Modal syntactic predicates and their semantics: an analysis of semantic paradoxes by means of well- and ill-foundedness

Boaz Darius Laan

Pembroke College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Philosophy

October 2025

Abstract

In contemporary philosophical discourse modality plays a central theoretical role. One approach to formalising modalities is what this thesis calls the mention predicate approach, where a modality is formalised as a first-order predicate that predicates on names of sentences / formulae, or names of propositions. There is a relative lack of philosophical interest in the mention predicate approach because it is prone to paradox, most notably Montague's Paradox. This thesis argues why this is unwarranted; and critically assesses and contributes to topics relating to the mention predicate approach. Chapter 3 of this thesis critically assesses how Stern 2014c; Stern 2015 resolves the paradoxes of the mention predicate approach. Chapter 4 of this thesis contributes to the categorisation of the paradoxes that the mention predicate approach is prone to. Moreover, this chapter contributes to the understanding of the extent and limitations of the semantics of the mention predicate approach. In particular, chapter 4 provides a limitative result which functions as a categorisation theorem indicating when paradox arises. Chapter 5 of this thesis provides an example of a modality in the philosophical literature that is most appropriately formalised using the mention predicate approach. In particular, chapter 5 critically assesses the account of modal potentialism in the philosophy of mathematics developed by Linnebo 2018. In doing so, chapter 5 shows that, as it stands, Linnebo 2018 provides an account of modal potentialism that is incoherent.

Modal syntactic predicates and their semantics: an analysis of semantic paradoxes by means of well- and ill-foundedness

Boaz Darius Laan

Pembroke College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Philosophy

October 2025

This thesis is dedicated to

Hendrik Post
&
Ellen Theresia Maria Laan

Acknowledgements

I am deeply grateful for the support, encouragement, strength, and inspiration of all those who have helped me along the journey of my DPhil and my thesis, without whom it would not have been possible. I never walked alone. Thank you!

First and foremost, I am immensely thankful to my supervisors Volker Halbach and Timothy Williamson; for invaluable discussions and comments; and for your patience, mentorship, guidance, and motivation. You gave me the space to carve out my own ideas, always taking me seriously by being tremendously generous with your time and by treating my work with your scrutiny. I learned so much from you philosophically, professionally, and personally.

I am also very thankful to James Studd, Øystein Linnebo, and Chris Scambler for their critical engagement, guidance and enthusiasm; to my examiners Alexander Paseau and Johannes Stern for taking me seriously as a scholar and welcoming me into scholarship; and to my former supervisors Yakov Kremnitzer and Carlo Nicolai for their mentorship and encouragement.

For insightful comments discussing parts of my thesis I would like to thank Tim Button, Joel David Hamkins, Paul Gorbow, Daniel Isaacson, Miguel Lopez Munive, Mariona Miyata-Sturm, Quinn Crawford IV, Daniel Rowe, Pranciškus Gricius, James Glover, Elisabetta Sassarini, Dominik Ehrenfels, Marco Grossi, Milan Hartwig, Davide Sutto, Robin Solberg; and the audiences of the Oxford Philosophy DPhil seminar, the KCL Logic Work-In-Progress seminar, the *Challenging The Infinite* conference in Oxford, and the CFORS graduate conference in Oslo.

Oxford would have been but a collection of old and dusty buildings if it wasn't for the people inhabiting them. Quinn Crawford IV, Mariona Miyata-Sturm, Baskaran Sripathmanathan, Lawrence Wang, Miguel Lopez Munive, Juliana Choi, José Gabriel Niño Barreat, Efimia Panagiotaki, Fiammetta Rosenblatt, Tom Ferdinandt, Rachel Gardner, Avin Houro, Charles Pidgeon, Elisabetta Sassarini, James Glover, Pranciškus Gricius, Marco Grossi, Milan Hartwig, Thomas Bullemore, Dominik Ehrenfels, Daniel Rowe, Robin Solberg, Pieter Garicano, Marten Garicano; thank you! You provided the much needed tension in the fabric.

Oxford is not my only home. Thank you Alexandros Doganis, Luke Piper, Zeki Son, and Hyung-In Kim for making London a warm embrace of old around the corner, every single time.

To Ellen Laan, Christa van Wijnbergen, Paul ’t Hart, Enrico Perotti, and Luis Garicano. You are my academic role models, whose integrity and curiosity made me believe I could be one, too.

To Emma Turmaine, Yoshua Maas, Jasper Herzstein, Britt Waterman, and Leah Waterman. You taught me that there is more to family than just blood; and what it means to be there when you are not.

To my family, and my parents. Thank you, pap en mam, for your love. For holding my hand, every step of the way, as I carve out my own path.

To my partner, Fê. Thank you for being my anchor. For the length of your continent. For inviting me to see my inner poet, and then inviting me to look outwards. Where our rivers meet.

To my brother, Joontje, the albatross. Who would I be without you?

Abstract

In contemporary philosophical discourse modality plays a central theoretical role. One approach to formalising modalities is what this thesis calls the mention predicate approach, where a modality is formalised as a first-order predicate that predicates on names of sentences / formulae, or names of propositions. There is a relative lack of philosophical interest in the mention predicate approach because it is prone to paradox, most notably Montague's Paradox. This thesis argues why this is unwarranted; and critically assesses and contributes to topics relating to the mention predicate approach. Chapter 3 of this thesis critically assesses how Stern 2014c; Stern 2015 resolves the paradoxes of the mention predicate approach. Chapter 4 of this thesis contributes to the categorisation of the paradoxes that the mention predicate approach is prone to. Moreover, this chapter contributes to the understanding of the extent and limitations of the semantics of the mention predicate approach. In particular, chapter 4 provides a limitative result which functions as a categorisation theorem indicating when paradox arises. Chapter 5 of this thesis provides an example of a modality in the philosophical literature that is most appropriately formalised using the mention predicate approach. In particular, chapter 5 critically assesses the account of modal potentialism in the philosophy of mathematics developed by Linnebo 2018. In doing so, chapter 5 shows that, as it stands, Linnebo 2018 provides an account of modal potentialism that is incoherent.

Contents

1 Introduction

Contents

1.1 Modal statements

In contemporary philosophical discourse, and in particular in the discourse of contemporary philosophical and mathematical logic, modality plays a central theoretical role. For example, the literature on modal logic contains a wealth of rich technical mathematical results, and a wealth of rich philosophical discussion. From a broad philosophical perspective, modal statements involve modal auxiliary verbs such as 'can', 'could', or 'must'; and express what can be, what could be, or what must be the case. Typical everyday modal statements include the following:

A) I could have had light blond hair.

B) You cannot trample the breeding grounds of endangered bird species.

C) Spaceships cannot travel faster than the speed of light.

D) 'If P, then P' must be true.

E) Popular music cannot be underground.

F) '2' could have meant 3.

G) 2 cannot equal 3.

Modal statements can often be formulated in terms of what is possible, or what is necessary. For example, we can reformulate the sentences A)–G) to get:

A′) It is possible for me to have light blond hair.

B′) It is not possible to trample the breeding grounds of endangered bird species.

C′) It is not possible that spaceships travel faster than the speed of light.

D′) It is necessary that 'if P, then P' is true.

E′) It is not possible that popular music is underground.

F′) It is possible that '2' means 3.

G′) It is not possible that 2 is equal to 3.

It is here that we see just how widespread modal statements are in contemporary philosophical discourse. Almost all philosophical disciplines are concerned with what is 'possible', given widely distinct and philosophically significant understandings of 'possible'. For example, there are distinct philosophically significant interpretations of 'could', 'cannot', 'must', 'possible', and 'necessary' under which the sentences A) – G) and A′) – G′) are true; had my light blond hair not darkened when growing up, I would have had light blond hair; it is wrong to trample the breeding grounds of endangered birds; it is a law of nature that nothing travels faster than the speed of light; it is in virtue of the logical form of the sentence 'if P, then P', as understood by logicians, that it is true; it lies in the meaning of the phrase 'underground music' that underground music is not popular; the freedom we have when interpreting formal symbols or syntax allows for interpreting the symbol '2' with what 3 means; and 2 is distinct from 3 because I

have provided a formal proof that $2 \neq 3$. Furthermore, the interpretations of 'could', 'cannot', 'must', 'possible', and 'necessary' need to differ to make all of the sentences A) – G) and A′) – G′) true. That is, there is no single interpretation of these words for which the sentences A) – G) and A′) – G′) are true. For example, if we read 'cannot' / 'not possible' as indicating what the limitations set by the laws of nature are; then B) and B′) are false, whereas C) and C′) are true. There is nothing according to the laws of nature stopping you from trampling the breeding grounds of endangered bird species. On the other hand, if we read 'cannot' / 'not possible' as indicating the limitations set by some appropriate normative framework; then B) and B′) are true, whereas C) and C′) are false. Normative frameworks tend to have no bearing on the laws of nature. We call these differing interpretations of 'could', 'cannot', 'must', 'possible', and 'necessary' different kinds of 'modality'; and they show us that the truths of sentences A) – G) and A′) – G′) have different explanations. Moreover, these different kinds of modalities show us that any proposed systematic account of modality with which to understand modal statements such as A) – G) and A′) – G′) ideally accommodates a wide range of differing explanations.

The widespread occurrences and central theoretical roles of diverse modal statements, such as sentences A) – G) and A′) – G′), in (contemporary) philosophical discourse has led to and motivates the (ongoing) philosophical analysis of modality; in order to further our understanding of modality, and to provide a systematic account of modality. Much work has already been done towards this analysis. The seminal works of Putnam 1972 and Kripke 1980 were among the first to analyse modality, which argue that there are at least two 'fundamental' kinds of modality; metaphysical modality (necessity), and epistemic modality (a priority).[1] Since the works of Putnam and Kripke, a wealth of literature has developed discerning different kinds of modality; such as the seminal work of Kratzer 1977, who offers a unified account of the meaning of 'must' and 'can' in the philosophy

[1] Specifically, it was called into doubt whether all and only the apriori truths are necessary. For example, assume the uncontroversial direction of Leibniz' laws about identity that identical things are indiscernible, i.e. instantiate the same properties. Further, assume the uncontroversial claim that, necessarily furze = furze. Since both furze and gorse are distinct names for *Ulex*, a genus of flowering plant comprising of thorny evergreen shrubs, we have that furze = gorse. Applying the two assumptions, we have that, necessarily furze = gorse. However, furze = furze is an apriori truth; whereas furze = gorse is not.

of language / linguistics. Metaphysical modality, deontic modality, nomic modality, logical modality, temporal modality, doxastic modality, and epistemic modality are all commonplace in the contemporary literature on modality, to name a few. They consider, respectively, what is possible or necessary [such as A) and A′)]; what is permitted or obligatory [such as B) and B′)]; what is possible or necessary according to the laws of nature [such as C) and C′)]; what statements are possible or necessary according to the laws of logic [such as D) and D′)]; what at some point / always in the future (or past) is the case; what is believed; and what is known. By abuse of terms, modalities are often called necessities, such as metaphysical necessity and deontic necessity.[2]

In this thesis, I will specifically be focusing on modality as understood in the discourse of contemporary philosophical and mathematical logic, where it plays a central theoretical role when we attempt to formalise other philosophical disciplines, in particular their relevant modalities, in some logical, or mathematical, framework. This is often called modal logic, and there is a wealth of rich technical results, including proof-theoretic and model-theoretic results about consistency, inconsistency, soundness, and completeness. Much philosophical discussion regarding modal logic focusses on the connection between an intended philosophical theory being formalised; and a logical framework formalising it. This discussion does not just pertain to the aptness, or more generally to what is sought after in a logical framework, having fixed an intended philosophical theory to be formalised. It also pertains to critically assessing and possibly tweaking an intended philosophical theory being formalised by means of a fixed logical framework formalising it, including analysing the coherence or consistency of said intended philosophical theory. In particular, in this thesis I will understand modality as the possible philosophical interpretations of syntactic formations (either syntactically primitive or complex) in modal logic that are intended to be interpreted as modalities.[3] In the following section

[2]However, not all modalities are straightforwardly understood as kinds of necessities / possibilities. For example, it is unclear whether we can talk about doxastic necessity or possibility, or epistemic necessity or possibility.

[3]Of course, this is not a fixed or unchanging understanding of modality. As stated, intended philosophical theories and formal frameworks influence each other and evolve in response to each other. Given an intended philosophical theory, we might find some formal modal framework with which to formalise said theory. In turn, (one of) the syntactic modal formation(s) of this formal framework might have a possible philosophical interpretation slightly different from the initial intended philosophical theory. Further, this philosophical

1.2 we will see some of what these syntactic modal formations can be.

1.2 Formalising modality

We call the that-clauses of modal statements, such as 'spaceships travel faster than the speed of light' in C′), the complement expression, or complement for short, of these modal statements. Given this, in contemporary modal logic there are two main approaches to the syntax of formalising modalities and modal statements, i.e. two main choices for what the syntactic expressions that are intended to be interpreted as modalities are.[4]

To introduce these approaches, we need the following definitions regarding the syntax of well-formed formulae. Firstly, we call a (possibly complex) logical expression of a first-order logical language an *n*-adic *sentential operator* iff it takes *n* well-formed formulae as arguments and outputs a new more complex well-formed formula. Thus, for example the logical connectives such as negation and conjunction are sentential operators. Secondly, we call a logical expression of a first- or higher-order language an *n*-ary predicate iff it takes *n* singular terms as arguments and outputs an atomic formula. Thus, for example '=' (identity) is a predicate.

Let us now consider these two main choices for what the syntactic expressions that are intended to be interpreted as modalities are. Firstly, the *operator approach* takes a syntactic modal formation to be a *sentential operator*.[5] Fixing a modality and a modal statement involving it, standardly a modal sentential operator formalising said modality

interpretation might itself have an analogue or generalisation whose formalisation requires tweaking the initial formal modal framework, leading to a new formal modal framework; and so forth. Similarly, the understanding of modality in the philosophical and mathematical logic literature has changed and evolved since modality was first analysed. Rather, with understanding modality as the possible philosophical interpretations of syntactic formations (either syntactically primitive or complex) in modal logic that are intended to be interpreted as modalities, I intend to signal which half of this cycle I will focus on in this thesis.

[4]This does not preclude of course that there could be further approaches to the syntax of formalising modalities and modal statements that are distinct to either of these; just that the two to follow are the main approaches.

[5]Unlike the logical connectives however, modal sentential operators are given non-truth-functional semantics.

is a primitive syntactic formation that; takes as argument the complement of said modal statement; and outputs said modal statement itself.[6]

Secondly, the *predicate approach* takes a syntactic modal formation to be a *predicate*. Fixing a modality and a modal statement involving it, a modal predicate formalising said modality standardly; takes as argument a singular term that, once a semantics is introduced, refers to the complement of said modal statement; and outputs said modal statement itself.[7] The predicate approach knows two main sub-approaches, revolving around how we should understand that, once a semantics is introduced, this singular term refers to the complement of said modal statement. In particular, these two sub-approaches understand this singular term as referring to the complement of said modal statement, qua *named sentence*; or qua *interpreted sentence*. In more detail, the first sub-approach takes a syntactic modal formation to be a first-order predicate that predicates on *names* of sentences or formulae, or *names* of propositions (i.e. names of the complements of relevant modal statements).[8] [9] [10] The second sub-approach takes a syntactic modal formation to be a higher-order predicate (or operation) that predicates on *interpreted* sentences or formulae, or propositions / facts / events / states of affairs *themselves* (i.e. what the complements of relevant modal statements express).

[6]This is not always the case, nor an exclusive requirement; and whether a modal sentential operator is primitive; complex; a monadic operator; a dyadic operator; or an n-adic operator for some other natural number n; depends on other general considerations, and on particular considerations involving the specific modality being formalised.

[7]This is not always the case, nor an exclusive requirement; and whether a modal predicate is primitive; complex; a unary predicate; a binary predicate; or an n-ary predicate for some other natural number n; and whether, once a semantics is introduced, the singular term(s) refer(s) to (parts of) the complement of the relevant modal statement; or to an entity (entities) of another kind that is (are) in some appropriate sense (senses) related to the complement of the relevant modal statement; depends on other general considerations, and on particular considerations involving the specific modality being formalised.

[8]Here and in the rest of this thesis, unless specified otherwise, 'predicating on x' is intended to be understood as 'predicating on the singular term 'x' *itself*'; and thus not as 'predicating on whatever the singular term 'x' refers to'.

[9]A name of a sentence / formula refers to the sentence / formula *itself*, qua syntactic object, not what the sentence / formula expresses. In natural language, the name of a sentence or formula is often denoted using quotation marks, as in 'You cannot trample the breeding grounds of endangered bird species'. Similarly, the name of a proposition (such as the proposition that spaceships travel faster than the speed of light) refers to 'that' plus the that-clause *itself* (such as 'that spaceships travel faster than the speed of light'), not what 'that' plus the that-clause expresses.

[10]The arity of such a first-order modal predicate in particular also depends on the distinction between de dicto and de re modalities, which we will cover in more detail at the end of this chapter.

Let us now make these approaches a bit more precise, by considering some examples.

- Under the operator approach a modal statement of the form 'It is necessary that A' for some natural language sentence A is standardly formalised as $\Box A'$; where '$\Box$' is the modal sentential operator; and A' is a first-order formalisation of the sentence A.
- Under the first sub-approach of the predicate approach a modal statement of the form 'It is necessary that A' for some natural language sentence A is standardly formalised as $\mathbf{N}\overline{\ulcorner A' \urcorner}$; where '$\mathbf{N}$' is the first-order modal predicate; and $\overline{\ulcorner A' \urcorner}$ is the singular term denoting the name of a first-order formalisation A' of the sentence A.
- Under the second sub-approach of the predicate approach a modal statement of the form 'It is necessary that A' for some natural language sentence A is standardly formalised as NA'; where 'N' is the higher-order modal predicate of type $t \to t$; and A' is of the type t of sentences and a higher-order formalisation of the sentence A.

It is important to note that if a first-order sentential modal operator is formalised in a higher-order language, it will also be formalised as a higher-order modal predicate. For example, '$\Box$' will be formalised as a predicate of type $t \to t$, since it maps sentences to sentences. Thus, the operator and 'predicate of propositions' approaches are very tightly linked, and the distinction between them is whether one decides to work in a first-order language, or a higher-order language.

Right off the bat, we must first address and clarify an initial objection against the 'predicate of names of sentences approach'. This kind of objection is discussed by Lewy 1947 in the context of truth, and is known as Lewy's argument.[11] The worry is that $\mathbf{N}\overline{\ulcorner A' \urcorner}$ is itself not a necessary claim, since $\ulcorner A' \urcorner$ could have meant something else. For example, "2 + 2 = 4' is necessary' is itself not necessary, because '2 + 2 = 4' could for example have meant 3 + 3 = 5. This becomes an issue when necessity claims are compounded, which we

[11] Lewy was not the first to discuss kind of objection. For a history of this kind of objection see Halbach 2001.

make much use of in contemporary philosophy, and in this thesis in particular. Luckily, this objection can be diffused. Unpacking the objection further, there are two ways in which, for example, '2 + 2 = 4' could for example have meant 3 + 3 = 5. The first is that the meanings of our terms stay fixed, but that the name '2 + 2 = 4' does not rigidly refer to the sentence '2 + 2 = 4' qua syntactic object. Thus, the name '2 + 2 = 4' could have referred to the sentence '3 + 3 = 5', which means 3 + 3 = 5 if the meanings of our terms stay fixed. We can respond to this line of reasoning by stipulating that when we name a sentence we choose a name of said sentence that rigidly refers to said sentence; and that our quotation '·' does the same.

The second way in which, for example, '2 + 2 = 4' could for example have meant 3 + 3 = 5, is when the meanings of our terms do not stay fixed. In such a case, it does not matter whether the names of sentences rigidly refer or not; '2' could have meant 3, and '4' could have meant 5, in which case '2 + 2 = 4' could have meant 3 + 3 = 5. This is slightly more tricky. To respond to this line of reasoning, we need a way to ensure that the meaning of the names of sentences stay constant under shifts in the circumstances of evaluation. We have two options to ensure this constancy of meaning, both of which suffice for the purposes of this thesis. The first option is due to Gupta 1978; Gupta 1980, and the second to Peacocke 1978. Gupta 1978; Gupta 1980 suggests that we understand "2 + 2 = 4' is true' as asserting

'2 + 2 = 4' is true where '2 + 2 = 4' means what it actually means.

Peacocke 1978 suggests a similar, but slightly different understanding, where "2 + 2 = 4' is true' is taken to assert

'2 + 2 = 4' is true in $\mathcal{L}$

for some fixed language $\mathcal{L}$, such as for example English. If we adopt either of these understandings for necessity, both ensure that "2 + 2 = 4' is necessary' is necessary. The main takeaway of this discussion on Lewy's argument, is that in the rest of this thesis the names of sentences rigidly refer, and that the sentences they refer to mean what they

actually do. For more on Lewy's argument in the context of modality, see Stern 2015, §1.2.

It is now important to introduce a distinction between these three approaches to formalising modal statements (two syntactically distinct approaches with one having two semantically distinct sub-approaches). Firstly, observe that, once a semantics is introduced, each modal sentential operator is interpreted as taking as arguments *interpreted* sentences or formulae (the complements of relevant modal statements); and outputting other *interpreted* sentences or formulae (the modal statements themselves); due to the syntactic behaviour of sentential operators. Furthermore, observing how the operator and 'predicate of propositions' approach are very tightly linked, each higher-order modal predicate is interpreted as a property of interpreted sentences or formulae (the complements of relevant modal statements). Thus, we see that the operator approach and the 'predicate of propositions' approach formalise a modal statement by formulating the relevant modality such that it concerns whatever aspect of reality the complement of said modal statement concerns. In reference to the use-mention distinction, we call a modality formalised with one of these two approaches a 'use modality', because they use the complements of relevant modal statements. In particular, we call the 'predicate of propositions' approach the use predicate approach.

