical singular terms. It may also be possible to extend it to cases involving indexical operators like "actually", such as "if it is raining, it is actually raining", using a strategy similar to the one used above for "now". (Interestingly, such a reading of the actuality operator is used in Soames (2007, p. 256)). But there are also cases of contingent a priori truths that do not involve any indexicals. E.g., Williamson (1986) has suggested that it is contingent and a priori that there is at least one believer, although no indexicals are needed to express this. There seems to be no way of distinguishing a *de se* and a *de re* reading of the relevant instances of the nesting problem, so the strategy outlined above will not work for these cases.

One way of answering this objection is to deny that these statements are a priori. This answer would be natural for Chalmers, who holds this position for independent reasons (see Chalmers (2006, section 4.1)). But we can also just accept that although it may sound strange, in a possibility in which there are no agents, it is a priori that there is at least one believer, although there is no believer. I believe that our intuitions concerning such particular cases are not very strong. In fact, in my answer to the second objection, I will suggest that our strongest reasons for wanting to accept ($A2$) are problematic quantified readings. It it also interesting to observe that while both nesting arguments can be constructed for the cases using indexicals discussed so far, Williamson's example cannot be used to construct an instance of the nesting problem for knowledge and temporal operators.

### 6.2.5 Objection 2: Quantifications

The second objection is this: even if we read all knowledge attributions and apriority claims thoroughly *de se*, the proposed solution to the nesting problem commits us to the falsity of (B2) and (C2), which by existential generalization commits us to the following two:

(D) Something is at some point in time both known and false.

(E) Something is possibly both a priori and false.

Assuming that the propositional entities quantified over are abstract, and therefore have eternal and necessary existence, the converse Barcan formula for propositional quantifiers is valid, and therefore (D) and (E) imply the following:

(D') At some point in time, some falsehood is known.

(E') Possibly, some falsehood is a priori.

However, we have very strong intuitions against these.

Although I think that the application of the converse Barcan formula is at least in need of further argument, there is a more basic problem with this objection, which already applies to (D) and (E). Abstractly, the problem is that the argument assumes that there is a single kind of entity which plays all the roles propositions are usually assumed to play, but as is often observed, this is quite implausible. (It should be noted that the position that denies the existence of such a kind, which is sometimes called *semantic pluralism*, very naturally combines with two-dimensional semantics; see Chalmers (2004, p. 167).)

More specifically, note that on the standard Lewisian account of *de se* attitudes, the object of a *de se* state of knowledge is a class of centered possible worlds. So since (D) and (E) are existential claims that involve predicating being known or being a priori of something, witnesses for them should be such classes. But they also predicate falsity of these things. And what could it mean to say that such a class is false at some point in time or some possible world? The semantic content of a phrase like "if it is raining now, it is raining" that is relevant for evaluating it under "always" or "necessarily" may not be contained in such a class of centered possible worlds. In other words, in a concrete instance like (B2) or (C2), knowledge or apriority may be predicated of some semantic content expressed by a phrase, while truth is predicated of another semantic content of the same phrase, which makes the existential generalization inapplicable.

This answers the second objection, and it touches an important methodological point concerning the nesting problems: we must be careful to separate our intuitions concerning specific instances of sentence schemas from those concerning their quantified generalizations. This also shows what was wrong with the initial argument for the plausibility of (A1) and (A2) in section 6.1.1. Moreover, it should not be argued that we have any reliable intuitions about *formulas* like $N1$ or $N2$ – since these involve propositional letters, any informal interpretation must be quantified in some way. This also gives us a second reason (besides the *de re* readings) why we intuitively want to accept premises (B2) and (C2) – we have strong intuitions for the generalized statements, and wrongly extend them to these instances.

# Chapter 7

# Conclusion

To conclude, I will first summarize what has been achieved in the thesis, and then sketch some ways of extending the logic that I have presented.

## 7.1 Summary

I have formalized the account of necessity, actuality and apriority given by two-dimensional semantics in a natural way as the propositional modal logic **2D**, using a semantics given by a class of frames with distinguished elements. I have axiomatized this logic, and noted some of its properties, which were in accord with common philosophical judgements.

Comparing this logic to the formal systems discussed in Davies and Humberstone (1980) and Restall (2010), I have argued that it is semantically equivalent to the second, and differs in a number of respects from the first. I have then given a lengthy argument for all the theorems of **2D** independently of two-dimensional semantics. For this, I first presented a minimal logic that should be uncontroversial, and showed that adding a number of axioms and a rule suffices to produce **2D**. For all of these except one axiom, it was not difficult to argue for their plausibility. The remaining axiom was a premise of the nesting argument, and I went on to argue that accepting it is the correct choice, still independently of two-dimensional semantics.

I hope that this research indicates that two-dimensional semantics can in fact be developed into a coherent systematic semantic theory. Furthermore, the fact that its logic of necessity, actuality and apriority can be completely motivated independently of the theory shows that a central aspect of two-dimensional semantics is compatible with the prevalent understanding of the three modalities involved. I hope that these observations strengthen the appeal of two-dimensional semantics.