On the other hand, we see that the 'predicate of names of sentences' approach formalises a modal statement by formulating the relevant modality such that it mentions, as opposed to uses, the complement of said modal statement. That is, the relevant modality concerns the name of the complement of said modal statement, which is distinct from whatever aspect of reality the complement of said modal statement concerns. Similarly, we call a modality formalised with the 'predicate of names of sentences' approach a 'mention modality', and in particular we call this approach the mention predicate approach.[12]

[12]Note that the three approaches to formalising modal statements discussed in the previous paragraphs are the main approaches; however they are not the only approaches, and the analysis provided in the previous paragraphs is not an exclusive analysis of the approaches to formalising modal statements. One could for example also consider a first-order modal predicate of propositions; which would then be classified as a use modality. Thus, and by the operator and use predicate approaches, note that the use-mention distinction between modalities is different from the operator-predicate distinction between modalities.

I am aware that the use-mention distinction is difficult to articulate, and that it is debated whether this distinction is a strict one or not. In particular, there are many borderline cases where it is unclear whether some string of symbols / sentence is being used, mentioned, or both. Consider the famous example by Quine 1980, p. 139: (1) 'Giorgione was so-called because of his size'. Here, we are not merely using the name Giorgione, but also mentioning it. For if we were merely using the name Giorgione, we should be able to substitute 'Giorgione' with 'Barbarelli', being names of the same person, to get: (1′) 'Barbarelli was so-called because of his size'. However, (1) is true, whereas (1′) is false. Another class of examples that are borderline cases between use and mention involve the phenomenon called 'mixed quotation'. Consider for example (2) 'Quine said that quotation 'has a certain anomalous feature''. According to some authors, such as Davidson 1979, (2) simultaneously uses 'has a certain anomalous feature' to say what Quine said; and mentions it to say which words Quine used in saying what he said. However, for other authors cases such as (2) involving mixed quotation are merely apparent borderline cases. Some authors, such as Recanati 2001, argue that mixed quotation is not a genuine semantic phenomenon at all. In particular, they argue that (2) only uses 'has a certain anomalous feature', and thus has the same semantic content as (2′) 'Quine said that quotation has a certain anomalous feature'. Not all authors go so far, and they argue that cases such as (2) involving mixed quotation are merely apparent borderline cases, even though the quotation marks in mixed quotation make a semantic contribution. For example, Cappelen and Lepore 2007 argue that (2) only mentions 'has a certain anomalous feature'.

Nonetheless, the formal constructions *themselves* of the operator, use predicate, and mention predicate approach are not such borderline cases; the former two clearly use, whereas the latter clearly mentions the complement of modal statements they occur in. Rather, the fact that the use-mention distinction is not a strict one makes it clear that it is sometimes a difficult choice which approach to use when formalising modal statements; precisely because all three approaches, whether explicitly or implicitly, rely on such a strict distinction between use and mention. This difficulty must clearly be

addressed. However, it lies beyond the scope of this thesis, and will be set aside for future work.

Furthermore, I am aware that introducing the terminology of 'use modality' and 'mention modality', in particular that of 'use modality', is a realist way of distinguishing the three approaches to formalising modal statements; we distinguish between a modality concerning whatever aspect of reality the complement of a relevant modal statement concerns; and a modality concerning the name of the complement of a relevant modal statement. This is realist in the sense that it implicitly assumes the existence of an independent / non-representational reality. Compare this with anti-realism, which denies either the independence / non-representationality of this reality, or its existence. This does not preclude the anti-realist from making a distinction similar to use and mention modalities; they can instead piggyback on whichever anti-realist way they prefer to make sense of the use-mention distinction. If one outright denies the use-mention distinction, one at the very least needs to provide some clear explanation for this denial. In such a case, the analysis and formalisation of modal statements must be informed by this explanation; and these could look quite different to those provided in this thesis and the literature. For example, the distinction between the use and mention predicate approaches could blur, or even dissolve, depending on the explanation given.

As stated, the operator approach, the mention predicate approach, and the use predicate approach are the three main approaches to formalising modalities and modal statements in contemporary modal logic. However, these three approaches are by no means equally supported; most philosophers are directly aware of the operator approach, only indirectly aware of the use predicate approach, and unaware of the mention predicate approach. Firstly, the operator approach is by far the most widely applied and investigated. Proponents include Montague 1963, Mulligan 2010, Prior 1971, Recanati 2000, and Williamson 2013, to name a few. Formal expositions of the operator approach were first given by the axiomatic treatments of C.I. C. I. Lewis 1912; C. I. Lewis 1914,

C. I. Lewis and Langford 1932, and MacColl 1906.[13] They all developed accounts of 'strict implication', aiming at avoiding the classical validities

$$(\phi \rightarrow \psi) \vee (\psi \rightarrow \phi) \tag{1.1}$$

$$\psi \rightarrow (\phi \rightarrow \psi) \tag{1.2}$$

$$\neg\phi \rightarrow (\phi \rightarrow \psi) \tag{1.3}$$

known as the Paradoxes of Material Implication. Instead of formalising implication using the standard truth-functional sentential operator '$\rightarrow$', i.e. the material conditional, C.I. Lewis proposed formalising implication using a non-truth-functional dyadic sentential operator '$\dashv\!\!\exists$' (whenever ϕ and ψ are well-formed formulae, then so is $\phi \dashv\!\!\exists \psi$), which he called strict implication. A monadic modal sentential operator '$\Box$' could then be defined as

$$\Box\phi := (\top \dashv\!\!\exists \phi)$$

C.I. Lewis[14] and others developed different calculi for strict implication, and thus the modal sentential operator, in subsequent years, which saw their application to epistemic and deontic modalities. Since then, the focus has shifted from taking strict implication as primitive to taking the modal sentential operator as primitive. Strict implication '$\dashv\!\!\exists$' can then be defined using the non-truth-functional modal sentential operator '$\Box$' and the truth-functional material conditional '$\rightarrow$' as

$$\phi \dashv\!\!\exists \psi := \Box(\phi \rightarrow \psi)$$

The operator approach[15] was further developed by the works of Hintikka 1957a; Hintikka 1957b; Kanger 1957; and in particular Kripke 1959; Kripke 1963a; Kripke 1963b;

[13] It is unclear whether these initial expositions were aware of the difference between operators and predicates: '[F]rom a historic perspective the distinction between use and mention only became clear during the 1930s subsequent to the work of Gödel, Tarski, and, especially, Carnap and Quine. But if no distinction is made between using sentences and mentioning them, then the distinction between operators and predicates of sentences will evidently be blurred' (Stern 2015, p. 24).

[14] C.I. Lewis is standardly taken to be the first to introduce a modal sentential operator. However, MacColl's exposition came 6 years earlier than C.I. Lewis', and according to Rahman and Redmond 2008 the introduction of strict implication is better attributed to MacColl.

[15] For the rest of this thesis, reference to the operator approach should be understood as taking the non-truth-functional monadic modal sentential operator as primitive as opposed to the non-truth-functional dyadic strict implication operator, unless stated otherwise.

Kripke 1965 with his development of possible-world semantics, providing soundness and completeness results with the proof-theoretic perspective analysed thus far.[16] In-depth formal expositions of the operator approach can be found in any textbook on modal logic, such as Blackburn et al. 2001, Chellas 1980, Hughes and Cresswell 1996, or Sider 2010.[17]

The use predicate approach has a rich history in modern modal logic and has been enjoying a recent resurgence, gaining much traction. It has deep connections with modal logicism, which too has recently gained much traction, where modalities are reduced to higher-order logic. Proponents include Bacon 2018 and Stalnaker 2023. An in-depth formal exposition of the use predicate approach can be found in Bacon 2023.[18]

The mention predicate approach likewise has a rich history in modern modal logic, stemming from criticisms from Carnap 1934 and Quine 1943; Quine 1947; Quine 1953 of C.I. Lewis and MacColl. There is much commonality between the mention predicate approach, and the rich literature on axiomatic syntactic theories of truth which treat truth as a predicate of names of sentences. Similarly to the use predicate approach, the mention predicate approach has been enjoying a recent resurgence; with active debate of the respective benefits and drawbacks between on the one hand proponents of the mention predicate approach and on the other proponents of the operator and use predicate approaches. Proponents of the mention predicate approach include Carnap 1934, Halbach 2021, Halbach and Leigh 2024, Quine, throughout his work, and Stern 2015. A formal exposition of the mention predicate approach can be found in Stern 2015.[19] For an

[16]Possible-world semantics has since been widely used as a semantics for modal logic, but possible worlds have been discussed in the philosophical literature for much longer, notably by Gottfried Leibniz (1951). Copeland 2002 provides an overview of the development of possible worlds in the period leading up to Kripke. Possible worlds were already discussed by Carnap 1946 and Meredith and Prior 1996 among others, influenced by the works of Charles Sanders Peirce (1994), and by Ludwig Wittgenstein in his *Tractatus* (2014).

[17]These historical notes are by no means meant to be exhaustive; for more on the development of modern modal logic, in particular the operator approach, see Goldblatt 2003; and for more on modalities before modern analytic philosophy and modern logic see M. Kneale and W. Kneale 1962.

[18]For more on the use of higher-order logic in (the development of) modern modal logic see Williamson 2013, §5.

[19]For more on the development of the mention predicate approach see Stern 2015, §2.1.

introduction to, and formal exposition of axiomatic syntactic theories of truth see Halbach 2014.

One of the benefits of the mention predicate approach that is often brought up when argued in favour of it, is its expressive richness, in particular its relative expressive richness to the operator and use predicate approaches. For example, the mention predicate approach allows for the formalisation of certain quantified statements involving modality that are widespread in philosophical discourse, such as 'Every analytic truth is necessary'. This statement is not formalisable in the operator or use predicate approaches; whereas it is formalisable in the mention predicate approach, by e.g. $\forall x(\mathbf{AT}(x) \rightarrow \mathbf{N}(x))$. However, an upshot of the way in which the mention predicate approach is expressively rich is that it is prone to paradox. There are many paradoxes in the mention predicate approach, the most notable of which is called Montague's Paradox. Kaplan and Montague 1960 and Montague 1963 showed that, akin to the Liar Paradox and Tarski's Undefinability of Truth, assuming weak and arguably natural modal principles in the mention predicate approach is inconsistent. The mention predicate approach has since fallen by the wayside; many philosophers have taken these paradoxes, in particular Montague's Paradox, to show that the mention predicate approach is ill-suited for formalising modal statements. However, this conclusion is too quick, and the moral of these paradoxes is more nuanced. As with all paradoxes, the paradoxes of the mention predicate approach must be addressed, and concessions must be made. Nevertheless, the mention predicate approach as a whole need not be be abandoned. For example, Stern 2014c; Stern 2015 shows that in the mention predicate approach we can consistently preserve many modal principles also available to the operator approach, if we make certain concessions on the structure of the names of sentences / names of propositions that the modal predicate predicates on. Stern's goal is 'to rebut Montague's assessment that virtually all of modal logic must be sacrificed, if we treat modalities syntactically [c.f. Montague 1963] — at least, if this assessment is understood in its straightforward, general way' (Stern 2014c, p. 561).

Furthermore, there is a methodological reason for not abandoning the mention predicate approach as a whole. The methodological approach of formalising philosophical notions

(in logic) has been much employed and defended, in particular by analytic philosophers. There are many benefits to this approach, the chief one being able to 'sharpen' philosophical notions by formalising them. There are many ways this sharpening can occur. For example, say we formalise some philosophical framework in a certain formal framework. We could then use the clarity of said formal framework to derive results or consequences, which can then be interpreted by, or translated back into, said philosophical framework. Often, these formal results or consequences are fruitful, surprising, or unwanted; and much easier to derive than trying to derive their philosophical interpretations / translations informally.[20] Moreover, if said formal framework is inconsistent, then something is wrong with the philosophical framework being formalised; and the root of the inconsistency points us towards the root of what is wrong with the philosophical framework.

To successfully sharpen philosophical notions by formalising them, however, requires us to 'successfully' / 'most appropriately' formalise them. That is, given a philosophical notion, we must choose an 'appropriate' / the 'most appropriate' formal framework to formalise said philosophical notion in. Given a philosophical framework, this will, among other things, consist in discerning its relevant structural features under consideration, and then abstracting said structural features by finding or constructing a formal framework that has strictly those structural features. I do not intend in this thesis to provide an account of what it precisely takes to choose an appropriate formal framework. Rather, the similarities of relevant structural features between the philosophical and formal frameworks under consideration is a necessary condition when choosing an appropriate formal framework.

In the case of formalising modalities, of the three approaches used in the literature, one approach mentions, and the other two approaches use the complements of modal statements being formalised. This distinction is a formally relevant one, precisely because of how structurally different the formal frameworks end up being depending on which

[20]This is a much-used approach employed by mathematicians. Say we have a mathematical theory $T1$ we know (very) little about; but we also have a mathematical theory $T2$ we know (very) much about, along with an invertible structure preserving mapping h between $T1$ and $T2$. We can then learn a lot about $T1$, by first going from $T1$ to $T2$ using h, then deriving some known result R, and finally going back to $T1$ using h^{-1}. This gives us a result R', which is strictly confined to the theory $T1$.

approach is used, such as susceptibility to paradox. The question then becomes whether it is closest to the intended interpretation of the modality being formalised that it be formalised as using, or mentioning the complements of modal statements it occurs in. In the case of metaphysical modality and nomic modality, many philosophers tend to think / argue that it is closest to their intended interpretations that they be formalised as using the complements of modal statements they occur in. Thus, we have a methodological reason for not abandoning the operator or use predicate approaches. As we shall see in chapter 5, I will argue that it is closest to the intended interpretation of the modality employed in the account of modal potentialism developed by Linnebo 2018 that it be formalised as mentioning the complements of modal statements it occurs in, despite the susceptibility to paradox. Thus, we have a methodological reason for not abandoning the mention predicate approach as a whole as a response to paradox.

1.3 The aim of this thesis

All in all, there is a relative lack of philosophical interest in the mention predicate approach, despite there being modalities in the literature that are most appropriately formalised using this approach. Therefore, in this thesis I will critically assess and contribute to topics relating to the mention predicate approach.

Chapter 2 of this thesis will provide the necessary logical background to be able to do this critical assessment. We will focus on formalising modalities as first-order unary predicates that predicate on names of sentences, and we will consider Peano Arithmetic as the theory of syntax we work in. We will then prove the central Diagonal Lemma and the Uniform Diagonal Lemma. Finally, we will use these Lemmas to prove Tarski's Undefinability of Truth, the Liar Paradox, and ultimately Montague's Paradox.

Chapter 3 of this thesis will critically assess how Stern 2014c; Stern 2015 resolves the paradoxes of the mention predicate approach. In particular, chapter 3 will argue that the framework of Stern 2014c; Stern 2015 does not provide a conception of the mention predicate approach. However, chapter 3 will also suggest in what tentative cases we

can nonetheless use the framework of Stern 2014c; Stern 2015, thereby providing a constructive takeaway.

Chapter 4 of this thesis will contribute to the categorisation of the paradoxes that the mention predicate approach is prone to; and to the understanding of the extent and limitations of the semantics of the mention predicate approach. In particular, chapter 4 will consider more paradoxes besides Montague's Paradox that the mention predicate approach is prone to. Furthermore, chapter 4 will generalise the classical possible-world predicate-semantics developed by Halbach and Leigh 2024 to multiple distinct first-order modal predicates and to two dimensional semantics. Finally, chapter 4 will provide a limitative result which will function as a categorisation theorem indicating when paradox arises in this model theory. As a result, chapter 4 will further our understanding of paradoxes involving multiple distinct first-order modal predicates, which have so far received little attention in the literature.

Chapter 5 of this thesis will provide an example of a modality in the philosophical literature that is most appropriately formalised using the mention predicate approach.[21] In particular, chapter 5 will critically assess the account of modal potentialism in the philosophy of mathematics developed by Linnebo 2018, focussing on the modality this account employs. In doing so, chapter 5 will show that, as it stands, Linnebo 2018 provides an account of modal potentialism that is incoherent due to the paradoxes the mention predicate approach is prone to. More generally, chapter 5 provides yet another warning of the risks of paradox, in particular what goes wrong if one does not use the most appropriate formalisation when formalising philosophical notions.

1.4 De dicto and de re modal statements

Before getting into the technical details of this thesis, it is important to address in a bit more detail precisely what kind of predicate the modal predicate is in the mention predicate approach. In particular, it is important to address whether it is a unary predicate, like a truth predicate; a binary predicate, like a satisfaction predicate; or an n-ary predicate

[21] This chapter has been published in a shortened version in Laan 2025.

for some other natural number n. In part, this choice will depend on the intended interpretation of the specific modality in question that is being formalised. However, with more generality, this choice will also depend on the distinction between what are called *de dicto* and *de re* modal statements. Recall that the mention predicate approach formalises a modality as a predicate that predicates on the names of the complements of modal statements involving said modality. De dicto modal statements then ascribe properties to names of closed sentences, relative to modalities; whereas de re modal statements ascribe properties to names of open formulae relative to names of sequences of objects relating to the free variables of said open formula, relative to modalities. In other words, the complement of a de dicto modal statement is a closed sentence; whereas the complement of a de re modal statement is an open formula.[22] Consider for example the two modal statements

H) Necessarily, Socrates is a Greek philosopher and a man.

I) Socrates is a Greek philosopher and necessarily a man.

To make the that-clauses more explicit, these two modal statements can be reformulated as

H′) It is necessary that Socrates is a Greek philosopher and a man.

I′) Socrates is a Greek philosopher. It is necessary that he is a man.

To make precise that the complement of H′) is a closed sentence and of I′) is an open formula, we can think of H′) and I′) as follows. H′) states that the named sentence [Socrates is a Greek philosopher and a man] has the property *x is necessary*; whereas I′) states that Socrates has the properties *x is a Greek philosopher*; and *x is a man is necessary*. Thus, H) is a de dicto modal statement, and I) is a de re modal statement.

[22]There used to be quite some discussion regarding the comprehensibility of de re modal statements, with the most notable critic being Quine (see for instance Quine 1953). As we shall see, in current philosophy the distinction between de dicto and de re modal statements is still very important for the mention predicate approach. However, for the operator and use predicate approaches this distinction has become less important in current philosophy.

In current philosophy this distinction is very important for the mention predicate approach. In the case of de dicto modal statements, it is the most intuitive and straightforward to formalise a modality occurring in a de dicto modal statement as a unary predicate that predicates on names of closed sentences or names of propositions. This is precisely how de dicto modal statements are formalised in the literature on the mention predicate approach (see for instance Stern 2015 or Halbach and Leigh 2024). However, the question of formalisation is less straightforward in the case of de re modal statements. Consider the name of an open formula; for example, 'x is a man'. When we name an open formula, we are not using the free variables involved. Instead, we are merely mentioning them. In particular, no named variables are interpreted when provided with a variable assignment. In other words, we cannot 'quantify into' unary predicates which predicate on names of open formulae. Rather, as Halbach 2021, Halbach and Leigh 2024, §3.5, and Stern 2015, §1.3; and further back Quine 1956; Quine 1977 suggest, de re modal statements seem to be best formalised using a binary satisfaction predicate that predicates on pairs consisting of names of formulae and names of variable assignments.[23] For reasons of technical simplicity, if we are only considering variable assignments for a finite number n of variables, de re modal statements seem to be best formalised using an $n + 1$-ary satisfaction predicate. Formalising modalities using a satisfaction predicate also allows us to formalise de dicto modal statements, if we also allow for names of closed formulae, i.e. sentences.

Most of the literature on the mention predicate approach, such as Stern 2015 and Halbach and Leigh 2024, focusses on a unary modal predicate, thus restricting its analysis of modality to de dicto modal statements. Though 'many questions [regarding modal satisfaction predicates] are still open and modal metaphysics could benefit from considering languages with great expressive power and predicates for de re modalities' (Halbach and Leigh 2024, p. 47); such authors that focus on unary modal predicates all allude to the possibility of generalising their frameworks for a satisfaction predicate, thus

[23] See Halbach 2021 for an account of such a binary modal satisfaction predicate, which covers some of the challenges that arise; including what choices there are when developing possible-world predicate-semantics regarding variable assignments that assign contingent objects to variables that do not exist at the world of evaluation.

accounting for de re modal statements. Following the literature, in this thesis modalities will be formalised using a unary predicate, unless stated otherwise.

2

Logical Background

Contents

To be able to critically assess and contribute to topics relating to the mention predicate approach, we must first provide the necessary logical background. Recall that the mention predicate approach formalises modalities as first-order predicates that predicate on names of sentences or names of propositions. In this thesis we will focus on multiple first-order predicates that predicate on names of sentences, as opposed to names of propositions, since this is the standard approach in the literature on the mention predicate approach.[1] Thus, in this chapter we will consider the theory of syntax we will work in, to make intelligible that the modal predicates apply to names of sentences. In particular, following

[1] Nothing much hinges on predicating over names of sentences, as opposed to names of propositions; so long as the names of propositions are sufficiently structured. How much structure is precisely needed will be further addressed later in this chapter. Note that this structure of the class of names of propositions depends on an account of what it is to be a proposition; but it is irrespective of an account of the identity conditions for propositions. For example, '$P \vee \neg P$' and '$P \rightarrow P$' are clearly different names, irrespective of whether the propositions $P \vee \neg P$ and $P \rightarrow P$ are identical, as coarse-grained theories of propositions would have it; or whether they are not identical, as more fine-grained theories of propositions would have it.

the broader literature we will work in Peano Arithmetic. We will consider unary modal predicates, akin to unary truth predicates; as opposed to binary modal predicates, akin to satisfaction predicates, or n-ary modal predicates.[2] To emphasise that these modal predicates apply to syntactic entities (namely, names of sentences), we will call them modal syntactic predicates. Furthermore, we will prove the Diagonal Lemma and the Uniform Diagonal Lemma (to be specified below), providing us with self-referential expressions in our syntax theory; which will play the central role in the (limitative) results we will prove in chapters 4 and 5. Finally, we will use these Lemmas to prove paradoxes relating to truth and modality, providing limitative results for their behaviour; Tarski's Undefinability of Truth, the Liar Paradox, and ultimately Montague's Paradox.