## 7.2 Extensions

There are many possibilities of extending **2D** by adding operators capturing additional concepts. Such extensions are interesting for a number of reasons: firstly, they can be seen as a continuation of the project of formalizing two-dimensional semantics. Secondly, they may show that two-dimensional semantics enables us to give good semantic accounts of notions that have so far proven difficult to analyze. Thirdly, such extensions may be applied to philosophical problems which are independent of two-dimensional semantics. I will give examples for such cases below. Finally, such extensions will bring difficulties, and solving these difficulties may help testing and clarifying the theory of two-dimensional semantics.

### 7.2.1 Propositional Quantifiers

A first extension would be to add quantifiers binding propositional letters. As we have seen in section 6.2.5, this may help to solve the nesting problem, by clearly differentiating between a formula like $\Diamond(Ap \land \neg p)$ and its existential generalization $\exists p \Diamond(Ap \land \neg p)$. Continuing the FWDE-based semantics given by MD, we might interpret propositional quantifiers as ranging over all subsets of points in a frame, but as indicated in section 6.2.5, this might capture only one of a number of semantic functions that can be performed by propositional quantifiers. Also, as noted in section 3.5.1, adding propositional quantifiers may make the logic sensitive to the differences between MD and RecD (or RD).

An application of such an extension to a philosophical issue that is independent of two-dimensional semantics can be found in Tharp's Theorems, which were presented in Tharp (1989), and discussed in Lewis (2002) and Humberstone (2004). Tharp makes three claims about the existence of certain truths. E.g., one of them says that every truth is a priori equivalent to a necessary truth. Using propositional quantifiers, this is straightforwardly formalized as $\forall p(p \to \exists q(\Box q \land A(p \leftrightarrow q)))$, and so such an extended logic of two-dimensional semantics would provide an independently motivated system in which these claims can be discussed formally.

### 7.2.2 Knowledge and Belief

Another natural extension would be to add operators for knowledge or belief. In the light of the discussion of the nesting problem, it is best to focus on *de se* attitudes. Although it is not obvious how the semantics of these operators should be defined, it would be natural to start from the account of *de se* attitudes in Lewis (1979), according to which the object of such an attitude is a class of centered worlds. Since these are formally represented by sets of points on the diagonal of a matrix FWDE, we can use such sets in some way to interpret operators representing attitudes. As remarked in connection with the nesting problem, such a semantics would of course not completely follow the capriciousness of natural languages concerning attitude reports, but for a propositional logic, this is not necessarily a defect. A system that roughly follows these lines has already been developed in Michels (2011).

If an operator *SK* formalizing "someone knows that" is added to the extension with propositional operators mentioned above, then a system results in which the knowability paradox (also called *Fitch's paradox*) can be discussed. This is an argument that recently has received considerable attention (see Salerno (2009)) which concludes that if some truth is not known, then there is a truth which cannot be known. In the proposed formal language, this could be represented by $\exists p(p \land \neg SKp) \to \exists p(p \land \neg\Diamond SKp)$. Interestingly, it could then also be contrasted with the variant in which possibility is replaced by conceivability. Further, the fact that the logic would also contain the actuality operator would make it easy to connect to the discussion in Edgington (1985), where it is argued that such an operator is needed for a proper discussion of the problem. A formalization of Fitch's paradox that is sensitive to issues of double-indexing has already been undertaken in Rabinowicz and Segerberg (1994), but a logic developed out of **2D** has the additional bonus of being independently motivated.

### 7.2.3 Predication and Quantification

As a final example for an extension of the logic of necessity, actuality and apriority as formalized by **2D**, I want to mention the move from a propositional logic to a predicate logic, which might be enriched with quantifiers over first-order variables, and even quantifiers over higher-order variables. To conveniently mark scope distinctions, it might also be interesting to add $\lambda$-abstractors. One of the many difficulties that can be expected in developing such an extension is that it requires us to make sense of quantification into the scope of "a priori", which ties in with the discussion of *de re* readings of claims about apriority in chapter 6.

In such a system, many claims made by two-dimensional semantics could be investigated formally in much greater detail than in the propositional languages discussed so far. Specifically questions concerning reference, which have been almost completely neglected in this thesis, would now become important. One of the most interesting aspects of such a logic would be that it would allow us to formally describe more intimately the connection between necessity and apriority which two-dimensional semantics draws, as mentioned in section 4.1.5.

# Appendix A

# Consequence

In section 1.2.3, I noted that there are two ways of defining consequence and validity in logics for indexicals, namely the general and the real-world definitions. In the matrix semantics for the logic of necessity, actuality and apriority discussed in chapter 3, we had the analogous options of using the class of frames $\mathsf{M}$ (corresponding to the general definitions) or the class of FWDEs $\mathsf{MD}$ (corresponding to the real-world definitions) as the semantics, which characterize the logics **2Dg** and **2D**, respectively. I have claimed that the definitions of real-world consequence and validity are philosophically more plausible than those of general consequence and validity, and that therefore, $\mathsf{MD}$ and **2D** should be seen as the philosophically relevant systems. However, in the literature on logics for indexicals, the real-world definitions are not universally accepted. As adopting them rather than the general definitions makes a substantial difference to the logic – we get additional validities, but we lose normality – the matter deserves more attention. I therefore use this appendix to argue for the real-world definitions.

In general, the general validities are a sublogic of the real-world validities. Correspondingly, as we have seen above, $\mathbf{2Dg} \subseteq \mathbf{2D}$. In this case, the question which definitions are correct comes down to whether it is philosophically correct to count the elements of **2D** that are not contained in **2Dg** as validities. Simple examples for such formulas are $T_@ = @p \rightarrow p$ and $T_A = Ap \rightarrow p$. The last formula in particular expresses something we have firm opinions about, and this gives us an easy way to decide between **2D** and **2Dg**: just as it is part of the logic of necessity that what is necessary is true, so it is part of the logic of apriority that what is a priori is true (cf. section 5.1.2). Therefore, $T_A$ must be a validity.

One may be tempted to conclude that **2D** must be correct. But this conclusion isn't quite supported by the argument. The observation that $T_A$ should be valid is fine as it goes, but it only shows that **2Dg** cannot be the correct logic of the three modalities. It may still be the case that $T_@$ should not be counted as valid. It would then follow that neither **2D** nor **2Dg** is an adequate logic. So the fact that $T_A$ should be seen as valid only shows that if one of **2D** and **2Dg** is correct, then it is **2D**. To convincingly defend **2D** against arguments against real-world validity, we have to show that it is correct to classify formulas representing indexical truths (truths that cannot be uttered falsely) like $T_@$ as valid formulas.

This issue is also important for another reason. In chapter 5, I assumed that the correct logic of necessity and actuality is **Act**, which is the logic of the two modalities according to real-world validity. This constituted part of my argument that **2D** is a plausible logic of necessity, actuality and apriority, independently of two-dimensional semantics. So it is important that formulas representing indexical truths like $T_{@}$ should in fact be valid. Therefore, I will argue for real-world validity in this appendix, in particular, I will argue that formulas representing indexical truths should be counted as valid. As a test-case, I will often use $T_{@}$, just because it is a very simple example.

In the next section, I will describe some positions and an argument from the literature on indexicality. In the succeeding section, I will state some considerations that indicate that the real-world definitions are the correct ones.

## A.1 Positions from the Literature

I will first give an overview of the literature on the dispute between general and real-world validity and consequence, and note that most authors do not give arguments for the definitions they adopt. I will then discuss a recent argument for general validity, and explain why I am not convinced by it.

### A.1.1 An Overview

Naturally, the first explicit occurrence of the question about the correct definitions of consequence and validity is in the early works on the logic and semantics of indexicals.

In some of these, such as Meredith and Prior (1965) or Hodes (1984), two notions are defined that correspond to general and real-world validity, without discussion whether one of them should be preferred over the other. Mostly, however, real-world validity or an analogous notion is assumed to be correct, without much discussion. E.g, this is done in Prior (1968a), Kamp (1971), Vlach (1973), Hazen (1978), or Kaplan (1978). However, some authors lean towards taking general validity to be the correct notion, such as Crossley and Humberstone (1977) and Davies and Humberstone (1980). The first of these also includes a short discussion about the two possible ways of defining validity, but just concludes that "[. . . ] some arguing would need to be done by anyone proposing to accord validity to such contingencies" [i.e. adopting real-world validity] (Crossley and Humberstone (1977, p. 15)).

In Kaplan (1989b), the definition corresponding to real-world validity is also presented as the correct one. In Kaplan (1989b, pp. 538–540), the failure of necessitation is discussed and motivated. Similarly, some peculiarities of the definition are discussed and made plausible in Kaplan (1989a, pp. 593–597), but neither of the texts contains a direct argument that the definition given is the correct one, rather than the one that would correspond to general validity.

An explicit discussion of the question can be found in Zalta (1988), who argues for real-world validity. This is criticised in Hanson (2006), where an argument for general validity is given. I will discuss this argument in the next section. Hanson's argument in turn is criticised in Nelson and Zalta (2010), who argue that Hanson's criticism of Zalta is incorrect, and that his argument for general validity fails. I largely agree with Hanson's criticism of Zalta's argument, but I neither agree with his positive argument nor with Nelson and Zalta's criticism of it.

The discussion in these papers focuses on what they call "Tarski's definition of logical truth", and how it should be extended to modal languages. I think this is the wrong approach – in what way Tarski's work on validity and consequence is adequate for extensional languages is itself a matter of debate, and how it is best extended to modal languages will only tell us something about validity and consequence for these languages in a very roundabout way. I believe that it is better to approach the question directly, without any appeal to Tarski's notion of logical truth, and I will do so below. Since I will argue for real-world consequence and validity there, I will present alternative arguments for the same conclusion as in Zalta (1988) and Nelson and Zalta (2010). I don't think it would be very interesting for me to elaborate on why I disagree with their arguments beyond the indications I have just given, especially since I don't disagree with their conclusion. But Hanson (2006) argues for general validity, and since I am not convinced by the arguments against it in Nelson and Zalta (2010), I have to explain why I think Hanson's argument fails. I will do so in the next section.

### A.1.2 An Argument for General Validity

The argument for general validity presented in Hanson (2006, pp. 445–447) can be summarized as follows:

(A) All sentences that are logical truths are analytic.

(B) Some sentences that are instances of real-world valid formulas are not analytic.