2.1 Logical pre-requisites

Let us first briefly consider the logical pre-requisites of this thesis.

A first-order logical language $\mathcal{L} := \{P_i\}_{i\in I} \cup \{f_j\}_{j\in J}$ consists of non-logical predicate symbols P_i indexed by I and non-logical function symbols f_j indexed by J. We will also allow 0-place predicate symbols, which will act as propositional sentence letters; and 0-place function symbols, which will act as constant symbols. We will often write $\mathcal{L} = \{P_i\}_{i\in I} \cup \{p_l\}_{l\in L} \cup \{f_j\}_{j\in J} \cup \{c_k\}_{k\in K}$ to indicate that $\{p_l\}_{l\in L} \subseteq \{P_i\}_{i\in I}$ are the propositional sentence letters; and $\{c_k\}_{k\in K} \subseteq \{f_j\}_{j\in J}$ are the constant symbols. We say that a language $\mathcal{L}_2$ expands the language $\mathcal{L}_1$ iff $\mathcal{L}_1 \subseteq \mathcal{L}_2$; and we say that $\mathcal{L}_2$ expands $\mathcal{L}_1$ with a set of symbols S iff $\mathcal{L}_1 \cap S \neq \emptyset$ and $\mathcal{L}_1 \cup S = \mathcal{L}_2$. We also then say that $\mathcal{L}_1$ reduces $\mathcal{L}_2$. Furthermore, we have countably many variables, denoted by x, y, z, x_1, x_2, etc.; and for the logical constants we use $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\forall$, $\exists$, and $=$.[3]

We define the terms, closed terms, sentences / expressions, formulae, and positive formulae of $\mathcal{L}$ in the usual way. Furthermore, the complexity of a (positive) formula is

[2] As discussed earlier, we will thus only account for de dicto modal statements. Due to limitations of space, the possibility of generalising the framework developed in this thesis for a satisfaction predicate, so as to account for de re modal statements, is left for future work.

[3] For our purposes it does not make a difference whether all of the logical constants are treated as primitive, or some of them are treated as abbreviations based on a truth-functionally complete set of logical constants.

also defined in the usual way. We denote individual terms and closed terms by s, t, s_1, s_2, etc.; and we denote individual sentences / expressions and formulae by $\phi, \psi, \lambda, \phi_1, \phi_2$, etc. Thus, the free and bound terms of a formula ϕ; and the notion of a term s being free for term t in ϕ; are defined in the usual way. We denote these syntactic categories of $\mathcal{L}$ by $\mathsf{Term}(\mathcal{L})$, $\mathsf{CTerm}(\mathcal{L})$, $\mathsf{Sent}(\mathcal{L})$ / $\mathsf{Exp}(\mathcal{L})$, and $\mathsf{Form}(\mathcal{L})$ respectively. We also denote all strings of symbols of $\mathcal{L}$ by $\mathsf{String}(\mathcal{L})$, and individual strings of symbols by e, l, e_1, e_2, etc.

We denote the set of all variables by Var; and we denote the set of free variables of a formula ϕ by $\mathsf{FV}(\phi)$. If $\{x_1, \ldots, x_n\} = \mathsf{FV}(\phi)$, for $n \in \mathbb{N}$, we abbreviate ϕ by $\phi(x_1, \ldots, x_n)$; or when it is clear from context what $x_1, \ldots, x_n$ are, by $\phi(\vec{x})$ for short. We also abbreviate $\forall x_1 \ldots \forall x_n \phi(x_1, \ldots, x_n)$ by $\forall \vec{x} \phi(\vec{x})$ for short. Moreover, we let $\phi(s/t)$ abbreviate the formula obtained by, whenever t is free and s is free for t in ϕ, substituting t by s. We abbreviate $(\ldots(\phi(s_1/t_1))\ldots)(s_m/t_m)$ by $\phi(s_1/t_1, \ldots, s_m/t_m)$; or when it is clear from context what $s_1, \ldots s_m$ and $t_1, \ldots, t_m$ are, by $\phi(\vec{s}/\vec{t})$ for short. If it avoids confusion, we also often drop '$\vec{t}$' and write $\phi(\vec{s})$.

Unless stated otherwise, in this thesis we work in Classical logic, both proof- and model-theoretically. Regarding the model-theoretic semantics, for a first-order logical language $\mathcal{L} = \{P_i\}_{i \in I} \cup \{p_l\}_{l \in L} \cup \{f_j\}_{j \in J} \cup \{c_k\}_{k \in K}$ an $\mathcal{L}$-model $\mathcal{M}$ is of the form

$$\mathcal{M} := \langle D; \{P_i^{\mathcal{M}}\}_{i \in I}; \{p_l^{\mathcal{M}}\}_{l \in L}; \{f_j^{\mathcal{M}}\}_{j \in J}; \{c_k^{\mathcal{M}}\}_{k \in K} \rangle \tag{2.1}$$

where

- $D \neq \emptyset$ is called the domain of $\mathcal{M}$.
- $P^{\mathcal{M}} \subseteq D^n$ is an n-ary relation for each n-ary predicate symbol P in $\mathcal{L}$, when $n > 0$.
- $p^{\mathcal{M}} \subseteq \{\emptyset\}$ for each 0-place predicate symbol p in $\mathcal{L}$; where $\{\emptyset\}$ is identified with 1, and $\emptyset$ with 0.
- $f^{\mathcal{M}} : D^k \to D$ is a k-ary function, i.e. $f^{\mathcal{M}} \subseteq D^{k+1}$, for each k-ary function symbol f in $\mathcal{L}$, when $k > 0$.
- $c^{\mathcal{M}} \in D$ is an element for each 0-place function symbol c in $\mathcal{L}$.

Note that $\mathcal{M}$ implicitly specifies the language $\mathcal{L}$; and thus we drop $\mathcal{L}$ and call $\mathcal{M}$ a model. We say that a model $\mathcal{M}_2 = \langle D_2; \{P_i^{\mathcal{M}_2}\}_{i\in I}; \{p_l^{\mathcal{M}_2}\}_{l\in L}; \{f_j^{\mathcal{M}_2}\}_{j\in J}; \{c_k^{\mathcal{M}_2}\}_{k\in K}\rangle$ *extends* a model $\mathcal{M}_1 = \langle D_1; \{P_i^{\mathcal{M}_1}\}_{i\in I}; \{p_l^{\mathcal{M}_1}\}_{l\in L}; \{f_j^{\mathcal{M}_1}\}_{j\in J}; \{c_k^{\mathcal{M}_1}\}_{k\in K}\rangle$ iff $D_1 \subseteq D_2$; $P_i^{\mathcal{M}_1} \subseteq P_i^{\mathcal{M}_2}$ for each $i \in I$; $p_l^{\mathcal{M}_1} = p_l^{\mathcal{M}_2}$ for each $l \in L$; $f_j^{\mathcal{M}_1} \subseteq f_j^{\mathcal{M}_2}$ for each $j \in J$; and $c_k^{\mathcal{M}_1} = c_k^{\mathcal{M}_2}$ for each $k \in K$. We say that a model $\mathcal{M}_2 = \langle D_2; \{P_i^{\mathcal{M}_2}\}_{i\in I_2}; \{p_l^{\mathcal{M}_2}\}_{l\in L_2}; \{f_j^{\mathcal{M}_2}\}_{j\in J_2}; \{c_k^{\mathcal{M}_2}\}_{k\in K_2}\rangle$ *expands* a model $\mathcal{M}_1 = \langle D_2; \{P_i^{\mathcal{M}_2}\}_{i\in I_1}; \{p_l^{\mathcal{M}_2}\}_{l\in L_1}; \{f_j^{\mathcal{M}_2}\}_{j\in J_1}; \{c_k^{\mathcal{M}_2}\}_{k\in K_1}\rangle$ iff $I_1 \subseteq I_2$; $L_1 \subseteq L_2$; $J_1 \subseteq J_2$; and $K_1 \subseteq K_2$. We also then say that $\mathcal{M}_1$ reduces $\mathcal{M}_2$.

Given a model $\mathcal{M}$, we call a function $g \subseteq \mathsf{Var} \times D$ a variable assignment on $\mathcal{M}$. We denote by g_u^x the variable assignment which maps x to u, and agrees with the variable assignment g on all other variables. Now, given an $\mathcal{L}$-model $\mathcal{M}$ and a variable assignment g on $\mathcal{M}$ we interpret the terms of $\mathcal{L}$ inductively as follows, which we denote by $[\cdot]_{\langle\mathcal{M},g\rangle}$:

- Atomic terms:
  - Given a variable x, we define $[x]_{\langle\mathcal{M},g\rangle}$ as $g(x)$.
  - Given a 0-place function symbol $c \in \mathcal{L}$, we define $[c]_{\langle\mathcal{M},g\rangle}$ as $c^{\mathcal{M}}$.
- Complex terms:
  - Given a k-ary function symbol f in $\mathcal{L}$, we define $[f(t_1, \ldots, t_k)]_{\langle\mathcal{M},g\rangle}$ as $f^{\mathcal{M}}([t_1]_{\langle\mathcal{M},g\rangle}, \ldots, [t_k]_{\langle\mathcal{M},g\rangle})$.

We interpret formulae of $\mathcal{L}$ with respect to $\mathcal{M}$ and g inductively as follows, which we denote by $\mathrm{V}_{\langle\mathcal{M},g\rangle}(\cdot)$:

- Atomic formulae:
  - $\mathrm{V}_{\langle\mathcal{M},g\rangle}(t_1 = t_2) = 1$ iff $[t_1]_{\langle\mathcal{M},g\rangle} = [t_2]_{\langle\mathcal{M},g\rangle}$.
  - $\mathrm{V}_{\langle\mathcal{M},g\rangle}(Pt_1 \ldots t_n) = 1$ iff $\langle [t_1]_{\langle\mathcal{M},g\rangle}, \ldots, [t_n]_{\langle\mathcal{M},g\rangle}\rangle \in P^{\mathcal{M}}$.
- Complex formulae:[4]
  - $\mathrm{V}_{\langle\mathcal{M},g\rangle}(\neg\phi) = 1$ iff $\mathrm{V}_{\langle\mathcal{M},g\rangle}(\phi) = 0$.

[4]There might be less clauses needed, depending on whether some of the logical constants are treated as abbreviations based on a truth-functionally complete set of logical constants.

- $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\phi \wedge \psi) = 1$ iff $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\phi) = 1$ and $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\psi) = 1$.
- $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\phi \vee \psi) = 1$ iff $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\phi) = 1$ or $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\psi) = 1$.
- $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\phi \rightarrow \psi) = 1$ iff $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\phi) = 0$ or $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\psi) = 1$.
- $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\phi \leftrightarrow \psi) = 1$ iff $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\phi) = \mathrm{V}_{\langle \mathcal{M},g\rangle}(\psi)$.
- $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\forall x\phi) = 1$ iff for all $u \in D$ we have that $\mathrm{V}_{\langle \mathcal{M},g^x_u\rangle}(\phi) = 1$.
- $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\exists x\phi) = 1$ iff for some $u \in D$ we have that $\mathrm{V}_{\langle \mathcal{M},g^x_u\rangle}(\phi) = 1$.

We say that a formula ϕ is valid in $\mathcal{M}$, denoted $\mathcal{M} \vDash \phi$, iff $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\phi) = 1$ for all variable assignments g. We say that ϕ is valid or logically true, denoted $\vDash \phi$, iff $\mathcal{M} \vDash \phi$ for every model $\mathcal{M}$. We say that ϕ is a logical consequence of a set of formulae Γ iff, whenever $\mathcal{M}$ is a model and g is a variable assignment such that $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\gamma) = 1$ for all $\gamma \in \Gamma$, we have that $\mathrm{V}_{\langle \mathcal{M},g\rangle}(\phi) = 1$.

Regarding the proof theory, I will not specify any particular Classical proof theory to be used, except for that it has to be sound and complete with respect to the above model-theoretic semantics, to cater to as many possible backgrounds; I will let the reader use the logical calculus they are comfortable with. At the end of chapter 5 we will also proof-theoretically work in the weaker Intuitionistic logic. Likewise, I will not specify any particular Intuitionistic proof theory to be used.[5] A formal proof of a formula ϕ' from a set of formulae S' will be given by an informal sketch with reference to the important ideas, and denoted by $S' \vdash \phi'$.

We say that $S \subseteq \mathsf{Sent}(\mathcal{L})$ is inconsistent iff for all $\phi \in \mathsf{Sent}(\mathcal{L})$ we have that $S \vdash \phi$; and consistent iff it is not inconsistent, i.e. iff there is some $\phi \in \mathsf{Sent}(\mathcal{L})$ such that $S \nvdash \phi$. Both for Classical logic and Intuitionistic logic, it can be shown that S is inconsistent iff *there is some* $\phi \in \mathsf{Sent}(\mathcal{L})$ such that $S \vdash \phi$ and $S \vdash \neg\phi$; and thus that S is consistent iff *for all* $\phi \in \mathsf{Sent}(\mathcal{L})$ we have that $S \nvdash \phi$ or $S \nvdash \neg\phi$. We say that S is maximally consistent iff it is consistent and for all $\phi \in \mathsf{Sent}(\mathcal{L})$ either $\phi \in S$ or $\neg\phi \in S$. We define the *theory of* S as $\mathsf{Th}_{\mathcal{L}}(S) := \{\phi \in \mathsf{Sent}(\mathcal{L}) \mid S \vdash \phi\}$. Thus, S is inconsistent iff $\mathsf{Th}_{\mathcal{L}}(S) = \mathsf{Sent}(\mathcal{L})$;

[5] For an example of an Intuitionistic proof theory, and how to strengthen it to a Classical proof theory, see the natural deduction systems of Troelstra and van Dalen 1988, §2.

and consistent iff $\mathsf{Th}_{\mathcal{L}}(S) \subsetneq \mathsf{Sent}(\mathcal{L})$. When it is clear from context what $\mathcal{L}$ is, we write $\mathsf{Th}(S)$. Note that the operation $\mathsf{Th}_{\mathcal{L}}(\cdot) : \mathcal{P}[\mathsf{Sent}(\mathcal{L})] \to \mathcal{P}[\mathsf{Sent}(\mathcal{L})]$ is a *closure operator*, letting $\mathcal{P}[\mathsf{Sent}(\mathcal{L})]$ denote the power set of $\mathsf{Sent}(\mathcal{L})$. That is, for any $S \subseteq \mathsf{Sent}(\mathcal{L})$ we have that

$$S \subseteq \mathsf{Th}(S) \qquad (\textit{Extensive})$$

$$S \subseteq S' \Rightarrow \mathsf{Th}(S) \subseteq \mathsf{Th}(S') \qquad (\textit{Monotonic})$$

$$\mathsf{Th}(\mathsf{Th}(S)) = \mathsf{Th}(S) \qquad (\textit{Idempotent})$$

Moreover, because proofs are of finite length, $\mathsf{Th}_{\mathcal{L}}(\cdot)$ is an *algebraic* closure operator. That is, for any $S \subseteq \mathsf{Sent}(\mathcal{L})$ we have that

$$\mathsf{Th}(S) = \bigcup_{S' \in \mathsf{Fin}(S)} \mathsf{Th}(S') \qquad (\textit{Algebraic})$$

where $\mathsf{Fin}(S) := \{S' \subseteq S \mid S' \text{ is finite}\}$.

We say that $\mathsf{E} \subseteq \mathsf{Sent}(\mathcal{L})$ is deductively closed, and we call E a theory, iff $\mathsf{E} \vdash \phi \Rightarrow \phi \in \mathsf{E}$. We say that the theory E_2 *extends* the theory E_1 iff E_1 and E_2 are formulated in some $\mathcal{L}$, and $\mathsf{E}_1 \subseteq \mathsf{E}_2$. We say that E_2 extends E_1 with some $S \subseteq \mathsf{Sent}(\mathcal{L})$ iff E_2 extends E_1 and $\mathsf{E}_2 = \mathsf{Th}(\mathsf{E}_1 \cup S)$. We say that the theory E_2 *expands* the theory E_1 iff E_1 is formulated in some $\mathcal{L}$, E_2 is formulated in some extension $\mathcal{L} \subseteq \mathcal{L}'$, and $\mathsf{Th}_{\mathcal{L}'}(\mathsf{E}_1) \subseteq \mathsf{E}_2$. We say that E_2 expands E_1 with some $S \subseteq \mathsf{Sent}(\mathcal{L}')$ iff E_2 expands E_1 and $\mathsf{E}_2 = \mathsf{Th}_{\mathcal{L}'}(\mathsf{E}_1 \cup S)$. We also then say that E_1 reduces E_2. Thus, if a theory E_2 extends a theory E_1, we also have that E_2 expands E_1. Finally, we say that a theory E contains some $S \subseteq \mathsf{Sent}(\mathcal{L})$ iff E is formulated in some extension $\mathcal{L} \subseteq \mathcal{L}'$ and $S \subseteq \mathsf{E}$.

2.2 Syntax theory

Let us now (briefly) consider the theory of syntax we work in. As the seminal work of Gödel 1931 first showed, syntax theory can be formalised in arithmetic, also called the arithmetisation of syntax theory. Thus, and following the broader literature on the mention predicate approach (and axiomatic syntactic theories of truth), we work in arithmetic as our theory of syntax. Formal expositions of the arithmetisation of syntax theory can be

found in graduate textbooks on mathematical logic, such as Boolos et al. 2007 or Monk 1976; more in-depth in Feferman 1960 or Hájek and Pudlák 1993; or in textbooks on axiomatic syntactic theories of truth, such as Halbach 2014.

To be able to work in arithmetic as our theory of syntax, we must first make arithmetic precise.

Definition 2.2.1 (The Language of Peano Arithmetic)**.** The language of Peano Arithmetic is a first-order language with identity, which we denote $\mathcal{L}_{\mathsf{PA}}$. The logical constants of $\mathcal{L}_{\mathsf{PA}}$ are $\neg$, $\wedge$, $\forall$, and $=$; with $\phi \vee \psi$; $\phi \rightarrow \psi$; $\phi \leftrightarrow \psi$; $\exists x \phi$; and $s \neq t$ standing in the standard way as abbreviations for $\neg(\neg\phi \wedge \neg\psi)$; $\neg(\phi \wedge \neg\psi)$; $\neg(\phi \wedge \neg\psi) \wedge \neg(\neg\phi \wedge \psi)$; $\neg\forall x \neg\phi$; and $\neg s = t$ respectively. Furthermore, $\mathcal{L}_{\mathsf{PA}}$ contains the constant symbol $\underline{0}$; the one-place function symbol S; and the two-place function symbols $+$ and $\times$. For the convenience of the results we will prove, we also add finitely many other function symbols to $\mathcal{L}_{\mathsf{PA}}$.[6]

Definition 2.2.2 (Peano Arithmetic)**.** Peano Arithmetic, denoted PA, is the theory of the following axioms in $\mathcal{L}_{\mathsf{PA}}$:

$$\forall x\, \mathsf{S}(x) \neq \underline{0} \tag{PA1}$$

$$\forall x \forall y (\mathsf{S}(x) = \mathsf{S}(y) \rightarrow x = y) \tag{PA2}$$

$$\forall x\, x + \underline{0} = x \tag{PA3}$$

$$\forall x\, x + \mathsf{S}(y) = \mathsf{S}(x + y) \tag{PA4}$$

$$\forall x\, x \times \underline{0} = \underline{0} \tag{PA5}$$

$$\forall x \forall y\, x \times \mathsf{S}(y) = (x \times y) + x \tag{PA6}$$

Axioms defining the remaining function symbols (PAF)

and the axiom schema

$$(\phi(\underline{0}) \wedge \forall x(\phi(x) \rightarrow \phi(\mathsf{S}(x))) \rightarrow \forall x \phi(x) \tag{Ind}$$

for $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}})$.

[6] We use these extra function symbols to represent primitive recursive functions. In particular, we add function symbols for what is minimally needed to do syntax theory in the object language.

We call the model $\mathbb{N}$ of PA the standard model of arithmetic; which has as domain ω (we will often use ω and $\mathbb{N}$ interchangeably), and interprets the constant symbol $\underline{0}$ as $0 \in \omega$, the function symbol S as the successor function in ω, the function symbol $+$ as the addition function in ω, and the function symbol $\times$ as the multiplication function in ω. That is, $\mathbb{N}$ is commonly taken to be the intended interpretation of the arithmetic vocabulary.

Another theory worth mentioning is Robinson Arithmetic.

Definition 2.2.3 (Robinson Arithmetic)**.** Robinson Arithmetic, denoted Q[7], is formulated in $\mathcal{L}_{\mathsf{PA}}$ and is the theory of the axioms $\mathsf{PA}1$- $\mathsf{PA}\mathrm{F}$ and the axiom

$$\forall x(x \neq \underline{0} \rightarrow \exists z\, \mathsf{S}(z) = x) \tag{Q1}$$

Q is interesting because it is a finitely axiomatisable fragment of PA that is much weaker than PA, where nonetheless Gödel's Incompleteness Theorems hold. In particular, syntax theory can be formalised in Q, showing that the induction schema is not necessary for the arithmetisation of syntax theory.[8] Nonetheless, following the broader literature on the mention predicate approach (and axiomatic syntactic theories of truth), we work in PA as our theory of syntax.[9]

Now that we have specified PA, let us consider how to formalise syntax theory in PA. Denote any language $\mathcal{L}$ countably expanding $\mathcal{L}_{\mathsf{PA}}$ by $\mathcal{L}^{+}_{\mathsf{PA}}$; denote the axioms of PA formulated in $\mathcal{L}^{+}_{\mathsf{PA}}$ by $\mathsf{PA}_{\mathcal{L}^{+}_{\mathsf{PA}}}$; and denote any theory extending $\mathsf{PA}_{\mathcal{L}^{+}_{\mathsf{PA}}}$ by $\mathsf{PA}^{+}_{\mathcal{L}^{+}_{\mathsf{PA}}}$. Then, as first shown by Gödel 1931, we can code $\mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$ in $\mathbb{N}$.[10] These codes are called Gödel numbers; and for $e \in \mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$ we denote the Gödel number of e by $\ulcorner e \urcorner$. Another important notion is that of the numeral of $n \in \mathbb{N}$, which is the expression given by n

[7] To avoid confusion, note that Boolos et al. 2007 denote Robinson Arithmetic by R; and denote by Q a theory slightly different to Robinson Arithmetic.

[8] The axioms of Q are the minimal axioms needed to prove that every recursive function is representable in PA.

[9] We could have also worked in any other theory of syntax that proves the appropriate analogues to the Diagonal Lemma and the Uniform Diagonal Lemma (to be specified below).

[10] There are many acceptable ways to code $\mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$ in $\mathbb{N}$. For details see the literature introduced at the beginning of this section. In particular, for examples of specific codings see Boolos et al. 2007, p. 188 and Cellucci 2022, §5.

applications of the function symbol S to the constant symbol $\underline{0}$, and abbreviated by $\overline{n}$. Taken together, the numeral of the Gödel number of some $e \in \mathsf{String}(\mathcal{L}^+_{\mathsf{PA}})$, abbreviated by $\overline{\ulcorner e \urcorner}$, functions as the name of the string e in the object theory $\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}}$. This will allow us to 'talk about' the expressions of (an extension of) PA within PA itself (and thus within said extension of PA). To make this 'talking about' precise, we require the following definition.

Definition 2.2.4 (Representability)**.** A k-ary function $f \subseteq \mathbb{N}^{k+1}$ is representable in PA iff there exists some $\phi(x_1, \ldots, x_k, y) \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}})$ such that for all $n_1, \ldots, n_k \in \mathbb{N}$

$$\mathsf{PA} \vdash \forall x(\phi(\overline{n_1}, \ldots, \overline{n_k}, x) \leftrightarrow x = \overline{f(n_1, \ldots, n_k)}) \tag{2.2}$$

A k-ary function constant $\underset{\bullet}{f}$ represents a k-ary function f in PA iff for all $n_1, \ldots, n_k \in \mathbb{N}$

$$\mathsf{PA} \vdash \underset{\bullet}{f}(\overline{n_1}, \ldots, \overline{n_k}) = \overline{f(n_1, \ldots, n_k)} \tag{2.3}$$

A k-ary relation $R \subseteq \mathbb{N}^k$ is weakly representable (or recursively enumerable) in PA iff there exists some $\phi(x_1, \ldots, x_k) \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}})$ such that for all $n_1, \ldots, n_k \in \mathbb{N}$

$$\langle n_1, \ldots, n_k \rangle \in R \Leftrightarrow \mathsf{PA} \vdash \phi(\overline{n_1}, \ldots, \overline{n_k}) \tag{2.4}$$

R is strongly representable in PA iff moreover

$$\langle n_1, \ldots, n_k \rangle \notin R \Rightarrow \mathsf{PA} \vdash \neg\phi(\overline{n_1}, \ldots, \overline{n_k}) \tag{2.5}$$

Note that a set $A \subseteq \mathbb{N}$ is a 1-ary relation, and thus weakly / strongly representable iff (2.4) / (2.4) & (2.5) hold(s).

Theorem 2.2.1. *Every recursive function is representable in* PA*. Furthermore,* $R \subseteq \mathbb{N}^k$ *is recursive iff* R *is strongly representable in* PA*.*

For a proof see the literature introduced at the beginning of this section. We say that a set of strings S is weakly / strongly representable iff $\ulcorner S \urcorner := \{\ulcorner e \urcorner \mid e \in S\} \subseteq \mathbb{N}$ is weakly / strongly representable. Then, for example, $\mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$, $\mathsf{Sent}(\mathcal{L}^+_{\mathsf{PA}})$ / $\mathsf{Exp}(\mathcal{L}^+_{\mathsf{PA}})$, $\mathsf{Term}(\mathcal{L}^+_{\mathsf{PA}})$, $\mathsf{CTerm}(\mathcal{L}^+_{\mathsf{PA}})$, and $\mathsf{String}(\mathcal{L}^+_{\mathsf{PA}})$ are primitive recursive, and thus representable in PA. Likewise, the set of axioms of $\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}}$ itself is primitive recursive, and thus

representable in PA. Furthermore, the standard syntactic operations on $\mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$ are primitive recursive, and thus representable in PA. The standard syntactic operations include the following:

I. The negation function; which takes as arguments the Gödel number of any $e \in \mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$; and outputs the Gödel number of the string which concatenates e on the end of '$\neg$'.

II. The conjunction function; which takes as arguments the Gödel numbers of any two $e_1, e_2 \in \mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$; and outputs the Gödel number of the string which first concatenates e_2 on the end of '$\wedge$', call it e_3, and then concatenates e_3 on the end of e_1.

III. The universal quantification function; which takes as arguments (i) the Gödel number of any $e \in \mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$ and (ii) the Gödel number of any variable α; and outputs the Gödel number of the string which first concatenates e on the end of α, call it e', and then concatenates e' on the end of '$\forall$'.

IV. The binary concatenation function; which takes as arguments the Gödel numbers of any two $e_1, e_2 \in \mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$; and outputs the Gödel number of the string which concatenates e_2 on the end of e_1.

V. The quotation function; which takes as argument the Gödel number of any $e \in \mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$; and outputs the Gödel number of the numeral of the Gödel number of e, i.e. $\ulcorner\overline{\ulcorner e \urcorner}\urcorner$.

VI. The ternary substitution function; which takes as arguments (i) the Gödel number of any $e \in \mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$, (ii) the Gödel number of any $l_1 \in \mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$ of length 1 occurring in e, and (iii) the Gödel number of any $e_1 \in \mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$; and outputs the Gödel number of the string which substitutes every occurrence of l_1 with e_1 in e.

We assume that we have chosen a 'natural' coding on $\mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$; and 'natural' representations of the preceding syntactic categories and operations; which will allow us

to prove statements of syntax theory in $\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}}$.[11] Later in this section we will briefly consider an example given by Feferman 1960 to illustrate why this 'naturalness' condition is important to set.

Furthermore, for convenience we assume that $\mathcal{L}_{\mathsf{PA}}$ contains unary function symbols representing (the characteristic functions of) $\mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$, $\mathsf{Sent}(\mathcal{L}^+_{\mathsf{PA}})$ / $\mathsf{Exp}(\mathcal{L}^+_{\mathsf{PA}})$, $\mathsf{Term}(\mathcal{L}^+_{\mathsf{PA}})$, $\mathsf{CTerm}(\mathcal{L}^+_{\mathsf{PA}})$, and $\mathsf{String}(\mathcal{L}^+_{\mathsf{PA}})$ for any of the extensions $\mathcal{L}^+_{\mathsf{PA}}$ of $\mathcal{L}_{\mathsf{PA}}$ that we will consider in this thesis; denoted by $\dot{\mathsf{Form}}(\mathcal{L}^+_{\mathsf{PA}})(\cdot)$, $\dot{\mathsf{Sent}}(\mathcal{L}^+_{\mathsf{PA}})(\cdot)$ / $\dot{\mathsf{Exp}}(\mathcal{L}^+_{\mathsf{PA}})(\cdot)$, $\dot{\mathsf{Term}}(\mathcal{L}^+_{\mathsf{PA}})(\cdot)$, $\dot{\mathsf{CTerm}}(\mathcal{L}^+_{\mathsf{PA}})(\cdot)$, and $\dot{\mathsf{String}}(\mathcal{L}^+_{\mathsf{PA}})(\cdot)$ respectively. Likewise, for convenience we assume that $\mathcal{L}_{\mathsf{PA}}$ contains function symbols representing the syntactic operations I-VI. That is, we assume that $\mathcal{L}_{\mathsf{PA}}$ contains function symbols representing; the negation function, denoted by $\dot{\neg}(\cdot)$; the conjunction function, denoted by $\dot{\wedge}(\cdot,\cdot)$; the universal quantification function, denoted by $\dot{\forall}(\cdot,\cdot)$; the binary concatenation function, denoted by $\dot{\frown}(\cdot,\cdot)$; the quotation function, denoted by $\dot{\mathsf{q}}(\cdot)$; and the ternary substitution function, denoted by $\dot{\mathsf{sub}}(\cdot,\cdot,\cdot)$.[12] We also introduce the following abbreviations for convenience:

- For $e \in \mathsf{String}(\mathcal{L}^+_{\mathsf{PA}})$, we let $\dot{\mathsf{dia}}(\overline{\ulcorner e \urcorner})$ abbreviate $\dot{\mathsf{sub}}(\overline{\ulcorner e \urcorner}, \overline{\ulcorner x \urcorner}, \dot{\mathsf{q}}(\overline{\ulcorner e \urcorner}))$.
- For $e, l_1, e_1, \ldots, l_k, e_k \in \mathsf{String}(\mathcal{L}^+_{\mathsf{PA}})$, we let $\dot{\mathsf{sub}}_k(\overline{\ulcorner e \urcorner}, \overline{\ulcorner l_1 \urcorner}, \overline{\ulcorner e_1 \urcorner}, \ldots, \overline{\ulcorner l_k \urcorner}, \overline{\ulcorner e_k \urcorner})$ inductively abbreviate:
  - $\dot{\mathsf{sub}}(\overline{\ulcorner e \urcorner}, \overline{\ulcorner l_1 \urcorner}, \overline{\ulcorner e_1 \urcorner})$, for $k = 1$.
  - $\dot{\mathsf{sub}}(\overline{\ulcorner \dot{\mathsf{sub}}_{j-1}(\overline{\ulcorner e \urcorner}, \overline{\ulcorner l_1 \urcorner}, \overline{\ulcorner e_1 \urcorner}, \ldots, \overline{\ulcorner l_{j-1} \urcorner}, \overline{\ulcorner e_{j-1} \urcorner})\urcorner}, \overline{\ulcorner l_j \urcorner}, \overline{\ulcorner e_j \urcorner})$ for $k = j > 1$.

  This lets us generalise the ternary substitution function to perform k-many substitutions at the same time.
- Assume $\phi(x_1, \ldots, x_k) \in \mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$ has no bound occurrences of $x_1, \ldots, x_k$. Then we let $\overline{\ulcorner \phi(\dot{\overline{x_1}}, \ldots, \dot{\overline{x_k}})\urcorner}$ abbreviate $\dot{\mathsf{sub}}_k(\overline{\ulcorner \phi(x_1, \ldots, x_k)\urcorner}, \overline{\ulcorner x_1 \urcorner}, \dot{\mathsf{q}}(\overline{x_1}), \ldots, \overline{\ulcorner x_k \urcorner}, \dot{\mathsf{q}}(\overline{x_k}))$. We also abbreviate this with $\overline{\ulcorner \phi(\dot{\vec{x}})\urcorner}$ for short.

[11] For more on 'natural' codings and representations see Halbach 2014, p. 33-35. For more on coding into non-standard models of arithmetic, and more generally non-standard theories of syntax that use such a coding, see Kaye 1991.

[12] That is, any $\mathcal{L}_{\mathsf{PA}}$-model, including $\mathbb{N}$, is restricted to interpreting these unary function symbols as the characteristic functions they represent.

We now have sufficient tools to prove, as promised, the Diagonal Lemma and the Uniform Diagonal Lemma.

Lemma 2.2.1 (Strong Diagonal Lemma)**.** *Assume $\phi(x) \in \mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$ has no bound occurrences of x. Then there is a closed term $t_{\phi(x)}$ such that*

$$\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \vdash t_{\phi(x)} = \overline{\ulcorner \phi(t_{\phi(x)}) \urcorner} \tag{DT}$$

Proof. Let $t_{\phi(x)}$ be given by $\dot{\mathsf{dia}}(\overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner})$. We then have that

$$\begin{aligned}
\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \vdash \dot{\mathsf{dia}}(\overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner}) &= \dot{\mathsf{sub}}(\overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner}, \overline{\ulcorner x \urcorner}, \dot{\mathsf{q}}(\overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner})) && (2.6)\\
&= \dot{\mathsf{sub}}(\overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner}, \overline{\ulcorner x \urcorner}, \overline{\ulcorner \overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner} \urcorner}) && (2.7)\\
&= \overline{\ulcorner \phi(\dot{\mathsf{dia}}(\overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner})) \urcorner} && (2.8)
\end{aligned}$$

Line (2.6) is an identity statement of the form $s = s$, since by definition $\dot{\mathsf{dia}}(\overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner})$ abbreviates the term $\dot{\mathsf{sub}}(\overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner}, \overline{\ulcorner x \urcorner}, \dot{\mathsf{q}}(\overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner}))$; line (2.7) holds because $\dot{\mathsf{q}}$ represents the quotation function; and line (2.8) holds because $\dot{\mathsf{sub}}$ represents the ternary substitution function, and because $\phi(x)$ has no bound occurrences of x. Q.E.D.

Lemma 2.2.2 (Diagonal Lemma)**.** *Assume $\phi(x) \in \mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$ has no bound occurrences of x. Then there is a sentence λ such that*

$$\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \vdash \lambda \leftrightarrow \phi(\overline{\ulcorner \lambda \urcorner}) \tag{2.9}$$

Proof. By Lemma 2.2.1 there is a closed term $t_{\phi(x)}$ given by $\dot{\mathsf{dia}}(\overline{\ulcorner \phi(\dot{\mathsf{dia}}(x)) \urcorner})$ such that

$$\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \vdash t_{\phi(x)} = \overline{\ulcorner \phi(t_{\phi(x)}) \urcorner} \tag{DT}$$

Let λ be given by $\phi(t_{\phi(x)})$. By the laws of identity, we then have that

$$\begin{aligned}
\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \vdash \lambda &\leftrightarrow \phi(t_{\phi(x)}) && (2.10)\\
&\leftrightarrow \phi(\overline{\ulcorner \phi(t_{\phi(x)}) \urcorner}) && (2.11)\\
&\leftrightarrow \phi(\overline{\ulcorner \lambda \urcorner}) && (2.12)
\end{aligned}$$

Q.E.D.

If the only function symbols of $\mathcal{L}_{\mathsf{PA}}$ were the arithmetic $\underline{0}$, S, $+$, and $\times$; then $\dot{\mathsf{dia}}(\overline{\ulcorner\phi(\dot{\mathsf{dia}}(x))\urcorner})$ would not be a closed term of $\mathcal{L}_{\mathsf{PA}}$; and thus (DT) would not be an identity statement proper. In particular, the Strong Diagonal Lemma would be false. However, recall that the ternary substitution function sub is primitive recursive, and thus representable in PA. In particular, dia is representable in PA by some $\psi(x, y) \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}})$. Now fix $\phi(x)$ as in the Diagonal Lemma, and let $\psi_\phi(x) \coloneqq \exists y(\psi(x, y) \wedge \phi(y))$. Letting $\lambda \coloneqq \psi_\phi(\overline{\ulcorner\psi_\phi(x)\urcorner})$, using the fact that $\psi(x, y)$ represents dia we have that $\mathsf{PA} \vdash \lambda \leftrightarrow \phi(\overline{\ulcorner\lambda\urcorner})$. That is, the Diagonal Lemma would still hold.

The Diagonal Lemma can be generalised to any formula ϕ with $k \in \mathbb{N}$ free variables.

Theorem 2.2.2 (K-ary Diagonal Lemma)**.** *Assume $\phi(x, y_1 \ldots, y_{k-1}) \in \mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$ has no bound occurrences of x. Then there is a $\lambda(\vec{y}) \in \mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$ such that*

$$\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \vdash \forall\vec{y}(\lambda(\vec{y}) \leftrightarrow \phi(\overline{\ulcorner\lambda(\vec{y})\urcorner}, \vec{y})) \tag{2.13}$$

The proof is the same as the proof of the Diagonal Lemma. Note that $\overline{\ulcorner\lambda(\vec{y})\urcorner}$ is a numeral, and thus the variables $\vec{y}$ do not occur in it, nor can be bound in it. For present purposes, we will also strengthen the Diagonal Lemma so that we *can* bind occurrences of $\vec{y}$ 'inside' $\overline{\ulcorner\lambda(\vec{y})\urcorner}$ by using $\overline{\ulcorner\lambda(\dot{\vec{y}})\urcorner}$.

Theorem 2.2.3 (Uniform Diagonal Lemma)**.** *Assume $\phi(x, y_1 \ldots, y_{k-1}) \in \mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$ has no bound occurrences of x. Then there is a $\lambda(\vec{y}) \in \mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$ such that*

$$\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \vdash \forall\vec{y}(\lambda(\vec{y}) \leftrightarrow \phi(\overline{\ulcorner\lambda(\dot{\vec{y}})\urcorner}, \vec{y})) \tag{2.14}$$

Proof. Consider $\phi(\dot{\mathsf{sub}}_{k-1}(x, \overline{\ulcorner y_1\urcorner}, \dot{\mathsf{q}}(\overline{y_1}), \ldots, \overline{\ulcorner y_{k-1}\urcorner}, \dot{\mathsf{q}}(\overline{y_{k-1}})), y_1, \ldots, y_{k-1}) \in \mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$. By the K-ary Diagonal Lemma we can find a $\lambda(\vec{y}) \in \mathsf{Form}(\mathcal{L}^+_{\mathsf{PA}})$ such that

$$\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \vdash \forall\vec{y}(\lambda(\vec{y}) \leftrightarrow \phi(\dot{\mathsf{sub}}_{k-1}(\overline{\ulcorner\lambda(\vec{y})\urcorner}, \overline{\ulcorner y_1\urcorner}, \dot{\mathsf{q}}(\overline{y_1}), \ldots, \overline{\ulcorner y_{k-1}\urcorner}, \dot{\mathsf{q}}(\overline{y_{k-1}})), \vec{y})) \tag{2.15}$$

This is precisely the claim

$$\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \vdash \forall\vec{y}(\lambda(\vec{y}) \leftrightarrow \phi(\overline{\ulcorner\lambda(\dot{\vec{y}})\urcorner}, \vec{y})) \tag{2.16}$$

because $\overline{\ulcorner\lambda(\dot{\vec{y}})\urcorner}$ abbreviates $\dot{\mathsf{sub}}_{k-1}(\overline{\ulcorner\lambda(\vec{y})\urcorner}, \overline{\ulcorner y_1\urcorner}, \dot{\mathsf{q}}(\overline{y_1}), \ldots, \overline{\ulcorner y_{k-1}\urcorner}, \dot{\mathsf{q}}(\overline{y_{k-1}}))$. Q.E.D.

Diagonalisation is infamous for being crucial to proving Gödel's First and Second Incompleteness Theorems (1931). They roughly state the following. Let $S \subseteq \mathsf{Sent}(\mathcal{L}^{+}_{\mathsf{PA}})$ be an extension of $\mathsf{PA}_{\mathcal{L}^{+}_{\mathsf{PA}}}$ with some recursively enumerable S^{-}, i.e. $S \coloneqq \mathsf{PA}_{\mathcal{L}^{+}_{\mathsf{PA}}} \cup S^{-}$. Assume that S is consistent. Then:

I. There are statements which are not provable from S, but nonetheless true.

II. S cannot prove its own consistency.

We will not go over the details of the proofs of these two theorems in this thesis, and I refer the interested reader to the references at the start of this section. However, it is important to mention some crucial constructions employed in the proofs, which make use of the arithmetisation of syntax theory. Firstly, by the fact that S is recursively enumerable because both $\mathsf{PA}_{\mathcal{L}^{+}_{\mathsf{PA}}}$ and S^{-} are; one can show that the set of all formulae provable from S, i.e. $\mathsf{Th}(S)$, is also recursively enumerable, i.e. representable in PA. Secondly, using the representability of $\mathsf{Th}(S)$ one can arithmetise the consistency assertion of S. Formally, for π a formula representing S; the complex formula representing $\mathsf{Th}(S)$ in the object language is abbreviated by $\mathbf{Bew}^{\pi}_{S}(x)$; and the arithmetised consistency assertion of S is denoted by $\mathbf{Con}^{\pi}_{S}$ and abbreviates $\neg\mathbf{Bew}^{\pi}_{S}(\overline{\ulcorner \bot \urcorner})$, for any contradiction $\bot$ such as $\underline{0} \neq \underline{0}$.

Two brief asides. Firstly, we can now consider, as promised, an example given by Feferman 1960 to illustrate what goes wrong if we do not have the 'naturalness' condition for (i) the coding on $\mathsf{String}(\mathcal{L}^{+}_{\mathsf{PA}})$, and (ii) the representations of the relevant syntactic categories and operations. Feferman notes that for a recursively enumerable $S \coloneqq \mathsf{PA}_{\mathcal{L}^{+}_{\mathsf{PA}}} \cup S^{-}$ Gödel's Second Incompleteness Theorem is sometimes loosely stated as $S \nvdash \mathbf{Con}_{S}$, where the formula π representing S is omitted. However, this omission introduces a significant ambiguity, given that $\mathbf{Con}^{\pi}_{S}$ differs as π differs. In particular, Feferman 1960 shows that one can find an 'unnatural' formula π^{*} such that (i) π^{*} represents S, and (ii) $S \vdash \mathbf{Con}^{\pi^{*}}_{S}$.