(C) Therefore, some sentences that are instances of real-world valid formulas are not logical truths.

As usual, Hanson takes sentences to be sentence *types*, and not particular tokens. Clearly, the conclusion (C) implies that something is wrong with the definition of real-world validity, and thereby constitutes an argument against real-world validity. Hanson assumes the first premise without argument, but substantiates the second premise by arguing that the following sentence is an instance of the existential claim:

(BP) "If Bush actually became President in 2001, then he became president in 2001."

Plausibly, (BP) is an instance of the real-world valid formula $T_{@}$. Hanson argues as follows that (BP) is not analytic: analyticity is truth in virtue of meanings. But (BP) is not true in virtue of its meanings alone, but only true in virtue of its meanings relative to an utterance context or an actual world-candidate – whatever is needed for the indexicality of "actually". Therefore (BP) is not analytic.

As truth in virtue of meanings and truth in virtue of meanings relative to a context is not explained further, it is difficult to assess this argument. Hanson writes that (BP) is not true in virtue of the meanings of its words since it "can only be evaluated relative to some possible world that is taken to be the actual world". Of course, in general, we have to take the context of utterance into account when evaluating the truth of a statement involving indexicals. But I fail to see how this observation bears on the issue of analyticity. Maybe Hanson's objection is this: since (BP) is a sentence type involving an indexical expression, it only has a truth-value relative to a context of utterance, so it is not true simpliciter. Therefore, it is not true, and so it cannot be true in virtue of meanings.

Since truth is not a property of sentence *types*, I agree that the sentence type (BP) is not true. But the reasoning from this to the claim that (BP) is not analytic is fallacious. Consider the following sentence type:

(RS) "If it is raining and there is a storm, then it is raining."

(RS) is not true simpliciter, since we can only evaluate it relative to a time and place (and possible world). So since it is not true, according to the earlier reasoning, it cannot be true in virtue of meanings, and therefore it cannot be analytic. But then according to premise (A), it is not a logical truth, and so classical propositional logic is inadequate since it counts $(p \wedge q) \rightarrow p$ as valid. Obviously, this conclusion is absurd, so something must be wrong with this line of argument.

I think that this just shows that either being true in virtue of meanings is not a way of being true, or analyticity is not truth in virtue of meanings. But that should not be surprising – being true is not a property of sentence types, but analyticity is. Probably, what I have presented is not what Hanson has in mind. But if it isn't, I don't know what might be. So I don't see why (BP) should not be analytic. But if Hanson's reasoning doesn't support the claim that (BP) is not analytic, then he has not given good reasons to take (B) to be true, and so he has not given a convincing argument for the conclusion that real-world validity is incorrect.

## A.2 Real-World Consequence

I believe that to decide between general and real-world consequence and validity, we need to be explicit about what a logic, understood as a mathematical construction, is for. The question is: why do we study these formal structures? Or as Kaplan (1989a, p. 596) puts it: what do you want to *do* with your logic? We have to be aware of the fact that formulas as mathematical objects, even if they are assigned some mathematical objects by another mathematical structure which we call a *formal semantics*, cannot be taken to *mean* something in the way utterances in natural language do. Of course, formal logics are not arbitrary mathematical structures, but how formal logics relate to meanings or natural languages is a non-trivial issue, and one that should be of central importance in the discussion about the correct definition of logical consequence.

Independently of the debate I am concerned with here, the importance of these issues has been stressed in Shapiro (1998) and Shapiro (2005). In these texts, Shapiro also provides a suggestion of how we can understand what we do when we do logic. He sketches the following idea: a formal logic is a model of natural language. Formulas represent sentential entities, and the formal consequence relation represents some relation between these entities. The expression "sentential entities" is meant to be ambiguous between a number of candidates, like sentence types, utterances, or propositions. Which of these are represented by formulas is something Shapiro leaves open, and I will do the same. However, I do think that this is an interesting question, especially in the context of logics for indexicals – see Russell (2008) for more on this issue. Shapiro notes that there might be more than one relation between these entities which we want to model with our formal system, which might all be reasonable precisifications of the meaning of "follows from". One can already read the introductory remarks of Tarski (2002 [1936]) in this light, and the thesis is more explicitly defended in Beall and Restall (2006). I will return to this issue in the concluding section.

I find Shapiro's picture compelling, and I will therefore assume that something of the sort is correct. I will first consider some relations that could be modeled by the formal consequence relation. This will not decide the issue between general and real-world consequence, but it will show us how some independently motivated accounts of what is modeled by a formal consequence relation decide the issue. I will then try to motivate real-world consequence by stating some of my intuitions. I am not unreservedly comfortable with this approach. It seems to me that a lot should be said about the details of the relation between formulas and what they are to represent, as well as the *uses* of logic, e.g. in assessing arguments. But this appendix is not the place to do so, so I will try to circumnavigate these issues.

By arguing for real-world consequence I only argue for the real-world definitions in the context of the logic of necessity and actuality. As it has been argued, e.g. in Predelli (1998), although the usual semantic account of indexicals gives us good reasons to take instances of $T_@$ to be true in virtue of meaning, the claim that a sentence like "I am here" is logically true is much more contentious. I therefore do not argue for the correctness of the definition of consequence and validity in Kaplan's logic of indexicals as presented in Kaplan (1989b), according to which the formula representing "I am here" is valid.