At first sight this might seem to refute Gödel's Second Incompleteness Theorem. However, Gödel and Feferman use different representing formulae. Gödel uses a 'natural' formula

to represent S, call it π', in his original formulation of his proof and shows that $S \nvdash \mathbf{Con}_S^{\pi'}$. Feferman uses π^* to represent S, and it is precisely the specific properties of π^* that ensure that $S \vdash \mathbf{Con}_S^{\pi^*}$. In particular, π^* and π' differ so much so that $\mathbf{Con}_S^{\pi^*}$ and $\mathbf{Con}_S^{\pi'}$ are different statements; and thus Gödel's Second Incompleteness Theorem is not refuted. To the contrary, Feferman 1960 proves that:

(i) if $\mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \subseteq S := \mathsf{PA}_{\mathcal{L}^+_{\mathsf{PA}}} \cup S^- \subseteq \mathsf{Sent}(\mathcal{L}^+_{\mathsf{PA}})$ is recursively enumerable, and

(ii) if S is represented by a formula π equivalent to a quantifier-free formula prefixed with existential and bounded universal quantifiers, which Feferman calls an *RE*-formula;

then $S \nvdash \mathbf{Con}_S^{\pi}$. This is a generalisation of Gödel's Second Incompleteness Theorem, because the representing formula π' that Gödel uses is a specific *RE*-formula; whereas π^* is not an *RE*-formula.

Summing up, the first takeaway of these observations is that Gödel's Second Incompleteness Theorem should be stated precisely as $S \nvdash \mathbf{Con}_S^{\pi'}$, instead of loosely as $S \nvdash \mathbf{Con}_S$. Though it is beyond the scope of this thesis to discuss why *RE*-formulae count as natural representing formulae, the second takeaway is that there are such things as 'natural' and 'unnatural' representing formulae; and that a 'natural' representing formula must be chosen when representing $\mathsf{Th}(S)$ and arithmetising the consistency assertion of S.

As a second aside, observe that the Diagonal Lemma in a sense does not depend on us using arithmetic as our theory of syntax; any proposed theory of syntax must at the very least represent the quotation and ternary substitution functions in the object language, thus proving diagonalisation. However, interestingly, the Diagonal Lemma is not wholly independent from arithmetic, either. Not only can syntax theory be formulated in arithmetic, as we have shown in this chapter, but arithmetic can also be formulated in syntax theory. For details on how this precisely works, see Quine 1946 or Halbach and Leigh 2024, §6.5.

Let us now take a step back and focus on diagonalisation again. Besides being central to proving Gödel's First and Second Incompleteness Theorems, diagonalisation plays

a central role in proving certain inconsistencies and paradoxes, such as the (limitative) results we will prove in chapters 4 and 5. Two such results are seminal in the literature, and provide important contextualisation to the mention predicate approach; Tarski's Undefinability of Truth (1983); and Montague's Paradox (1960); (1963).

We will first consider a paradox relating to truth; Tarski famously proved that truth is undefinable.

Theorem 2.2.4 (Tarski's Undefinability of Truth)**.** *There is no first-order formula* $T(x) \in \mathsf{Form}(\mathcal{L}^{+}_{\mathsf{PA}})$ *such that, for any* $\phi \in \mathsf{Sent}(\mathcal{L}^{+}_{\mathsf{PA}})$*, we have that*

$$\mathsf{PA}_{\mathcal{L}^{+}_{\mathsf{PA}}} \vdash T(\overline{\ulcorner \phi \urcorner}) \leftrightarrow \phi \tag{2.17}$$

Proof. Assume for contradiction that there is such a first-order formula $T(x)$. By the Diagonal Lemma there is a fixed-point λ such that

$$\mathsf{PA}_{\mathcal{L}^{+}_{\mathsf{PA}}} \vdash \lambda \leftrightarrow \neg T(\overline{\ulcorner \lambda \urcorner}) \tag{2.18}$$

By (2.17) we also have that

$$\mathsf{PA}_{\mathcal{L}^{+}_{\mathsf{PA}}} \vdash T(\overline{\ulcorner \lambda \urcorner}) \leftrightarrow \lambda \tag{2.19}$$

Taking (2.18) and (2.19) together we have that

$$\mathsf{PA}_{\mathcal{L}^{+}_{\mathsf{PA}}} \vdash T(\overline{\ulcorner \lambda \urcorner}) \leftrightarrow \neg T(\overline{\ulcorner \lambda \urcorner}) \tag{2.20}$$

which is a contradiction. Q.E.D.

Tarski's Undefinability of Truth is also often stated as a limitative result for a naive axiomatic syntactic theory of truth, know as the Liar Paradox.

Theorem 2.2.5 (Liar Paradox)**.** *Let* $\mathcal{L}^{\mathbf{T}}_{\mathsf{PA}}$ *denote the language that expands* $\mathcal{L}_{\mathsf{PA}}$ *with a unary predicate* $\mathbf{T}$*. Furthermore, let* $\mathsf{PA}_{\mathbf{T}}$ *denote the theory of the axioms of* PA *formulated in* $\mathcal{L}^{\mathbf{T}}_{\mathsf{PA}}$*. Then, any theory* E *that contains* $\mathsf{PA}_{\mathbf{T}}$ *and the schema*

$$\mathbf{T}\overline{\ulcorner \phi \urcorner} \leftrightarrow \phi \tag{$\mathbf{T}$-schema}$$

where $\phi \in \mathsf{Sent}(\mathcal{L}^{\mathbf{T}}_{\mathsf{PA}})$ *is inconsistent.*

Proof. The proof is structurally the same as for the previous formulation of Tarski's Undefinability of Truth. Specifically, by the Diagonal Lemma there is a fixed-point λ such that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg \mathbf{T}^{\ulcorner}\overline{\lambda}^{\urcorner} \tag{2.21}$$

By the **T**-schema, we also have that

$$\mathsf{E} \vdash \mathbf{T}^{\ulcorner}\overline{\lambda}^{\urcorner} \leftrightarrow \lambda \tag{2.22}$$

Taking (2.21) and (2.22) together we have that

$$\mathsf{E} \vdash \mathbf{T}^{\ulcorner}\overline{\lambda}^{\urcorner} \leftrightarrow \neg \mathbf{T}^{\ulcorner}\overline{\lambda}^{\urcorner} \tag{2.23}$$

which implies E is inconsistent. Q.E.D.

Thus, the Liar Paradox shows us that truth as a predicate of names of sentences, couched in a suitable syntax theory, cannot behave the way we naively expect truth to behave; satisfying the **T**-schema. This important result has lead the literature on truth to either argue that truth is not after all a predicate of names of sentences; or to argue that we should somehow weaken the **T**-schema. Axiomatic syntactic theories of truth, such as explored by Halbach 2014, opt for the latter.

2.3 Introducing a modal syntactic predicate

Let us now introduce a single first-order unary modal syntactic predicate, which we will denote **N**, to our syntax theory PA.

Definition 2.3.1. Let $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}}$ denote the language that expands $\mathcal{L}_{\mathsf{PA}}$ with a unary predicate **N**. Furthermore, let $\mathsf{PA}_{\mathbf{N}}$ denote the theory of the axioms of PA formulated in $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}}$.

It is clear that the Liar Paradox suitably modified provides a limitative result for extending $\mathsf{PA}_{\mathbf{N}}$. However, the Liar Paradox provides a truth-theoretic interpretation of a unary syntactic predicate due to the biconditional in **T**-schema. Furthermore, **T**-schema would not be suitable for a modal interpretation of a unary syntactic predicate; it ensures the unary syntactic predicate is the trivial modality.

Nonetheless, as discussed in the introduction, there are also paradoxes and inconsistencies involving a unary syntactic predicate of a more modal flavour. The most notable such paradox is called Montague's Paradox, proved by Kaplan and Montague 1960 and Montague 1963. They proved a limitative result for a naive formalisation of the mention predicate approach, and thus for $\mathsf{PA}_{\mathbf{N}}$.

Theorem 2.3.1 (Montague's Paradox)**.** *Let* E *be a theory that meets the following conditions:*

i. E *contains* $\mathsf{PA}_{\mathbf{N}}$.

ii. E *contains the schema*

$$\mathbf{N}\overline{\ulcorner\phi\urcorner} \rightarrow \phi \tag{$\mathsf{T}_{\mathbf{N}}$}$$

where $\phi \in \mathsf{Sent}(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})$.[13]

iii. E *is closed under the 'necessitation' rule*

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\phi\urcorner} \tag{$\mathsf{NEC}_{\mathbf{N}}$}$$

where $\phi \in \mathsf{Sent}(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})$.[14]

Then E *is inconsistent.*

Proof. By the Diagonal Lemma there is a fixed-point λ such that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner} \tag{2.25}$$

We then have the following:

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner} \qquad (2.25) \qquad (2.26)$$

$$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner} \rightarrow \lambda \qquad \mathsf{T}_{\mathbf{N}} \qquad (2.27)$$

[13] The $\mathsf{T}_{\mathbf{N}}$-schema is the predicate analogue of the schema of the same name in the operator approach. This is not to be confused with Tarski's **T**-schema.

[14] E being closed under the necessitation rule also means that we can iteratively apply the necessitation rule. More specifically, we have that

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\phi\urcorner} \Rightarrow \mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\mathbf{N}\overline{\ulcorner\phi\urcorner}\urcorner} \Rightarrow \mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\mathbf{N}\overline{\ulcorner\mathbf{N}\overline{\ulcorner\phi\urcorner}\urcorner}\urcorner} \Rightarrow \ldots \tag{2.24}$$

where at each step we observe that E is closed under the necessitation rule.

$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner} \rightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}$	(2.26), (2.27)	(2.28)
$\mathsf{E} \vdash \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}$	(2.28)	(2.29)
$\mathsf{E} \vdash \lambda$	(2.26), (2.29)	(2.30)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner}$	$\mathsf{NEC}_{\mathbf{N}}$, (2.30)	(2.31)

Lines (2.29) and (2.31) imply E is inconsistent. Q.E.D.

Montague's Paradox is a strengthening of the Liar Paradox and Tarski's Undefinability of Truth; if we reformulate the biconditional in the **T**-schema as the conjunction of two conditionals, we still get an inconsistency if we weaken the right-to-left conditional into a rule, giving us $\mathsf{NEC}_{\mathbf{N}}$, while retaining the left-to-right conditional in $\mathsf{T}_{\mathbf{N}}$.

Recall from the Introduction (1) that many philosophers have taken Montague's Paradox to show that the mention predicate approach is ill-suited for formalising modal statements; the adequacy criteria for a modal (syntactic) predicate must be weaker than both $\mathsf{T}_{\mathbf{N}}$ and $\mathsf{NEC}_{\mathbf{N}}$, but the principles of factivity (expressed by $\mathsf{T}_{\mathbf{N}}$[15]) and the necessity of tautologies (expressed by $\mathsf{NEC}_{\mathbf{N}}$) are taken to be fundamental to (adequate formalisations of) many modalities, such as metaphysical modality. The operator and mention predicate approaches, on the other hand, have roughly speaking much less strict limitations on the modal principles they can express, and they express most modal principles philosophers are interested in.[16]

However, it is misplaced to conclude from Montague's Paradox that the adequacy criteria for a modal (syntactic) predicate must be weaker than both $\mathsf{T}_{\mathbf{N}}$ and $\mathsf{NEC}_{\mathbf{N}}$. Taking a closer look at the proof of Montague's Paradox, we see that it has four components: instances

[15] If one introduces a truth predicate **T**, predicating over syntactic entities, on top of a modal syntactic predicate **N**, one could also express factivity by the schema $\mathbf{N}\overline{\ulcorner\phi\urcorner} \rightarrow \mathbf{T}\overline{\ulcorner\phi\urcorner}$ (or even by the single axiom $\forall x(\mathbf{N}x \rightarrow \mathbf{T}x)$). However, for **T** to be a truth predicate, one must also assume certain principles for **T**. In particular, for λ the diagonal sentence in the proof of Montague's Paradox, one should not be able to show $\vdash \mathbf{T}\overline{\ulcorner\lambda\urcorner} \rightarrow \lambda$. Otherwise, using $\mathbf{N}\overline{\ulcorner\lambda\urcorner} \rightarrow \mathbf{T}\overline{\ulcorner\lambda\urcorner}$ and Modus Ponens one would regain the instance of $\mathsf{T}_{\mathbf{N}}$ used in the proof of Montague's Paradox. Stern 2014a; Stern 2014b and Stern 2015, §4 explore what principles can be 'safely' assumed for **T** and **N** in this setting.

[16] I say roughly speaking, because there are nuances that, due to concerns of space, in this thesis cannot have the exposition they deserve. Such nuances include the limitations of which modal frame correspondences (that is, which properties of the accessibility relation of a possible-world frame) the operator approach can express (see Blackburn et al. 2001); or whether the logic of the broadest necessity in higher-order logic of Bacon 2018; Bacon 2023 is S4 or S5 (see Bacon 2018; Bacon 2023).

of $\mathsf{T}_\mathbf{N}$, $\mathsf{NEC}_\mathbf{N}$, the Diagonal Lemma, and Classical logic.[17] Montague's Paradox must certainly be addressed, and concessions must be made, as with all paradoxes. Thus, we must instead treat Montague's Paradox as a limitative result, and conclude that we must give up on (the problematic instances of) one or more of these four components.

To sum up this chapter, we provided the necessary logical background to be able to now critically assess and contribute to topics relating to the mention predicate approach. We focussed on formalising modalities as first-order unary predicates that predicate on names of sentences, since this is the standard approach in the literature on the mention predicate approach. In particular, we considered Peano Arithmetic as the theory of syntax we will work in, again following the broader literature. Central to the (limitative) results we will prove in chapters 4 and 5, we proved the Diagonal Lemma and the Uniform Diagonal Lemma, providing us with self-referential expressions in our syntax theory. Finally, we used these Lemmas to prove paradoxes relating to truth and modality, providing limitative results for their behaviour; Tarski's Undefinability of Truth, the Liar Paradox, and ultimately Montague's Paradox.

[17] Observing closely, one can see that the proof of Montague's Paradox provided here only requires Intuitionistic Logic. Furthermore, at the end of chapter 5 we will see that the Diagonal Lemma can also be proved in Intuitionistic Logic. Thus, these four components can be weakened to instances of $\mathsf{T}_\mathbf{N}$, $\mathsf{NEC}_\mathbf{N}$, the Diagonal Lemma, and Intuitionistic Logic. However, until the end of chapter 5 we work in Classical Logic in this thesis.

3

Blocking Diagonalisation To Avoid Paradox

Contents

At the end of the previous chapter 2 we observed that the four components involved in Montague's Paradox are instances of $\mathsf{T}_\mathbf{N}$, $\mathsf{NEC}_\mathbf{N}$, the Diagonal Lemma, and Classical logic. Thus, to avoid Montague's Paradox, we must give up on (the problematic instances of) one or more of these four components. In this chapter we will consider what happens if we give up on the Diagonal Lemma, and inevitably, why this does not provide a means of addressing the paradoxes of the mention predicate approach.

To make sense of and contextualise what we mean by giving up on the Diagonal Lemma, we will first consider an example of a framework that does precisely this; Stern 2014c; Stern 2015 provides an account of the mention predicate approach that

'blocks' diagonalisation by allowing the names of sentences to have only very little structure. Stern then shows that one can construct a class of possible-world models for multiple first-order modal syntactic predicates such that validity with respect to frames where the accessibility relations are restricted in different ways allows one to express many modal principles also expressible in the multimodal operator approach, including $\mathsf{T}_{\mathbf{N}}$ and $\mathsf{NEC}_{\mathbf{N}}$. Before critically assessing Stern's framework, it is important to understand the dialectical role of his framework. In particular, Stern's goal is 'to rebut Montague's assessment that virtually all of modal logic must be sacrificed, if we treat modalities syntactically [c.f. Montague 1963] — at least, if this assessment is understood in its straightforward, general way' (Stern 2014c, p. 561). Thus, Stern's goal is not to provide an account of how one 'should' understand the mention predicate approach; rather, it is an account of how one 'could' understand the mention predicate approach. Stern 2015, §2.3 himself argues that an expressively rich framework, and in particular diagonalisation, is necessary to how we 'should' understand the mention predicate approach.

In this chapter, I will be critically assessing Stern's framework. More specifically, I will be critically assessing whether his framework is an account of how one even 'could' understand the mention predicate approach. That is, whether it is a conception of the mention predicate approach in the first place. Firstly, we will see that Stern's framework, and blocking diagonalisation more generally, when suitably generalised to account for de re modal statements are still susceptible to (a version of) Montague's Paradox; they do not succeed in providing a uniform strategy for avoiding the paradoxes of the mention predicate approach. Secondly, we will also see that the intended interpretation of Stern's framework is committed to diagonalisation, which is in tension with that his framework blocks diagonalisation; we cannot add statements to his framework that are true to his intended interpretation. Taking these two arguments together, we see that Stern's framework does not provide a conception of the mention predicate approach; and blocking diagonalisation does not provide a means of addressing the paradoxes of the mention predicate approach. We will finish this chapter by considering in what tentative cases we can nonetheless use Stern's framework, thereby providing a constructive takeaway of his framework.

3.1 Stern's framework

Let us now examine precisely how Stern 2014c; Stern 2015 'blocks' diagonalisation. He does this by making two distinct moves. Let us focus on a single first-order unary modal syntactic predicate $\mathbf{N}$,[1] and recursively introduce to $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}}$ a class of quotation constant symbols $\ulcorner\phi\urcorner$ for every sentence ϕ of this expanded language; and call this language $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$. In more detail:

Definition 3.1.1 (Term, formula, and quotation degree for $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$)**.** The terms and formulae of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$; and the 'quotation degree' $qd : \mathsf{Form}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}] \cup \mathsf{Term}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}] \to \mathbb{N}$, which is a function that determines the maximal depth of embedded quotations in terms and formulae; are defined simultaneously by induction:

I. If s is a variable or constant symbol of $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}}$, then s is a term of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and $qd(s) = 0$;

II. If p is a 0-place predicate symbol of $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}}$, then p is a 0-place predicate symbol of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and $qd(p) = 0$;

III. If $s_1, \ldots, s_k$ are terms of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and f is a k-place ($k > 0$) function symbol of $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}}$, then $f(s_1, \ldots, s_k)$ is a term of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and $qd(f(s_1, \ldots, s_k)) = max(qd(s_1), \ldots, qd(s_k))$;

IV. If $s_1, \ldots, s_k$ are terms of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and P is a k-place ($n > 0$) predicate symbol of $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}}$ (including '='), then $Ps_1 \ldots s_k$ is a formula of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and $qd(Ps_1 \ldots s_k) = max(qd(s_1), \ldots, qd(s_k))$;

V. If ϕ is a formula of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and x is a variable, then $\neg\phi$ and $\forall x\phi$ are formulae of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and $qd(\neg\phi) = qd(\forall x\phi) = qd(\phi)$;

VI. If ϕ and ψ are formulae of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$, then $\phi \wedge \psi$ is a formula of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and $qd(\phi \wedge \psi) = max(qd(\phi), qd(\psi))$;

VII. The other logical constants are treated as abbreviations using '$\neg, \wedge, \forall$' in the usual way;

[1] We do this for ease of presentation, even though the semantics of Stern 2014c; Stern 2015 considers a truth predicate and finitely many first-order unary modal syntactic predicates $\mathbf{N}_i$, for $1 \leq i \leq n$. The multimodal aspect of the predicate-semantics of Stern 2014c; Stern 2015 is more relevant if we compare Stern's approach to the generalisation we will construct in chapter 4 of the possible-world predicate-semantics developed by Halbach and Leigh 2024 to multiple first-order unary modal syntactic predicates.

VIII. If ϕ is a sentence of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$, then $\ulcorner\phi\urcorner$ is a quotation constant symbol of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and $qd(\ulcorner\phi\urcorner) = qd(\phi) + 1$.

The quotation constant symbols will act as the names of the sentences of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$, and are thus what $\mathbf{N}$ will be able to predicate on. Furthermore, observe that we only introduce quotation constant symbols for the sentences of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$, and not for the formulae or arbitrary strings of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$. Thus, $\mathbf{N}$ will be able to predicate on all and only the names of the sentences of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$. Following much of the literature on the mention predicate approach, Stern explains that for reasons of simplicity he restricts his investigation in this way to only account for de dicto modalities (Stern 2015, §1.3).

Now assume that we are assessing what the behaviour of our modal syntactic predicate might be. In particular, we are assessing whether we can emulate Montague's Paradox in this new setting.[2] Therefore, let E be a theory that satisfies the following conditions:

i. E contains the axioms of PA formulated in $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$.

ii. E contains the axiom schema

$$\mathbf{N}\ulcorner\phi\urcorner \rightarrow \phi \tag{$\mathsf{T}_{\mathbf{N}}^{\mathcal{S}}$}$$

where $\phi \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}]$.

iii. E is closed under the necessitation rule

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}\ulcorner\phi\urcorner \tag{$\mathsf{NEC}_{\mathbf{N}}^{\mathcal{S}}$}$$

where $\phi \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}]$.

Since E contains the axioms of PA, we can still technically prove the Diagonal Lemma. However, we can only prove the Diagonal Lemma for the Gödel numbers of formulae. That is, following the proof of Montague's Paradox, by the Diagonal Lemma there is a λ such that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner} \tag{3.1}$$

[2]We could also assess whether any other paradox from the 'standard' mention predicate approach detailed in Definition 2.3.1, like the Liar Paradox, can be emulated in this new setting. However, we are focussing on emulating Montague's Paradox due to its historical significance with regards to the reception of the mention predicate approach, and the fact that Stern is responding to Montague.