A terminological note: I have used "consequence" and "validity" for the relations and properties as defined in formal logics. I will now also use "consequence" for the relation between sentential entities that is being modeled by a logic. Sometimes, I will say that a conclusion "follows from" some premises, or that an argument is "valid". These are all meant to express the same thing. Note especially that the validity of formulas and the validity of arguments (outside the formal logic) are distinct notions.

### A.2.1 Consequence Relations

I will now consider a few options for the consequence relation that we might want to model with the consequence relation in a formal logic. The most common proposals for consequence use some version of modalized truth-preservation. This abstract characterization is called the "Generalized Tarski Thesis" in Beall and Restall (2006), who state it as follows:

(GTT) An argument is valid if and only if, in every case in which the premises are true, so is the conclusion.

We can now give different accounts of what it is to be a case, and correspondingly, what it is to be true in a case, which gives us different relations that can be modeled by a logic.

A natural way to understand cases is to take them to be possible ways for the world to be, in the sense that the relevant modality is metaphysical necessity.

It is clear that since it is possible that it is actually raining without it raining, $p$ should not be a consequence of $@p$, and in general, that this relation must be modeled by the general consequence relation. But as noted in Humberstone (2004, p. 23), this instance of (GTT) is implausible for reasons that have no direct connection to indexicality. He presents the following argument from Peacocke (1976): John is drinking water; therefore John is drinking $H_2O$. Since water is necessarily $H_2O$, it is necessarily the case that if John is drinking water, he is drinking $H_2O$. So according to the understanding of consequence under consideration, the argument is valid. However, as argued by Humberstone, it is quite implausible that this argument is valid, as the corresponding conditional

(JW) "If John is drinking water, then John is drinking $H_2O$"

is not a priori.

As Humberstone notes, this example suggests that the cases in (GTT) should be ideally conceivable possibilities, in the sense that the relevant modality is apriority. This rules out the problematic argument, and in general, it ensures that the conditional corresponding to any valid argument is a priori. Also, since it is a priori that if it is actually raining then it is raining, according to this proposal, $p$ should be counted as a consequence of $@p$. So if we think of logic as modeling the relation of a priori truth-preservation, the correct definition of consequence seems to be real-world consequence.

There is also the so-called *interpretational account of consequence*. According to it, an argument is valid if and only if truth is preserved from the premises to the conclusion under any reinterpretation of the non-logical constants. Unless we take "water" and "$H_2O$" to be logical constants, this solves the problem with Humberstone's example, as some substitution instances of (JW) are false. And it is also easy to see that according to the interpretational account, "it is raining" is a consequence of "it is actually raining" if we take implication and actuality as logical constants – whatever we insert for "it is raining", the conclusion will be true if the premise is. Again, this points to real-world consequence.

As nice as these considerations are for the definition of real-world consequence, they do not constitute a solid argument for it. Someone might easily come up with an analysis of consequence according to which it does not follow from John drinking water that John is drinking $H_2O$, although it also doesn't follow from it actually raining that it is raining. An example is truth preservation which must hold apriorily necessarily. That is, an argument may be said to be valid if and only if it is a priori that it is necessary that if the premises are true, so is the conclusion. In fact, this is the relation modeled by the class of frames $\mathsf{M}$ in the matrix semantics of chapter 3. This proposal solves Humberstone's worry, since although it is necessary that if John drinks water he drinks $H_2O$, the fact that this is necessary is not a priori. And it doesn't give us real-world consequence, since it is not apriorily necessary that if it is actually raining, it is raining – this follows from the fact that what is a priori is true, and that it is not necessary that if it is actually raining, it is raining.

To resolve the question about general and real-world consequence, we could therefore try to argue for one specific instance of the (GTT), or some other explication of the relation that is modeled by the formal consequence relation. But this paper is not the place for such a difficult task. Rather, I will try to motivate real-world consequence independently of any such specification.

### A.2.2 Logic and Reasoning

To motivate real-world consequence, I want to go back to a question that was mentioned earlier: what do we want to do with the logic? At least one of the most central applications of a logic is to give a criterion of when a piece of reasoning is correct. By way of example, this opinion is expressed in the following quote by John Etchemendy:

> Logic is in part a service discipline, providing precise, idealized models of valid and invalid reasoning, models which in turn help us to describe and understand the process of rational investigation, whether in mathematics, the sciences, or everyday life. (Etchemendy (2008))

This understanding of the function of logic points to real-world consequence, since the inferences that are licensed by real-world consequence in addition to those licensed by general consequence are all intuitively valid pieces of reasoning. In general, this can be seen as a consequence of the fact that we are always in a position to know (apriorily) that we are in the actual world. I will illustrate this with our test-case. In my opinion, (1) is a perfectly good piece of reasoning, and moreover, it adheres to the same standards of reasoning as the arguments (2) and (3):

(1) It is actually raining, therefore it is raining.

(2) John knows that it is raining, therefore it is raining.

(3) It is raining and there is a storm, therefore it is raining.

Not only does it adhere to the same standards, it also seems to me that the reasoning is good for analogous reasons: (3) follows on account of the concept of conjunction, (2) follows on account of the concept of knowledge, and (1) follows on account of the concept of actuality. (You can replace "the concept of ..." by the "the meaning of '...'" if this makes you more comfortable.) If logic is indeed about reasoning, then a logic of some notions should capture how we can reason *with the concepts it formalizes*. So just as a logic of conjunction must relate by the consequence relation what represents (3) (e.g. $\{p \wedge q\} \vDash p$) and a logic of knowledge must relate by the consequence relation what represents (2) (e.g. $\{K_j p\} \vDash p$), so a logic of actuality must relate by the consequence relation what represents (1) (e.g. $\{@p\} \vDash p$).

Of course, one might think that (1) is not sufficiently analogous to (2) and (3). Clearly, formal logic need not capture every pattern of reasoning that is good in some sense. So one might argue that although (1) is a good piece of reasoning, it is not good in the relevant sense, and therefore should not be licensed by *logic*. Such an argument might be possible, but it is quite unclear to me what the difference could be. So I take it that I have at least established that *prima facie*, real-world consequence is more plausible than general consequence.

### A.2.3 Monism and Pluralism

Some people are logical monists – they think that there is one consequence relation which should be modeled by logics. This has been an implicit assumption in my presentation so far. But there are also logical pluralists – people who think that there are several relations which are all equally admissible precisifications of the intuitive concept of consequence. Most prominently, logical pluralism has been defended by Beall and Restall (2006). For monists and pluralists, the discussion in this appendix will indicate different things.

I have argued that once we start being careful about what is modeled by a logic, the natural option which gives us general consequence (namely necessary truth-preservation) is problematic, and that some other natural options give us real-world consequence. Although there are further options that give us general consequence, we have additional reasons for preferring real-world consequence, as it licenses patterns of reasoning for which it seems that they should be licensed by logic.

For logical monists, this should provide some reasons for real-world consequence, or at the very least, show the need for good arguments for general consequence. On the other hand, for logical pluralists, although there may be some relations of consequence that are more adequately modeled by general consequence, I have presented good reasons to be interested in relations that are better modeled by real-world consequence.

In either case, I have given good reasons to work with real-world validity and consequence, and to take indexical truths like the ones formalized by $T_{@}$ to be logical truths. On the one hand, this shows that adopting $\mathsf{MD}$ and **2D** for the matrix semantics discussed in chapter 3 is the correct choice. On the other hand, it shows that the choice made in chapter 5 to adopt the logic **Act** of real-world validities is correct as well.

# Bibliography

C. Anthony Anderson. Toward a logic of a priori knowledge. *Philosophical Topics*, 21:1–20, 1993.

JC Beall and Greg Restall. *Logical Pluralism*. Oxford: Oxford University Press, 2006.

Patrick Blackburn and Maarten Marx. Remarks on Gregory's "actually" operator. *Journal of Philosophical Logic*, 31:281–288, 2002.

Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge: Cambridge University Press, 2001.

Alexander Chagrov and Michael Zakharyaschev. *Modal Logic*, volume 35 of *Oxford Logic Guides*. Oxford: Oxford University Press, 1997.

David J. Chalmers. *The Conscious Mind*. Oxford: Oxford University Press, 1996.

David J. Chalmers. Epistemic two-dimensional semantics. *Philosophical Studies*, 118:153–226, 2004.

David J. Chalmers. The foundations of two-dimensional semantics. In Manuel García-Carpintero and Josep Macià, editors, *Two-Dimensional Semantics*, pages 55–140. Oxford: Oxford University Press, 2006.

David J. Chalmers. Propositions and attitude ascriptions: A Fregean account. Forthcoming in *Noûs*, available online in advance of print. `doi:10.1111/j.1468-0068.2010.00788.x`, 2011.

John N. Crossley and Lloyd Humberstone. The logic of "actually". *Reports on Mathematical Logic*, 8:11–29, 1977.

Martin Davies. *Meaning, Quantification, Necessity*. London: Routledge & Kegan Paul, 1981.

Martin Davies. Reference, contingency, and the two-dimensional framework. *Philosophical Studies*, 118:83–131, 2004.

Martin Davies and Lloyd Humberstone. Two notions of necessity. *Philosophical Studies*, 38:1–30, 1980.

Martin Davies and Daniel Stoljar. Introduction. *Philosophical Studies*, 118: 1–10, 2004.

Josh Dever. Low-grade two-dimensionalism. *Philosophical Books*, 48:1–16, 2007.

Dorothy Edgington. The paradox of knowability. *Mind*, 94:557–568, 1985.

Dorothy Edgington. Two kinds of possibility. *Aristotelian Society Supplementary Volume*, 78:1–22, 2004.

John Etchemendy. Reflections on consequence. In Douglas Patterson, editor, *New Essays on Tarski and Philosophy*, pages 263–299. Oxford: Oxford University Press, 2008.

Gareth Evans. Reference and contingency. *Monist*, 62:161–189, 1979.

Kit Fine. Introduction. In *Modality and Tense*, pages 1–16. Oxford: Oxford University Press, 2005.

Gottlob Frege. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50, 1892.