Observe that if we had

$$\mathsf{E} \vdash \lfloor \lambda \rfloor = \overline{\ulcorner \lambda \urcorner} \tag{3.2}$$

we could recover an analogue of Montague's Paradox in this new setting using $\mathsf{T}^{S}_{\mathbf{N}}$ and $\mathsf{NEC}^{S}_{\mathbf{N}}$; by substitution on (3.1) using (3.2) we would get

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg \mathbf{N} \lfloor \lambda \rfloor \tag{3.3}$$

and the analogue of Montague's Paradox would then be proved by following the analogues of lines (2.26) - (2.31). Thus, the first move Stern makes is to block (3.2) by ensuring that the models he constructs do not identify the interpretations of quotation constant symbols with their Gödel numbers. However, this does not block the analogue of Montague's Paradox just yet. For if we had diagonalisation on the *quotation constant symbols* themselves, we would be able to directly show (3.3). Thus, the second move Stern makes is to block (3.3) by ensuring that the models he constructs cannot satisfy a sufficient fragment of syntax theory / arithmetic on the quotation constant symbols to satisfy diagonalisation on them.

Schweizer 1992, p. 7 succinctly states what Montague's Paradox, and other paradoxes such as the Liar Paradox, rely on:

> Thus there are really two independent assumptions built into the above phenomenon of modal self-reference:
>
> (a) the possession of a class of terms structurally rich enough to do arithmetic and to sustain the diagonal lemma, and
>
> (b) the use of these terms as the privileged names of syntactical objects in defining the modal logic.

Stern 'blocks' diagonalisation by denying assumption (b); the 'privileged names of syntactic objects in defining the modal logic' are not 'structurally rich enough to do arithmetic and to sustain the diagonal lemma'.[3]

To make the previous assessments more precise, we will now consider the class of possible-world models that Stern constructs in (2014); (2015). To reiterate, this construction is a

[3]Tarski 1983 also 'blocks' diagonalisation by denying assumption (b) when he types the truth predicate **T** (for any object language L, the truth for L can only be formulated in a meta-language M that contains L and a truth predicate $\mathbf{T}_L$ for L); M only contains privileged names for sentences of L.

modal generalisation of the Revision Theory of Truth following Gupta 1982; and satisfies the Adequacy Condition.

Definition 3.1.2 (Proper premodel)**.** A proper premodel of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ is a pair $\mathcal{M} := \langle D, I\rangle$; with domain $\mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}] \subsetneq D$; and interpretation function I defined on $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}} \setminus \{\mathbf{N}\}$, where I satisfies the following conditions:

I. $I(\ulcorner\phi\urcorner) = \phi$ for every $\phi \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}]$;

II. If a term s is not a quotation constant symbol, then $I(s) \notin \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}]$;

III. If P is a k-place predicate symbol of $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}}$ (excluding '=') and $\phi_i \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}]$ for $1 \leq i \leq k$, then $\langle \phi_1, \ldots, \phi_i, \ldots, \phi_k\rangle \in I(P)$ iff for all $\phi_i' \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}]$ we have that $\langle \phi_1, \ldots, \phi_i', \ldots, \phi_k\rangle \in I(P)$.

IV. If f is a k-place ($k > 0$) function symbol, letting $Ran(I(f))$ be the range of $I(f)$,[4] we have that $Ran(I(f)) \cap \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}] = \emptyset$. Furthermore, if $\phi_i \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}]$ for $1 \leq i \leq k$, then $I(f)(\phi_1, \ldots, \phi_i, \ldots, \phi_k) = I(f)(\phi_1, \ldots, \phi_i', \ldots, \phi_k)$ for all $\phi_i' \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}]$.

Importantly, premodels block (3.2) and (3.3). Firstly, observe that I and II imply that no premodel interprets a quotation constant symbol and a term that is not quotation constant symbol as the same object. In particular, for all quotation constant symbols $\ulcorner\phi\urcorner$ the schema

$$\forall x((x = \underline{0} \vee \exists y(x = \mathsf{S}(y))) \rightarrow x \neq \ulcorner\phi\urcorner) \tag{3.4}$$

is valid in the class of premodels. Instantiating each instance of the schema (3.4) with $\overline{\ulcorner\phi\urcorner}$ we get that the schema $\overline{\ulcorner\phi\urcorner} \neq \ulcorner\phi\urcorner$ is valid, i.e. the premodels do not identify the interpretations of quotation constant symbols with their Gödel numbers. Thus, the premodels block (3.2). Likewise, for any other suitable coding

$$g : \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}] \rightarrow \{\overline{n} \mid n \in \mathbb{N}\} \tag{3.5}$$

we have that the schema $g(\phi) \neq \ulcorner\phi\urcorner$ is valid. Even further, for any suitable coding

$$g^* : \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}] \rightarrow \mathsf{Term}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}] \setminus \{\ulcorner\phi\urcorner \mid \phi \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}]\} \tag{3.6}$$

[4] $Ran(I(f)) = \pi_{k+1}(I(f)) := \{\pi_{k+1}(a) \mid a \in I(f)\}$, where π_{k+1} is the $(k+1)$-th projection function.

such that the (a) (sub)class of premodels satisfies a sufficient fragment of syntax theory / arithmetic on $g^*(\mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}]) \coloneqq \{g^*(\phi) \mid \phi \in \mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}]\}$ to satisfy diagonalisation on

$g^*(\mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}])$; we have that the schema $g^*(\phi) \neq \ulcorner\phi\urcorner$ is valid.

Secondly, observe that III and IV state that the interpretations of predicates (except for $\mathbf{N}$) and functions in premodels do not discriminate between sentences in $\mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}]$; IV ensures that no premodel interprets a *function symbol* of $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}$ as a syntactic operation on $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}$; and III ensures that no premodel interprets a *complex formula* of $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}} \setminus \{\mathbf{N}\}$ as a syntactic operation on $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}$. Therefore, the premodels cannot satisfy a sufficient fragment of syntax theory / arithmetic on the quotation constant symbols to satisfy diagonalisation on them. Thus, the premodels block (3.3).

As Stern points out, the restrictions in the definition of a premodel are too strong, and can be weakened in different ways (Stern 2014c, p. 558). For example, certain syntactic operations, such as the negation or conjunction functions, could be safely introduced to Stern's construction. However, Stern notes that it is an open question just how much structure the class of quotation constant symbols can have in his construction; or more generally while avoiding paradox. At the very least, it is clear that either the quotation function, or the ternary substitution function, defined on the quotation constant symbols cannot be representable in the object language theory; otherwise we could directly prove the Diagonal Lemma, and thereby as a corollary (3.3).

Stern finally extends premodels to models by giving a revisionary interpretation of $\mathbf{N}$. He does this as follows.

Definition 3.1.3 (Modal premodel frames and evaluation functions)**.** Let $\mathcal{W}$ be a set of premodels of $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}$, and R be a binary relation on $\mathcal{W}$. Then we call $F \coloneqq \langle \mathcal{W}, R\rangle$ a modal premodel frame of $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}$. Furthermore, we write $\mathcal{M}R\mathcal{M}'$ for $\langle \mathcal{M}, \mathcal{M}'\rangle \in R$, and define $R(\mathcal{M}) \coloneqq \{\mathcal{M}' \in \mathcal{W} \mid \langle \mathcal{M}, \mathcal{M}'\rangle \in R\}$. Lastly, we call a function $f : \mathcal{W} \to \mathcal{P}[\mathsf{Sent}((\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}})]$ an evaluation function relative to F. We denote by Val_F the set of all evaluation functions relative to F.

Definition 3.1.4 (Models of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$)**.** Let F be a modal premodel frame of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ and $f \in Val_F$. Then F and f induce models $\mathcal{M}^f := \langle \mathcal{M}, Y_{\mathcal{M}^f} \rangle$ of $(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}$ for every $\mathcal{M} \in \mathcal{W}$. Here,

$$Y_{\mathcal{M}^f} := \bigcap_{\mathcal{M}' \in R(\mathcal{M})} f(\mathcal{M}') \tag{3.7}$$

is the extension of the modal syntactic predicate **N**.

Definition 3.1.5 (Modal jump)**.** Let F be a frame. The modal jump relative to F is an operation $\Xi_F : Val_F \to \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}]$ such that for all $f \in Val_F$ and $\mathcal{M} \in \mathcal{W}$ we have that

$$\Xi_F(f)(\mathcal{M}) := \{\phi \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}] \mid \mathcal{M}^f \vDash \phi\} \tag{3.8}$$

By transfinite recursion we define iterative applications of Ξ_F, for α an ordinal:

$$\Xi_F^{\alpha}(f) := \begin{cases} f & \text{for} \quad \alpha = 0 \\ \Xi_F(\Xi_F^{\beta}(f)) & \text{for} \quad \alpha = \beta + 1 \\ g \in Val_F & \text{for} \quad \alpha \text{ a limit ordinal} \end{cases} \tag{3.9}$$

where for all $\mathcal{M} \in \mathcal{W}$

$$g(\mathcal{M}) := \{\phi \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}] \mid \exists \delta(\phi \in \bigcap_{\delta \leq \beta < \alpha} \Xi_F^{\beta}(f)(\mathcal{M}))\} \tag{3.10}$$

Since $\Xi_F^{\alpha}(f) \in Val_F$ for any α, iteratively applying the modal jump operation relative to F to some $f \in Val_F$ results in a revisionary sequence of interpretations of **N** given by

$$\bigcap_{\mathcal{M}' \in R(\mathcal{M})} \Xi_F^{\alpha}(f)(\mathcal{M}') \tag{3.11}$$

for any $\mathcal{M} \in \mathcal{W}$. This iteration reaches a unique fixed point at ω:

Lemma 3.1.1. *Let F be a modal premodel frame and $f, g \in Val_F$. Then for all $\mathcal{M} \in \mathcal{W}$, all $n \in \mathbb{N}$, and all ordinals α: if $\alpha > n+1$, then for all $\phi \in \{\psi \in \mathsf{Sent}[(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}})^{\mathcal{S}}] \mid qd(\psi) \leq n\}$ we have that*

$$\phi \in \Xi_F^{n+2}(f)(\mathcal{M}) \Leftrightarrow \phi \in \Xi_F^{\alpha}(g)(\mathcal{M}) \tag{3.12}$$

Proof. For a proof of this, and other results of Stern discussed in this chapter, see his proofs in (2014); (2015). Q.E.D.

Corollary 3.1.1. *Let F be a modal premodel frame. Then for all $f, g \in Val_F$ and all $\phi \in \mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}]$*

$$\Xi^{\omega}_F(f) = \Xi^{\omega+1}_F(g) \tag{3.13}$$

$$\mathcal{M}^{\Xi^{\omega}_F(f)} \vDash \phi \Leftrightarrow \phi \in \Xi^{\omega}_F(f)(\mathcal{M}) \tag{3.14}$$

In particular, corollary 3.1.1 implies that there is a unique fixed point $g \in Val_F$ of the operation $\Xi_F(\cdot)$, i.e. $\Xi_F(g) = g$. Given this uniqueness, we can now define the following:

Definition 3.1.6 (Proper model)**.** Let F be a modal premodel frame, g be the unique fixed point of the operation $\Xi_F(\cdot)$, and $\mathcal{M} \in \mathcal{W}$. Then we call the model $\mathcal{M}^g$ induced by F and g a proper model of $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}$. Because of the uniqueness of g, $\mathcal{M}^g$ is also unique. Let $\phi \in \mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}]$. If $\mathcal{M}^g \vDash \phi$ for all $\mathcal{M} \in \mathcal{W}$ we write that $F \vDash \phi$. For a class of frames $\mathfrak{F}$, if $F \vDash \phi$ for all $F \in \mathfrak{F}$ we write that $\mathfrak{F} \vDash \phi$.

Theorem 3.1.1. *Let F be a modal premodel frame, and let $\phi, \psi \in \mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}]$. Then:*

$$F \vDash \mathbf{N}\ulcorner\phi \rightarrow \psi\urcorner \rightarrow (\mathbf{N}\ulcorner\phi\urcorner \rightarrow \mathbf{N}\ulcorner\psi\urcorner) \tag{3.15}$$

$$F \vDash \phi \Rightarrow F \vDash \mathbf{N}\ulcorner\phi\urcorner \tag{3.16}$$

Theorem 3.1.2. *Let $\mathfrak{F}$ be the class of all modal premodel frames F where $\forall \mathcal{M} \in \mathcal{W}_F\ \mathcal{M} \in R_F(\mathcal{M})$. These frames are called reflexive. Letting $\phi \in \mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}]$, we then have that*

$$\mathfrak{F} \vDash \mathbf{N}\ulcorner\phi\urcorner \rightarrow \phi \tag{3.17}$$

We will now introduce the machinery necessary to make the Adequacy Condition precise. Recall it states that validity with respect to modal premodel frames where the accessibility relation is restricted in different ways allows one to express most modal principles also expressible in the operator approach.

Definition 3.1.7 (first-order propositional modal operator logic and possible-world operator-semantics)**.** What follows is a brief exposition of relevant notions from first-order propositional modal operator logic, which treats modality as a sentential operator and

thus falls under the operator approach. In particular, we consider the possible-world O-semantics for a modal sentential operator.

- The language $L_\Box$ of propositional modal operator logic contains no function or k-place ($k > 0$) predicate symbols; and contains countably many 0-place predicate symbols, i.e. propositional sentence letters. The set of all propositional sentence letters is denoted SL. Furthermore, the logical constants of $L_\Box$ are $\neg$, $\wedge$, and $\Box$. The formulae of $L_\Box$ are defined in the usual way.

- We call a function $i : SL \to 2$ (where $2 = \{0, 1\}$) that interprets the propositional sentence letters a premodel of $L_\Box$. The set of all premodels is denoted $2^{SL} := \{i \mid i : SL \to 2\}$.

- We call a pair $F_\Box := \langle W, R\rangle$ a possible-world frame of $L_\Box$; where W is some set of labels, called 'worlds', and R is a binary relation on W. As before, we write wRv for $\langle w, v\rangle \in R$, and define $R(w) := \{v \in W \mid \langle w, v\rangle \in R\}$.

- We call a tuple $\mathcal{M}_\Box := \langle W, R, I\rangle$ a possible-world model of $L_\Box$; where W and R form a frame $\langle W, R\rangle$, and $I : W \to 2^{SL}$ maps worlds to premodels. The set of all such I is denoted $(2^{SL})^W$.

- Complex formulae are evaluated as either true or false at a world in a model. Let $\mathcal{M}_\Box = \langle W, R, I\rangle$ be a model. The valuation function $V_I : W \to (\mathsf{Sent}(L_\Box) \to \{0, 1\})$ is defined inductively as follows:

  - $V_I(w, p) = I(w, p)$ if $p \in SL$
  - $V_I(w, \neg\phi) = 1 - V_I(w, \phi)$
  - $V_I(w, \psi_1 \wedge \psi_2) = Min(V_I(w, \psi_1), V_I(w, \psi_2))$
  - $V_I(w, \Box\phi) = 1$ iff $[I(v, \phi) = 1$ for all $v \in R(w)]$.

- Let $\phi \in \mathsf{Sent}(L_\Box)$, $\mathcal{M}_\Box = \langle W, R, I\rangle$ be a model, and $\mathfrak{F}_\Box$ be some class of frames. Then:

- If $V_I(w, \phi) = 1$ for some $w \in W$, we say that ϕ is true at w and write $\mathcal{M}_\Box(w) \vDash \phi$.
- If $V_I(w, \phi) = 1$ for all $w \in W$, we say that $\mathcal{M}_\Box$ makes ϕ true and write that $\mathcal{M}_\Box \vDash \phi$.
- If $\langle W, R, I\rangle \vDash \phi$ for all $I \in (2^{SL})^W$, we say that $\langle W, R\rangle$ makes ϕ true and write $\langle W, R\rangle \vDash \phi$.
- If $F_\Box \vDash \phi$ for all $F_\Box \in \mathfrak{F}_\Box$, we say that $\mathfrak{F}_\Box$ makes ϕ valid and write $\mathfrak{F}_\Box \vDash \phi$.
- If $F_\Box \vDash \phi$ for all frames $F_\Box$, we say that ϕ is valid and write $\vDash \phi$.

Definition 3.1.8 (Property Φ)**.** Let $\mathfrak{F}$ be a class of modal premodel frames as in definition 3.1.3; and let $\mathfrak{F}_\Box$ be a class of possible-world frames as in definition 3.1.7; such that the accessibility relations of all $F \in \mathfrak{F}$ and $F_\Box \in \mathfrak{F}_\Box$ satisfy property Φ. We then say that $\mathfrak{F}$ and $\mathfrak{F}_\Box$ have property Φ.

Definition 3.1.9 (Translation from $L_\Box$ to $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^S$)**.** We call a function $* : L_\Box \to \mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^S]$ a realisation. A function $H^* : \mathsf{Sent}(L_\Box) \to \mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^S]$ is a translation function iff the following conditions are satisfied, for any $\phi, \psi_1, \psi_2 \in \mathsf{Sent}(L_\Box)$:

- $H^*(p) = *(p)$ if $p \in SL$.
- $H^*(\bot) = \bot$.
- $H^*(\neg\phi) = \neg H^*(\phi)$.
- $H^*(\psi_1 \wedge \psi_2) = H^*(\psi_1) \wedge H^*(\psi_2)$.
- $H^*(\Box\phi) = \mathbf{N}\ulcorner H^*(\phi)\urcorner$.

We can now make the Adequacy Condition precise.

Theorem 3.1.3 (Semantic Adequacy Condition)**.** *Let $\mathfrak{F}^\Phi_\Box$ be the class of all possible-world frames as in definition 3.1.7 with property Φ. Likewise, let $\mathfrak{F}^\Phi$ be the class of all modal premodel frames as in definition 3.1.3 with property Φ. Then for all $\phi \in \mathsf{Sent}(L_\Box)$*

$$\mathfrak{F}^\Phi_\Box \vDash \phi \Leftrightarrow \textit{for all realisations } * [\mathfrak{F}^\Phi \vDash H^*(\phi)] \tag{3.18}$$

Corollary 3.1.2 (Proof-theoretic Adequacy Condition)**.** *Let* $S \subset \mathsf{Sent}(L_\Box)$ *be proof-theoretically sound and complete with respect to the class of all possible-world frames as in definition 3.1.7 with property* Φ*. Let* $\mathfrak{F}^\Phi$ *be the class of all modal premodel frames as in definition 3.1.3 with property* Φ*. Then for all* $\phi \in \mathsf{Sent}(L_\Box)$

$$S \vdash \phi \Leftrightarrow \textit{for all realisations} * [\mathfrak{F}^\Phi \vDash H^*(\phi)] \tag{3.19}$$

Having made the Adequacy Condition precise, we can see why Stern's construction can only express *most*, and not all, modal principles also expressible in the operator approach; the translation Stern employs maps modal *propositional* logic to $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^S$. Thus, results 3.1.3 and 3.1.2 only prove that Stern's construction can express the modal principles that propositional modal operator logic can express, which are precisely the modal principles that can, given possible-world O-semantics, be captured by strictly restricting the accessibility relation.[5] In effect this is because Stern's construction only allows one to formalise de dicto modal statements, and not de re modal statements, with the modal syntactic predicate **N**.

3.2 Problems with Stern's framework

Now that we have in detail considered Stern's possible-world model construction in (2014); (2015), in particular how he 'blocks' diagonalisation, we can consider why Stern's framework does not provide a conception of the mention predicate approach; and why blocking diagonalisation does not provide a means of addressing the paradoxes of the mention predicate approach. We will consider two arguments to this end.

3.2.1 The problem of formalising de re modal statements

Firstly, recall that Stern explains that for reasons of simplicity he restricts his investigation to de dicto modalities (Stern 2015, §1.3); and recall that in the mention predicate approach de re modal statements seem to be best formalised using a binary modal satisfaction

[5] Compare this with for example the Barcan ($\mathsf{BF}_\Box$) and Converse Barcan ($\mathsf{CBF}_\Box$) Formulae below, which are the modal principles that are captured by domains shrinking or growing respectively along the accessibility relation.

predicate that predicates on pairs consisting of names of formulae and names of variable assignments. Just like with a unary modal syntactic predicate, such a binary modal satisfaction predicate is susceptible to inconsistency and paradox. Some of these results make crucial use of syntactic machinery, i.e. syntactic or arithmetic results such as the representability of the ternary substitution function or the Diagonal Lemma, to arrive at inconsistency or paradox. Thus, generalising Stern's construction in (2014); (2015) for a modal satisfaction predicate, where diagonalisation into the modal satisfaction predicate is 'blocked', would likewise avoid such paradoxes.

However, there are also paradoxes and inconsistencies involving a satisfaction predicate that do not rely on syntactic machinery. Along a truth-theoretic interpretation of a satisfaction predicate, consider for example the following paradox, which is Russell's Paradox or Tarski's Undefinability of Truth for a satisfaction predicate.

Theorem 3.2.1 (Undefinability of Satisfaction)**.** *Consider a first-order language* $\mathcal{L}_{\mathsf{Sat}}$ *with identity that contains a binary predicate symbol* $\mathsf{Sat}(x, y)$*. Just like in Definition 3.1.1, for each formula* ϕ *of* $\mathcal{L}_{\mathsf{Sat}}$ *there is a closed term* $\ulcorner\phi\urcorner$ *in* $\mathcal{L}_{\mathsf{Sat}}$*. Then any theory* E *extending the schema*

$$\forall x(\mathsf{Sat}(\ulcorner\phi(x)\urcorner, x) \leftrightarrow \phi(x)) \tag{TS}$$

is inconsistent.

Proof. The proof is structurally the same as the proof of Russell's Paradox. We instantiate $\phi(x)$ in TS with $\neg\mathsf{Sat}(x, x)$; and then take $\ulcorner\neg\mathsf{Sat}(x, x)\urcorner$ as the value of x:

$$\begin{aligned} &\mathsf{E} \vdash \mathsf{Sat}(\ulcorner\neg\mathsf{Sat}(x, x)\urcorner, \ulcorner\neg\mathsf{Sat}(x, x)\urcorner) \\ &\qquad\leftrightarrow \\ &\qquad\neg\mathsf{Sat}(\ulcorner\neg\mathsf{Sat}(x, x)\urcorner, \ulcorner\neg\mathsf{Sat}(x, x)\urcorner) \end{aligned}$$

This is a contradiction. Q.E.D.

Firstly, as alluded to, looking at the above proof of Theorem 3.2.1 one sees that this inconsistency result does not rely on any syntactic machinery. In particular, the schema TS is too weak to prove such results as diagonalisation. Secondly, note that Theorem 3.2.1

involves a binary satisfaction predicate that does not predicate on variable assignments, but rather over the values of the specific variables x and y. Hence, such a binary satisfaction predicate is weaker than a binary satisfaction predicate that predicates on variable assignments; and thus Theorem 3.2.1, appropriately modified, also holds for such a binary satisfaction predicate.

There are also more complicated and involved paradoxes and inconsistencies involving a satisfaction predicate in the literature. Consider for example the paradox provided by Halbach and Zhang 2017, which is a non-arithmetic version of Visser's paradox in Visser 1989.

Theorem 3.2.2 (Visser without Gödel)**.** *Consider a first-order language $\mathcal{L}_{\mathsf{Sat}}$ with identity that contains a binary predicate symbol $<$ and a quaternary predicate symbol $\mathsf{Sat}_x(y, z, w)$. We say that variable x is in 'index' position. Just like in Definition 3.1.1, for each formula ϕ of $\mathcal{L}_{\mathsf{Sat}}$ there is a closed term $\ulcorner\phi\urcorner$ in $\mathcal{L}_{\mathsf{Sat}}$. Now consider the following two axioms, and axiom schema*

$$\forall x \exists y\ x < y \tag{SER}$$

$$\forall x \forall y \forall z (x < y \rightarrow (y < z \rightarrow x < z)) \tag{TRANS}$$

$$\forall x \forall y (\mathsf{Sat}_y(\ulcorner\phi(x, y)\urcorner, x, y) \leftrightarrow \phi(x, y)) \tag{VS}$$

where any occurrence of a variable in index position in $\phi(x, y)$ is bound; and furthermore any occurrence of the quantifiers $\forall v$ or $\exists v$ that binds an occurrence of the variable v in index position is restricted by $v > y$. Here, $\forall v > y\ \psi$ is understood in the usual way as $\forall v(y < v \rightarrow \psi)$; and $\exists v > y\ \psi$ as $\exists v(y < v \wedge \psi)$.

Then the schema VS *is consistent; whereas any theory* E *extending the schema* VS*, and axioms* SER *and* TRANS *is inconsistent.*

Proof. For a detailed proof see Halbach and Zhang 2017. To sketch the consistency of VS, if we assume $<$ is converse well-founded, i.e. for any set of objects we have a $<$-maximal element, by induction on $(<)^{-1}$ we can construct models of VS.

To sketch the inconsistency, we instantiate $\phi(x, y)$ in VS with $\psi(x, y) \coloneqq \forall z > y \neg \mathsf{Sat}_z(x, x, z)$. Then for arbitrary a, assuming $\mathsf{Sat}_a(\ulcorner\psi(x, y)\urcorner, \ulcorner\psi(x, y)\urcorner, a)$ leads to a contradiction; by using VS, SER, TRANS, and VS again. Thus, for any a we have that $\neg\mathsf{Sat}_a(\ulcorner\psi(x, y)\urcorner, \ulcorner\psi(x, y)\urcorner, a)$ is the case. However, using VS a final time we conclude that $\mathsf{Sat}_a(\ulcorner\psi(x, y)\urcorner, \ulcorner\psi(x, y)\urcorner, a)$.

Q.E.D.

Again, the proof of Theorem 3.2.2 does not rely on any syntactic machinery; and appropriately modified holds for a binary satisfaction predicate that predicates on variable assignments.

Theorems 3.2.1 and 3.2.2 provide truth-theoretic interpretations of a satisfaction predicate due to the biconditional in the schemata TS and VS. However, neither the principle TS nor VS would be suitable for a modal interpretation of a satisfaction predicate; they ensure the satisfaction predicate is the trivial modality. Thus, these paradoxes do not yet pose a problem for generalising Stern's construction for a modal satisfaction predicate. However, there are also paradoxes and inconsistencies involving a satisfaction predicate of a more modal flavour that do not rely on syntactic machinery.

Theorem 3.2.3 (Montague's Paradox for Satisfaction)**.** *Consider a first-order language $\mathcal{L}_{\mathsf{Sat}}$ with identity that contains a binary predicate symbol $\mathsf{Sat}(x, y)$. Just like in Definition 3.1.1, for each formula ϕ of $\mathcal{L}_{\mathsf{Sat}}$ there is a closed term $\ulcorner\phi\urcorner$ in $\mathcal{L}_{\mathsf{Sat}}$. Let* E *be a theory that extends the schema*

$$\forall x(\mathsf{Sat}(\ulcorner\phi(x)\urcorner, x) \rightarrow \phi(x)) \qquad (\mathsf{T}_{\mathsf{Sat}})$$

and is closed under the rule

$$\mathsf{E} \vdash \phi(a/x) \Rightarrow \mathsf{E} \vdash \mathsf{Sat}(\ulcorner\phi(x)\urcorner, a) \qquad (\mathsf{NEC}_{\mathsf{Sat}})$$

when a is an arbitrary closed term. Then E *is inconsistent.*

Proof. The proof is a slight modification of the proof of Theorem 3.2.1, in structurally the same way that the proof of Montague's Paradox is a slight modification of the proof

of Tarski's Undefinability of Truth. Again, we instantiate $\phi(x)$ in $\mathsf{T}_{\mathsf{Sat}}$ with $\neg\mathsf{Sat}(x, x)$; and then take $\ulcorner\neg\mathsf{Sat}(x, x)\urcorner$ as the value of x:

$$\mathsf{E} \vdash \mathsf{Sat}(\ulcorner\neg\mathsf{Sat}(x, x)\urcorner, \ulcorner\neg\mathsf{Sat}(x, x)\urcorner)$$

$$\rightarrow$$

$$\neg\mathsf{Sat}(\ulcorner\neg\mathsf{Sat}(x, x)\urcorner, \ulcorner\neg\mathsf{Sat}(x, x)\urcorner) \tag{3.20}$$

$$\mathsf{E} \vdash \neg\mathsf{Sat}(\ulcorner\neg\mathsf{Sat}(x, x)\urcorner, \ulcorner\neg\mathsf{Sat}(x, x)\urcorner) \quad (3.20) \tag{3.21}$$

$$\mathsf{E} \vdash \mathsf{Sat}(\ulcorner\neg\mathsf{Sat}(x, x)\urcorner, \ulcorner\neg\mathsf{Sat}(x, x)\urcorner) \quad \mathsf{NEC}_{\mathsf{Sat}}, (3.21) \tag{3.22}$$

Lines (3.21) and (3.22) are contradictory. Q.E.D.

Firstly, Theorem 3.2.3 provides a modal interpretation of a satisfaction predicate; $\mathsf{T}_{\mathsf{Sat}}$ is a modification of $\mathsf{T}_{\mathbf{N}}$ with a satisfaction predicate, and thus is a first pass at capturing factivity for a modality; and $\mathsf{NEC}_{\mathsf{Sat}}$ is a modification of $\mathsf{NEC}_{\mathbf{N}}$ with a satisfaction predicate, and thus is a first pass at capturing the modal status of logical truths. Secondly, the proof of Theorem 3.2.3 similarly does not rely on any syntactic machinery; and appropriately modified holds for a binary satisfaction predicate that predicates on variable assignments.

Thus, Theorem 3.2.3 also holds when generalising Stern's construction by means of a binary modal satisfaction predicate that predicates on variable assignments, noting that this generalisation still blocks syntactic machinery. As we have seen simply introducing a separate class of quotation constant symbols avoids paradox involving a unary modal syntactic predicate, and allows one to express with a unary predicate the modal principles that propositional modal operator logic can express. However, simply introducing a separate class of quotation constant symbols does not yet avoid paradox involving a binary modal satisfaction predicate, in particular Montague's Paradox for Satisfaction; and thus does not yet allow one to express with a binary satisfaction predicate the modal principles that first-order modal operator logic can express. Therefore, in generalising Stern's construction one must resort to addressing paradoxes and inconsistencies involving a binary modal satisfaction predicate in a substantially different way to how Stern 2014c; Stern 2015 addresses structurally the same paradoxes and inconsistencies involving a

unary modal syntactic predicate. This is precisely because blocking syntactic machinery does not avoid structurally the same paradoxes for such a generalisation, as Theorem 3.2.3 shows. Thus, something else must give. As a consequence, it is a non-trivial matter, if at all possible, for such a generalisation to account for de re modal statements by obtaining an adequacy condition for first-order modal operator logic.

Provided that generalising Stern's construction will account for both de dicto and de re modal statements, addressing structurally the same paradoxes and inconsistencies in a substantially different way does not provide a uniform strategy for avoiding paradox; and casts doubt on the suitability and efficacy, and hence viability, of Stern's initial construction in (2014); (2015), which accounts for only de dicto modal statements, as a conception of the mention predicate approach. More generally, we see that no matter how one blocks diagonalisation for a unary modal syntactic predicate, when such a framework is generalised by means of a binary modal satisfaction predicate it is still susceptible to structurally the same paradoxes, if need be by suitably modifying Theorem 3.2.3. As above, such a generalisation must resort to addressing structurally the same paradoxes and inconsistencies in a substantially different way to those involving a unary modal syntactic predicate. Again as above, given that such a generalisation will account for both de dicto and de re modal statements, addressing structurally the same paradoxes and inconsistencies in a substantially different way casts doubt on blocking diagonalisation as a means of addressing the paradoxes of the mention predicate approach.

3.2.2 The importance of de re modal statements

This is a good point in this thesis to go into detail why it is important to generalise frameworks for the mention predicate approach to also account for de re modal statements (even though due to limitations of space the framework developed in this thesis is not generalised in such a way, and only accounts for de dicto modal statements). That is because much of current philosophy revolves around de re modal statements / principles. Consider for example the necessitism / contingentism debate, which lies at the centre of modern metaphysics. As Williamson puts it, necessitism is the claim that 'it is necessary that everything is such that it is necessary that something is identical with it'

(Williamson 2013, p. 3). Contingentism, on the other hand, is the negation of necessitism. Modern metaphysics is deeply entwined with modal logic; the operator and use predicate approaches are often used to formalise and investigate claims of metaphysics, where the modality involved is understood as metaphysical modality. In particular, in first-order modal operator logic the necessitism / contingentism debate revolves around the following two claims, known as the Barcan schema ($\mathsf{BF}_\Box$) and the Converse Barcan schema ($\mathsf{CBF}_\Box$) respectively:

$$\Diamond\exists x\phi \rightarrow \exists x\Diamond\phi \qquad (\mathsf{BF}_\Box)$$

$$\exists x\Diamond\phi \rightarrow \Diamond\exists x\phi \qquad (\mathsf{CBF}_\Box)$$

Both of these claims are de re, since the modal complement of the sentence '$\exists x\Diamond\phi$' is the open formula ϕ, assuming there is no vacuous quantification. For example, in 'something is such that it is possible that it is green' (an instance of '$\exists x\Diamond\phi$') we are making a modal claim about some thing, to wit that it is green; we are not making a modal claim about a sentence / proposition, unlike in 'it is possible that something is green' (an instance of '$\Diamond\exists x\phi$').

$\mathsf{BF}_\Box$ expresses the de re modal principle that if something possibly exists, it also exists. Given the possible-world O-semantics, the domains of accessible possible worlds can only shrink; one can show that all instances of $\mathsf{BF}_\Box$ are valid on a frame $\langle W, R\rangle$ iff [for any worlds $w, u \in W$, if $u \in R(w)$, then $D_u \subseteq D_w$]. $\mathsf{CBF}_\Box$ is the contraposition of $\mathsf{BF}_\Box$, and expresses the de re modal principle that if something exists, it also possibly exists. Given the possible-world O-semantics, the domains of accessible possible worlds can only grow; one can show that all instances of $\mathsf{CBF}_\Box$ are valid on a frame $\langle W, R\rangle$ iff [for any worlds $w, u \in W$, if $u \in R(w)$, then $D_w \subseteq D_u$]. Necessitists who use modal logic when doing metaphysics accept and defend the conjunction of the schemata $\mathsf{BF}_\Box$ and $\mathsf{CBF}_\Box$. On the contrary, contingentists who use modal logic when doing metaphysics reject and argue against either $\mathsf{BF}_\Box$; $\mathsf{CBF}_\Box$; or both. For more on the necessitism / contingentism debate, see Williamson 2013.

As another example consider the relativism / absolutism debate about quantification, which is relevant to modern philosophy of language / logic, and is closely related to, though subtly different from, the necessitism / contingentism debate. Relativism about quantification naively states that no use of the universal and existential quantifier is unrestricted / absolute, in the sense of ranging over an unrestricted domain of absolutely everything. Absolutism about quantification on the other hand states that some uses of the universal and existential quantifier are unrestricted / absolute. This naively captures the relativist position, because a crucial aspect of the relativist programme in the first place is the requirement to formulate their view in a coherent way. As it stands, we cannot understand naive formulations of relativism, like the one given here, as saying that *absolutely* no use of the quantifier is unrestricted / absolute, i.e. as employing an unrestricted quantifier; such a formulation would precisely be a counterexample to what itself is trying to claim. Thus, naive formulations of relativism must employ a restricted quantifier, as the relativist would have it. However, then it is unclear why there should not be some larger domain where one does have uses of the universal and existential quantifier that are unrestricted / absolute.

Studd 2019 calls this reply to naive formulations of relativism the *objection from ineffability*; and he covers options available to the relativist to reply to this objection. One such option is modal potentialism, which is a position in the philosophy of mathematics that we will cover in more detail in chapter 5. Roughly, potentialism is the view that mathematical objects are generated successively, and that certain generative processes are incompletable. Thus, certain totalities of mathematical objects are indefinitely extensible; for any such totality, new objects can be generated. Modal potentialism explicates potentialism using first-order modal operator logic, taking a modal object language as primitive for mathematics. How the primitive modality should be interpreted is a crucial part of the modal potentialist position and the debate surrounding its coherence as a position; and different modal potentialists have different views regarding how the primitive modality should be interpreted. Modal potentialists tend to accept and defend $\mathsf{CBF}_{\Box}$ as part of their axiomatisation of modal potentialism, and the de re modal principle it expresses. Furthermore, modal potentialists tend to accept and defend the de re modal

principle that there are things that do not exist, but only possibly exist.[6] Thus, the universal quantifier '$\forall$' cannot quantify over 'absolutely everything'. Though Studd uses modal potentialism as a way to explicate relativism, it is important to note that not all modal potentialists are relativists. For example, Linnebo 2018, §3.6 argues that absolutism can be retrieved with the 'modalised quantifier' '$\Box\forall$'. Studd argues that such 'hybrid relativism' faces much of the same criticism the relativist, whether hybrid or 'thoroughgoing' (both Studd's terminology), launches at the absolutist (Studd 2019, §7.5). For more on the relativism / absolutism debate see Studd 2019.

3.2.3 The problem of Stern's intended interpretation

A second argument for why Stern's framework does not provide a conception of the mention predicate approach; and why blocking diagonalisation does not provide a means of addressing the paradoxes of the mention predicate approach; is that they are precisely not syntactic enough. Recall that the mention predicate approach formalises a de dicto modal statement by formulating the relevant modality as a first-order predicate that predicates on names of sentences.[7] In the framework we set up in this thesis, we embrace diagonalisation in the formalisation of the mention predicate approach. On the contrary, Stern's framework blocks diagonalisation in the formalisation of the mention predicate approach. However, given the meta-linguistic nature of the mention predicate approach, is it even philosophically warranted to block diagonalisation?

To answer this question, recall that Stern blocks diagonalisation by ensuring that the class of quotation constant symbols in his framework is structurally too poor to support any syntax theory, precisely to avoid paradox. In effect, when we add the 'quotation constant symbols' in the construction of $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}$, we are simply adding countably many constants which we can biject into the sentences of $(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}$. The rest of Stern's framework lays

[6] To reiterate, this is a very rough sketch of the view, and a lot of important detail has been skipped over. We will cover modal potentialism in more detail in chapter 5, and therefore I contend this rough sketch is adequate for the current purpose of providing an example of how current philosophy often revolves around de re modal statements / principles.

[7] As stated earlier in the previous chapter 2, nothing much hinges on predicating over names of sentences, as opposed to names of propositions. In particular, I also claim the following argument against Stern, if we substitute 'names of sentences' with 'names of propositions'.

no further constraints on what kind of structure this bijection must preserve, precisely because it cannot preserve too much structure, in particular diagonalisation. This shows that $\ulcorner\phi\urcorner$ in a sense is a misleading characterisation of the quotation constant symbols; the object theory itself does not know these constants are of syntax. We could have also introduced constants c_n for $n \in \mathbb{N}$, and because both $\mathsf{Sent}[(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})^{\mathcal{S}}]$ and $\{c_n \mid n \in \mathbb{N}\}$ are countable, we can find an appropriate bijection. Even if we consider Stern's open question of how much syntax theory we can nonetheless introduce in his framework; in such an enriched framework the class of quotation constant symbols must be structurally too poor to support *all* syntax theory, on pain of paradox.

However, looking at Stern's meta-theory and model theory, he *intends* to interpret the quotation constant symbols as *names of sentences*, qua syntactic objects. This is precisely because he intends to provide an account of the mention predicate approach; modality is meta-linguistic, where 'modal notions are best conceived as predicates applicable to names of sentences' (Stern 2015, p. 1). That is, modality predicates on syntactic objects. By the construction of syntax theory as the theory of (the manipulation of) syntactic objects, syntactic objects are a class of objects that supports all syntax theory. Furthermore, it is clear that quotation and substitution are crucial aspects of the manipulation of syntactic objects, and thus syntax theory. For example, one might naturally ask how many letters the sentence 'It is possible for me to have light blond hair' contains. Further, one might naturally ask, if we replace 'light' with 'dark' in the previous sentence, how many letters *that* sentence contains. Intending to predicate on names of sentences commits one to diagonalisation; and more specifically, intending to predicate on names of sentences commits one to representing the quotation and ternary substitution functions defined on these names. Not representing quotation or substitution in a theory of syntax is like doing arithmetic without multiplication. That is, we can think of multiplication as a kind of substitution. For example, multiplying two by three can be seen as taking a plurality of two singulars, substituting each of the two singulars by a plurality of three singulars, and counting how many singulars there are in the final plurality.

Thus, Stern has an intended interpretation of the quotation constant symbols in his meta-theory (they are syntactic objects); but we cannot add sentences to the object theory that are true with respect to said intended interpretation (e.g. sentences representing quotation and substitution); for else the object theory would be inconsistent. It is one thing to not add all true sentences with respect to a certain intended interpretation of some class of objects to some object theory describing them. This is often very desirable, and it is what we do when we generalise or abstract away structure. However, in such a case the point is precisely that adding true sentences with respect to any intended interpretation of some class of objects that satisfies the relevant structure results in a consistent theory. It is another thing entirely when adding certain true sentences with respect to a certain intended interpretation of some class of objects to some object theory describing them results in inconsistency. In particular, it shows that the object theory is inadequate at capturing the behaviour of said class of objects as they are intended to be interpreted.

Therefore, either Stern has provided a fruitful account of a modal predicate that predicates on members of a class of representational objects that are *not* syntactic, *nor* structurally rich enough to support syntax theory, at the very least diagonalisation; or Stern has provided an account of the mention predicate approach that fails to address the paradoxes of the mention predicate approach. That is, contrary to what Stern sets out to do, Stern has not provided an account of the mention predicate approach that succeeds in addressing the paradoxes of the mention predicate approach. It is in this sense that Stern's construction is not syntactic enough.

However, one might wonder whether there is anything constructive we can learn from Stern's framework. In particular, a natural question one might ask is whether we can make sense of a class of representational objects that are not syntactic, nor structurally rich enough to support syntax theory; or at the very least not structurally rich enough to support diagonalisation.

3.2.4 Generalising the use-mention distinction

As introduced in this thesis, the mention predicate approach focusses on linguistic entities as the objects predicated over by modal predicates. However, linguistic entities are not the only representations in philosophy we are interested in. In particular, we can generalise the use-mention distinction, to not just consider expressions of some language, but any class of representations for which we have, or could possibly have, a semantic account. This not only covers representations of reality, but also representations of representations, or representations of fictitious stories, etc. Thus, we can understand using a representation as being concerned with whatever said representation concerns, or 'depicts'. That is, as being concerned with the semantic content of said representation. On the other hand, we can understand mentioning a representation as being concerned with said representation itself, qua being a representation.