D. M. Gabbay, A. Kurucz, F. Wolter, and M. Zakharyaschev. *Many-Dimensional Modal Logics: Theory and Applications*, volume 148 of *Studies in Logic and the Foundations of Mathematics*. Amsterdam: Elsevier, 2003.

Daniel Gallin. *Intensional and Higher-Order Modal Logic*. Amsterdam: North-Holland, 1975.

Dominic Gregory. Completeness and decidability results for some propositional modal logics containing "actually" operators. *Journal of Philosophical Logic*, 30:57–78, 2001.

Dominic Gregory. Iterated modalities, meaning and a priori knowledge. *Philosophers' Imprint*, 11:1–11, 2011.

Peter Gärdenfors. On the extensions of S5. *Notre Dame Journal of Formal Logic*, 14:277–280, 1973.

William H. Hanson. Actuality, necessity, and logical truth. *Philosophical Studies*, 130:437–459, 2006.

Allen Hazen. Expressive completeness in modal language. *Journal of Philosophical Logic*, 5:25–46, 1976.

Allen Hazen. The eliminability of the actuality operator in propositional modal logic. *Notre Dame Journal of Formal Logic*, 19:617–622, 1978.

Harold T. Hodes. Axioms for actuality. *Journal of Philosophical Logic*, 13: 27–34, 1984.

G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic*. London: Routledge, 1996.

Lloyd Humberstone. Two-dimensional adventures. *Philosophical Studies*, 118: 17–65, 2004.

Hans Kamp. Formal properties of 'now'. *Theoria*, 37:227–273, 1971.

David Kaplan. On the logic of demonstratives. *Journal of Philosophical Logic*, 8:81–98, 1978.

David Kaplan. Afterthoughts. In Joseph Almog, John Perry, and Howard Wettstein, editors, *Themes from Kaplan*, pages 565–614. Oxford: Oxford University Press, 1989a.

David Kaplan. Demonstratives. In Joseph Almog, John Perry, and Howard Wettstein, editors, *Themes from Kaplan*, pages 481–563. Oxford: Oxford University Press, 1989b.

Marcus Kracht. *Tools and Techniques in Modal Logic*, volume 142 of *Studies in Logic and the Foundations of Mathematics*. Amsterdam: Elsevier, 1999.

Marcus Kracht and Frank Wolter. Properties of independently axiomatizable bimodal logics. *Journal of Symbolic Logic*, 56:1469–1485, 1991.

Saul A. Kripke. *Naming and Necessity*. Cambridge, MA: Harvard University Press, 1980 [1972]. First published in *Semantics of Natural Language*, edited by Donald Davidson and Gilbert Harman, pages 253–355, 763–769, Dordrecht: D. Reidel, 1972.

David Lewis. Anselm and actuality. *Noûs*, 4:175–188, 1970.

David Lewis. *Counterfactuals*. Oxford: Basil Blackwell, 1973.

David Lewis. Attitudes *de dicto* and *de se*. *Philosophical Review*, 88:513–543, 1979.

David Lewis. Tharp's third theorem. *Analysis*, 62:95–97, 2002.

C. A. Meredith and A. N. Prior. Modal logic with functorial variables and a contingent constant. *Notre Dame Journal of Formal Logic*, 6:99–109, 1965.

Robert Michels. Soames's argument 1 against strong two-dimensionalism. Forthcoming in *Philosophical Studies*, available online in advance of print. `doi: 10.1007/s11098-011-9746-x`, 2011.

Richard Montague. Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. In *Proceedings of a Colloquium on Modal and Many-Valued Logics*, volume 16 of *Acta Philosophica Fennica*, pages 153–167. Helsinki: Societas Philosophica Fennica, 1963.

Michael Nelson and Edward N. Zalta. A defense of contingent logical truths. Forthcoming in *Philosophical Studies*, available online in advance of print. `doi: 10.1007/s11098-010-9624-y`, 2010.

Christopher Peacocke. What is a logical constant? *Journal of Philosophy*, 73: 221–240, 1976.

John Perry. The problem of the essential indexical. *Noûs*, 13:3–21, 1979.

Stefano Predelli. Utterance, interpretation and the logic of indexicals. *Mind & Language*, 13:400–414, 1998.

A. N. Prior. "Now". *Noûs*, 2:101–119, 1968a.

A. N. Prior. "Now" corrected and condensed. *Noûs*, 2:411–412, 1968b.

W. V. Quine. Three grades of modal involvement. In *The Ways of Paradox and Other Essays, Revised and Enlarged Edition*, pages 158–176. Cambridge, MA: Harvard University Press, 1976 [1953]. First published in *Proceedings of the XIth International Congress of Philosophy* 14, pages 65–81, 1953.

Wlodek Rabinowicz and Krister Segerberg. Actual truth, possible knowledge. *Topoi*, 13:101–115, 1994.

Greg Restall. Proofnets for S5: Sequents and circuits for modal logic. In Costas Dimitracopoulos, Ludomir Newelski, Dag Normann, and John R. Steel, editors, *Logic Colloquium 2005*, volume 28 of *Lecture Notes in Logic*, pages 151–172. Cambridge: Cambridge University Press, 2008.

Greg Restall. Truth values and proof theory. *Studia Logica*, 92:241–264, 2009.

Greg Restall. A cut-free sequent system for two-dimensional modal logic, and why it matters. Typescript, available at `http://consequently.org/papers/cfss2dml.pdf`, 2010.

Bertrand Russell. On denoting. *Mind*, 14:479–493, 1905.

Gillian Russell. One true logic? *Journal of Philosophical Logic*, 37:593–611, 2008.

Joe Salerno, editor. *New Essays on the Knowability Paradox*. Oxford: Oxford University Press, 2009.

Nathan Salmon. The logic of what might have been. *Philosophical Review*, 98: 3–34, 1989.

Laura Schroeter. Two-dimensional semantics. In Edward N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Winter 2010 edition, 2010.

Schiller Joe Scroggs. Extensions of the Lewis system S5. *Journal of Symbolic Logic*, 16:112–120, 1951.

Krister Segerberg. *An Essay in Classical Modal Logic*. Number 13 in Filosofiska Studier. Uppsala: Uppsala Universitet, 1971.

Krister Segerberg. Two-dimensional modal logic. *Journal of Philosophical Logic*, 2:77–96, 1973.

Stewart Shapiro. Logical consequence: Models and modality. In Matthias Schirn, editor, *The Philosophy of Mathematics Today*, pages 131–156. Oxford: Oxford University Press, 1998.

Stewart Shapiro. Logical consequence, proof theory, and model theory. In Stewart Shapiro, editor, *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 651–670. Oxford: Oxford University Press, 2005.

Scott Soames. *Reference and Description: The Case Against Two-Dimensionalism*. Princeton: Princeton University Press, 2005.

Scott Soames. Kripke, the necessary aposteriori, and the two-dimensionalist heresy. In Manuel García-Carpintero and Josep Macià, editors, *Two-Dimensional Semantics*, pages 272–292. Oxford: Oxford University Press, 2006.

Scott Soames. Actually. *Proceedings of the Aristotelian Society Supplementary Volume*, 81:251–277, 2007.

Robert C. Stalnaker. Assertion. In Peter Cole, editor, *Syntax and Semantics, Volume 9: Pragmatics*, pages 315–332. New York: Academic Press, 1978.

Alfred Tarski. On the concept of following logically. *History and Philosophy of Logic*, 23:155–196, 2002 [1936]. Originally published in Polish and German in 1936.

Leslie Tharp. Three theorems of metaphysics. *Synthese*, 81:207–214, 1989.

Frank Vlach. *'Now' and 'Then': A Formal Study in the Logic of Tense Anaphora.* PhD thesis, University of California, Los Angeles, 1973.

Timothy Williamson. The contingent a priori: Has it anything to do with indexicals? *Analysis*, 46:113–117, 1986.

Timothy Williamson. Indicative versus subjunctive conditionals, congruential versus non-hyperintensional contexts. *Philosophical Issues*, 16:310–333, 2006.

Edward N. Zalta. Logical and analytic truths that are not necessary. *Journal of Philosophy*, 85:57–74, 1988.

Lennart Åqvist. Modal logic with subjunctive conditionals and dispositional predicates. *Journal of Philosophical Logic*, 2:1–76, 1973.

# List of Notation

**Formulas**

| | | |
|---|---|---|
| $K_\Box$ | $\Box(p \to q) \to (\Box p \to \Box q)$ | page 12 |
| $N1$ | $Ap \to \Box Ap$ | page 22 |
| $N2$ | $\Box(Ap \to p)$ | page 22 |
| $N3/TP$ | $Ap \to \Box p$ | page 22/page 33 |
| $T_\Box$ | $\Box p \to p$ | page 23, Definition 3.5 |
| $5_\Box$ | $\Diamond p \to \Box \Diamond p$ | page 23, Definition 3.5 |
| $D_@$ | $@p \to \neg @ \neg p$ | page 23, Definition 3.5 |
| $D_{c@}$ | $\neg @ \neg p \to @p$ | page 23, Definition 3.5 |
| $I1$ | $\Box p \to @p$ | page 23, Definition 3.5 |
| $I2$ | $@p \to \Box @p$ | page 23, Definition 3.5 |
| $4_A$ | $Ap \to AAp$ | page 23, Definition 3.5 |
| $I3$ | $Ap \to @p$ | page 23, Definition 3.5 |
| $I4$ | $A(@p \to p)$ | page 23, Definition 3.5 |
| $HP$ | $\Box p \to Ap$ | page 33 |
| $KP^+$ | $\Box p \to \Box Ap$ | page 33 |
| $CP^+$ | $(\Box p \vee \Box \neg p) \to A(\Box p \vee \Box \neg p)$ | page 33 |
| $CP_1^+$ | $\Box p \to A(p \to \Box p)$ | page 34 |
| $I \triangle N$ | $p \to \triangle Np$ | page 62 |

Formulas whose names are indexed by a modality, like $5_\Box$, are to be read as schemas, where the modality can be replaced by another one. E.g., $5_A$ is $Cp \to ACp$. Further, I use the following convention: if $X$ denotes a formula $\varphi$, then $\ulcorner \Box X \urcorner$ denotes the formula $\Box\varphi$. E.g., $\Box D_A$ is $\Box(Ap \to Cp)$.