For example, consider the following representation: a black and white camera picture of a tree. Assuming some account of the semantics of pictures, such as Abusch 2020 or Malinas 1991; if we are to interpret what is true in this picture, or what it depicts, such as that there is a tree, we are using the picture. On the other hand, we are mentioning the picture when we say that the picture is black and white, or that it is grainy / high definition. As alluded to, the 'picture' need not be interpreted as also providing information about the world. For example, if we are to interpret what is true in *The Unicorn Tapestries* (c. 1495-1505) (a series of seven tapestries showing, among other things, a group of noblemen and hunters pursuing a unicorn), or what it depicts, such as that there is a unicorn, we are using these tapestries. On the other hand, we are mentioning these tapestries when we say that they are woven in wool, metallic threads, and silk; or that they can be found in the Met Cloisters in New York.[8]

[8]One might have a worry that we cannot make a strict distinction between use and mention for a piece of art like a tapestry, or a painting. For example, consider the claim: 'The unicorn in *The Unicorn Tapestries* is white'. Are we using the tapestry, because it depicts a white unicorn; or are we mentioning the tapestry, because the tapestry depicts the unicorn with white dye? Similar ambiguities arise if we consider any other tapestry or painting, and claims about the colour of what they depict, if anything. However, this lack of strict distinction between use and mention for a piece of art is not a quirk unique to the specific class of representations of pieces of art; recall from the introduction that we have a similar lack of strict distinction between use and mention for natural language as well. Thus, there is a broader worry regarding the non-strict distinction between use and mention in general. Moreover, as in the introduction, the fact that

As another example, consider the following representation: a word file saved on a computer. We are using this word file when we claim that it contains the string '12345', or a portion of Shakespeare's *Hamlet*. On the other hand, we are mentioning this word file when we say that it is smaller than 100kb, or that it was created on 17.06.2024. More generally, computer scientists speak of *metadata*, or 'data about data', of some file, which provides precisely all information that mentions said file.

Furthermore, it is not obvious that diagonalisation is unique to syntactic entities, qua representations. In particular, to represent the quotation and ternary substitution functions is necessary for syntax theory, but it is not obvious that it is sufficient. Presumably, a suitable theory of camera pictures can also represent these functions; pictures can occur within pictures, and it is plausible that we can substitute parts of pictures for yet other pictures. In particular, in such a case we would be able to prove diagonalisation. A prime example of the picture equivalent of a self-referential sentence is what is known in art theory as the Droste effect; a picture recursively appearing within itself, where one would expect a picture to appear.[9] This effect is named after a brand of cocoa called Droste, which employed this effect in the packaging of their products. We can similarly see the Droste effect in the print *Print Gallery* (1956) by M.C. Escher (1956). The Droste effect in turn is an example of a *mise en abyme*, which generalises the effect to film and literary theory as well. Thus, it is a non-trivial question how much structure representations of non-syntactic classes have.

Recall that the linguistic entities that the modal predicate in the mention predicate approach predicates on are names of the complements of modal statements involving

---

the use-mention distinction in general is not a strict one makes it clear that it is sometimes a difficult choice which of the operator, use predicate, and mention predicate approaches to use when formalising modal statements. This worry must clearly be addressed; however, addressing this worry lies beyond the scope of this thesis, and will be set aside for future work.

[9]Note that it is more apt to call the Droste effect the picture equivalent of a *self-referential* sentence, as opposed to a *diagonal* sentence. Technically, any provable sentence where we do not employ substitution as in the Diagonal Lemma is non-self-referential (such as the axioms of PA), and is a diagonal sentence for the provability predicate $\mathbf{Bew}(x)$. However, it is clear that the Droste effect is more than just a fixed point *tout court*; rather, it is a fixed point of a special kind that refers to itself. That said, I do not aim here to give an account of the notion of self-reference (or what special kind of fixed point the Droste effect is); see instead Grabmayr et al. 2023, Halbach and Visser 2014a; Halbach and Visser 2014b, and Picollo 2018; Picollo 2020.

the modality being formalised. However, there might also be modalities with intended interpretations most appropriately formalised using a modal predicate that mentions members of some other class of representations for which we have, or could possibly have, a semantic account. Call such modalities representational modalities. These representations would be related in some appropriate way to the complements of modal statements involving a representational modality being formalised. It is important to flag that this is very exploratory, and I have no examples of representational modalities. However, as a theoretical exercise, this suggests the mention predicate approach could be generalised such that modal predicates mention members of any class of representations for which we have, or could possibly have, a semantic account.

As a first attempt at this generalisation, let us go back to Stern's construction in (2014); (2015). Assume that we are trying to formalise a representational modality. Further assume that we have some theory of (the structure of) the representations being mentioned, and some associated semantic account. Recall that diagonalisation plays the central role in the (limitative) results we proved at the end of chapter 2, and will prove in the chapters 4 and 5. Therefore, if we further assume that these representations in question are sufficiently structured so that we could prove diagonalisation on them; then for a unary modal predicate mentioning these representations we would also be able to prove analogues of the paradoxes of diagonalisation and the results proved in this thesis. In particular, we see that nothing much hinges on predicating over names of sentences as opposed to names of propositions, so long as the names of propositions are sufficiently structured, i.e. prove diagonalisation. Thus, let us instead lastly assume that these representations are structurally too poor to prove diagonalisation on them; and that a unary predicate is a suitable formalisation of the representational modality.[10] Then Stern's framework points towards, or suggests the feasibility of formalising said representational modality such that we can avoid paradox and consistently preserve all modal principles also available to the operator approach. This is a constructive takeaway of Stern's framework.

[10]Taking into account the intended interpretation of the representational modality, and whether there are analogues of de dicto and de re modal statements involving said representational modality.

To sum up this chapter, we considered what happens if we give up on the Diagonal Lemma and why this does not provide a means of addressing the paradoxes of the mention predicate approach. We first considered an example of a framework that gives up on the Diagonal Lemma; Stern 2014c; Stern 2015 provides an account of the mention predicate approach that blocks diagonalisation and constructs a class of possible-world models for this account that satisfies the Adequacy Condition. We then saw that Stern's framework, and blocking diagonalisation more generally, does not succeed in providing a uniform strategy for avoiding the paradoxes of the mention predicate approach. I further argued that diagonalisation is a necessary aspect of any formalisation of the mention predicate approach. We thus concluded that Stern's framework does not provide a conception of the mention predicate approach; and that blocking diagonalisation does not provide a means of addressing the paradoxes of the mention predicate approach. We finished this chapter by tentatively observing that we can nonetheless use Stern's framework when formalising a representational modality that mentions a class of representations that are structurally too poor to prove diagonalisation on them.

4

Categorising The Paradoxes Of The Mention Predicate Approach

Contents

At the end of chapter 2 we observed that the four components involved in Montague's Paradox are instances of $\mathsf{T_N}$, $\mathsf{NEC_N}$, the Diagonal Lemma, and Classical logic. In the previous chapter 3 we concluded that giving up on the Diagonal Lemma is not an option. Instead, in this thesis diagonalisation is fully embraced, playing the central role in the (limitative) results we proved at the end of chapter 2, and will prove in this chapter 4 and the following chapter 5. So where do we go from here?

4.1 More paradoxes: unary and merging

So far, we have only been focussing on Montague's Paradox. However, as alluded to in the introduction, the mention predicate approach is prone to a whole family of paradoxes, of which Montague's Paradox is the most notable; and many of these paradoxes involve conditions other than $\mathsf{T_N}$ or $\mathsf{NEC_N}$. Consider for example the following paradox unrelated to Montague's Paradox, which is a generalisation of Gödel's original Second Incompleteness Theorem (1931), and a reformulation of Löb's Theorem (1955).[1]

Theorem 4.1.1 (Gödel-Löb)**.** *Let* E *be a theory that meets the following conditions:*

i. E *contains* $\mathsf{PA_N}$.

ii. E *contains the schemata*

$$\mathbf{N}\overline{\ulcorner \phi \rightarrow \psi \urcorner} \rightarrow (\mathbf{N}\overline{\ulcorner \phi \urcorner} \rightarrow \mathbf{N}\overline{\ulcorner \psi \urcorner}) \tag{$\mathsf{K_N}$}$$

$$\mathbf{N}\overline{\ulcorner \phi \urcorner} \rightarrow \neg \mathbf{N}\overline{\ulcorner \neg \phi \urcorner} \tag{$\mathsf{D_N}$}$$

$$\mathbf{N}\overline{\ulcorner \phi \urcorner} \rightarrow \mathbf{N}\overline{\ulcorner \mathbf{N}\overline{\ulcorner \phi \urcorner} \urcorner} \tag{$\mathsf{4_N}$}$$

where $\phi, \psi \in \mathsf{Sent}(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})$.[2]

iii. E *is closed under the necessitation rule*

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}\overline{\ulcorner \phi \urcorner} \tag{$\mathsf{NEC_N}$}$$

where $\phi \in \mathsf{Sent}(\mathcal{L}^{\mathbf{N}}_{\mathsf{PA}})$.

Then E *is inconsistent.*

Proof. By the Diagonal Lemma there is a λ such that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg \mathbf{N}\overline{\ulcorner \lambda \urcorner} \tag{4.1}$$

[1] See also Friedman and Sheard 1987, p. 14. In particular, they prove that T-Cons, T-Rep, and T-intro are jointly inconsistent; where T-Cons is $\mathsf{D_N}$, T-Rep is $\mathsf{4_N}$, T-Intro is $\mathsf{NEC_N}$, and their base theory contains A10, which is $\mathsf{K_N}$.

[2] The $\mathsf{K_N}$-, $\mathsf{D_N}$-, and $\mathsf{4_N}$-schemata are the predicate analogues of the schemata of the same name in the operator approach.

We then have the following:

$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}$	(4.1)	(4.2)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda \leftrightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner}$	$\mathsf{NEC}_{\mathbf{N}}$, (4.2)	(4.3)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner} \leftrightarrow \mathbf{N}\overline{\ulcorner\neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner}$	$\mathsf{K}_{\mathbf{N}}$, (4.3)	(4.4)
$\mathsf{E} \vdash \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner} \leftrightarrow \neg\mathbf{N}\overline{\ulcorner\neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner}$	(4.4)	(4.5)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner} \rightarrow \neg\mathbf{N}\overline{\ulcorner\neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner}$	$\mathsf{D}_{\mathbf{N}}$	(4.6)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner} \rightarrow \mathbf{N}\overline{\ulcorner\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner}$	$\mathsf{4}_{\mathbf{N}}$	(4.7)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner} \rightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}$	(4.7), (4.6), (4.5)	(4.8)
$\mathsf{E} \vdash \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}$	(4.8)	(4.9)
$\mathsf{E} \vdash \lambda$	(4.9), (4.2)	(4.10)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner}$	$\mathsf{NEC}_{\mathbf{N}}$, (4.10)	(4.11)

Lines (4.9) and (4.11) imply E is inconsistent. Q.E.D.

Call paradoxes involving a single unary modal syntactic predicate unary paradoxes. Thus, Montague's Paradox and Theorem 4.1.1 (Gödel-Löb) are unary paradoxes. However, unary paradoxes are not the only paradoxes; there are also paradoxes involving multiple distinct unary modal syntactic predicates, call these merging paradoxes. The merging paradoxes have so far received little attention in the literature, even though they are just as important, if not more important, to avoid as the unary paradoxes. They are arguably more important to avoid, since the merging paradoxes are more general than the unary paradoxes. It is clear that any framework able to formulate merging paradoxes is also able to formulate unary paradoxes, and thus the merging paradoxes are at least as rich in kind as the unary paradoxes. Furthermore, the merging paradoxes are strictly richer in kind than the unary paradoxes. That is, there are merging paradoxes that are truly 'merging', where no paradox arises if we focus on the fragments involving only a single unary modal syntactic predicate.[3]

[3]For more on when merging paradoxes are truly 'merging', see Fischer and Stern 2015.

For example, similarly to how the typed hierarchy of truth of Tarski 1983 is a way of avoiding Tarski's Undefinability of Truth / the Liar Paradox, Montague's Paradox can be avoided by typing, where the restriction of $\mathsf{T}_{\mathbf{N}}$ and $\mathsf{NEC}_{\mathbf{N}}$ to formulae not containing $\mathbf{N}$ results in a consistent theory. However, paradox can arise yet again if we are not careful how we generalise such typing when considering multiple distinct unary modal syntactic predicates.

Definition 4.1.1. Let $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\,\mathbf{N}_2}$ denote the language that expands $\mathcal{L}_{\mathsf{PA}}$ with the unary predicates $\mathbf{N}_1$ and $\mathbf{N}_2$. Furthermore, let $\mathsf{PA}_{\mathbf{N}_1,\,\mathbf{N}_2}$ denote the theory of the axioms of PA formulated in $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\,\mathbf{N}_2}$.

Theorem 4.1.2 (Naive Typing of Montague's Paradox)**.** *This Theorem is a piece of folklore. See for instance Halbach 2006. Let* E *be a theory that meets the following conditions:*

i. E *contains* $\mathsf{PA}_{\mathbf{N}_1,\,\mathbf{N}_2}$.

ii. E *contains the schema*

$$\mathbf{N}_1\overline{\ulcorner\phi\urcorner} \rightarrow \phi \tag{$\mathsf{T}_{\mathbf{N}}$}$$

and is closed under the necessitation rule

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\phi\urcorner} \tag{$\mathsf{NEC}_{\mathbf{N}_1}$}$$

where for both $\phi \in \mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_2})$.

iii. E *contains the schema*

$$\mathbf{N}_2\overline{\ulcorner\phi\urcorner} \rightarrow \phi \tag{$\mathsf{T}_{\mathbf{N}_2}$}$$

and is closed under the necessitation rule

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}_2\overline{\ulcorner\phi\urcorner} \tag{$\mathsf{NEC}_{\mathbf{N}_2}$}$$

where for both $\phi \in \mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1})$.

Then E *is inconsistent. However, the theory* E_1 *of* $\mathsf{PA}_{\mathbf{N}_1,\,\mathbf{N}_2}$ *extended with* $\mathsf{T}_{\mathbf{N}_1}$ *and closed under the necessitation rule* $\mathsf{NEC}_{\mathbf{N}_1}$*, where both conditions are restricted to formulae in* $\mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_2})$*; and the theory* E_2 *of* $\mathsf{PA}_{\mathbf{N}_1,\,\mathbf{N}_2}$ *extended with* $\mathsf{T}_{\mathbf{N}_2}$ *and closed under the*

necessitation rule $\mathsf{NEC}_{\mathbf{N}_2}$*, where both conditions are restricted to formulae in* $\mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1})$*; are both consistent. Moreover, the theory* E' *of* $\mathsf{PA}_{\mathbf{N}_1,\,\mathbf{N}_2}$ *extended with* $\mathsf{T}_{\mathbf{N}_1}$ *and* $\mathsf{T}_{\mathbf{N}_2}$ *and closed under the necessitation rules* $\mathsf{NEC}_{\mathbf{N}_1}$ *and* $\mathsf{NEC}_{\mathbf{N}_2}$*, where all four conditions are restricted to formulae not containing either* $\mathbf{N}_1$ *or* $\mathbf{N}_2$*, is consistent.*

Proof. The well known consistency of the typed restriction of the **T**-schema to formulae not containing **T** has as corollaries the consistencies of E_1 and E_2, where **T** is replaced by $\mathbf{N}_1$ and $\mathbf{N}_2$ respectively. The consistency of E' follows in a similar fashion from the consistency of combining multiple **T**-schemata for distinct predicates $\mathbf{T}_i$, each restricted to not containing any of the $\mathbf{T}_j$.[4]

For the inconsistency of E, by the Diagonal Lemma there is a λ such that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1(\overline{\ulcorner\mathbf{N}_2\urcorner}\mathbin{\dot{\frown}}\dot{\mathsf{q}}(\overline{\ulcorner\lambda\urcorner})) \tag{4.12}$$

Because $\dot{\mathsf{q}}$ represents the quotation function, from (4.12) we have that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1(\overline{\ulcorner\mathbf{N}_2\urcorner}\mathbin{\dot{\frown}}\overline{\ulcorner\overline{\ulcorner\lambda\urcorner}\urcorner}) \tag{4.13}$$

Because $\dot{\frown}$ represents the binary concatenation function, from (4.13) we have that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \tag{4.14}$$

We then have the following:

$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner}$	(4.14)	(4.15)
$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \rightarrow \mathbf{N}_2\overline{\ulcorner\lambda\urcorner}$	$\mathsf{T}_{\mathbf{N}_1}$	(4.16)
$\mathsf{E} \vdash \mathbf{N}_2\overline{\ulcorner\lambda\urcorner} \rightarrow \lambda$	$\mathsf{T}_{\mathbf{N}_2}$	(4.17)
$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \rightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner}$	(4.16), (4.17), (4.15)	(4.18)
$\mathsf{E} \vdash \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner}$	(4.18)	(4.19)
$\mathsf{E} \vdash \lambda$	(4.19), (4.15)	(4.20)
$\mathsf{E} \vdash \mathbf{N}_2\overline{\ulcorner\lambda\urcorner}$	$\mathsf{NEC}_{\mathbf{N}_2}$, (4.20)	(4.21)

[4]For these consistencies see for instance Halbach 2006.

$$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad \mathsf{NEC}_{\mathbf{N}_1}, (4.21) \qquad (4.22)$$

Lines (4.19) and (4.22) imply E is inconsistent. Q.E.D.

The Naive Typing of Montague's Paradox shows us that we cannot 'naively' type in the same way for multiple distinct unary modal syntactic predicates like we do for a single such predicate. The problem is that, where we for example do not allow $\mathbf{N}_1$ to occur in ϕ in $\mathsf{NEC}_{\mathbf{N}_1}$ and $\mathsf{T}_{\mathbf{N}_1}$; $\mathbf{N}_1$ can still be 'mentioned' in ϕ in $\mathsf{NEC}_{\mathbf{N}_1}$ and $\mathsf{T}_{\mathbf{N}_1}$, by being nested under multiple quotations. This is precisely what is being exploited in the inconsistency proof above, where $\mathbf{N}_1$ occurs in λ; and $\mathbf{N}_1$ is merely mentioned in $\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}$ and $\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner}$, allowing us to use $\mathsf{NEC}_{\mathbf{N}_1}$ and $\mathsf{T}_{\mathbf{N}_1}$ in lines (4.22) and (4.16) respectively. There are different ways one could nonetheless appropriately type in the multimodal setting, so as to avoid the above 'mentioning' phenomenon and to avoid paradox. For instance, one could form a Tarskian hierarchy of predicates by inductively introducing $\mathbf{N}_1^n$ and $\mathbf{N}_2^n$ simultaneously, for $n \in \mathbb{N}$; where $\mathbf{N}_1^n\overline{\ulcorner\phi\urcorner}$ and $\mathbf{N}_2^n\overline{\ulcorner\phi\urcorner}$ are well-formed iff ϕ does not contain $\mathbf{N}_1^l$ or $\mathbf{N}_2^l$ for $l \geq n$. For more on how to succesfully type multiple distinct unary syntactic predicates see Paseau 2009; which is a reply to Halbach 2008; which itself is a reply to Paseau 2008.

Appropriately typing is one way in which Montague's Paradox can be avoided in the multimodal setting. Such typing involves restricting factivity, for example by restricting $\mathsf{T}_{\mathbf{N}_1}$ to formulae not containing $\mathbf{N}_1$ or $\mathbf{N}_2$. However, one might hope that there is another way of avoiding Montague's Paradox in the multimodal setting that does not come at the cost of restricting factivity. In particular, one might hope that the combination of multiple unary modal syntactic predicates recovers unrestricted factivity, thereby merely weakening as opposed to abandoning unrestricted factivity. For example, one might argue that if something is apriori true, then it is the case. However, we still run into paradox.

Theorem 4.1.3 (Bimodal Montague's Paradox)**.** *This Theorem is a syntactic predicate version of Fischer and Stern 2015, Theorem 11. Let* E *be a theory that meets the following conditions:*

i. E *contains* $\mathsf{PA}_{\mathbf{N}_1,\,\mathbf{N}_2}$.

ii. E *contains the schema*

$$\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\phi\urcorner}\urcorner} \rightarrow \phi \tag{BT}$$

where $\phi \in \mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\,\mathbf{N}_2})$.

iii. E *is closed under the necessitation rules*

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\phi\urcorner} \tag{$\mathsf{NEC}_{\mathbf{N}_1}$}$$

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}_2\overline{\ulcorner\phi\urcorner} \tag{$\mathsf{NEC}_{\mathbf{N}_2}$}$$

where $\phi \in \mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\,\mathbf{N}_2})$.

Then E *is inconsistent.*[5]

Proof. Just like in the proof of the Naive Typing of Montague's Paradox, by the Diagonal Lemma there is a λ such that

$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner}$	(4.14)	(4.23)
$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \rightarrow \lambda$	(BT)	(4.24)
$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \rightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner}$	(4.24), (4.23)	(4.25)
$\mathsf{E} \vdash \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner}$	(4.25)	(4.26)
$\mathsf{E} \vdash \lambda$	(4.26), (4.23)	(4.27)
$\mathsf{E} \vdash \mathbf{N}_2\overline{\ulcorner\lambda\urcorner}$	$\mathsf{NEC}_{\mathbf{N}_2}$, (4.27)	(4.28)
$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner}$	$\mathsf{NEC}_{\mathbf{N}_1}$, (4.28)	(4.29)

Lines (4.26) and (4.29) imply E is inconsistent. Q.E.D.

Furthermore, the Bimodal Montague's Paradox can be generalised to any finite combination of unary modal syntactic predicates.

[5]Similarly, E is inconsistent if (BT) is replaced with

$$\mathbf{N}_2\overline{\ulcorner\mathbf{N}_1\overline{\ulcorner\phi\urcorner}\urcorner} \rightarrow \phi$$

Definition 4.1.2. Let $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$ denote the language that expands $\mathcal{L}_{\mathsf{PA}}$ with $n \in \mathbb{N}$ distinct unary predicates $\mathbf{N}_i$, for $1 \leq i \leq n$.

Theorem 4.1.4 (Multimodal Montague's Paradox)**.** *Let* E *be a theory that meets the following conditions:*

i. E *contains the axioms of* PA *formulated in* $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$.

ii. E *contains the schema*

$$\mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\overline{\phi}\urcorner^{\cdots}}\urcorner\urcorner \rightarrow \phi \tag{MT_n}$$

where $\phi \in \mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$.

iii. E *is closed under the necessitation rules*

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}_i\ulcorner\overline{\phi}\urcorner \tag{$\mathsf{NEC}_{\mathbf{N}_i}$}$$

where $\phi \in \mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$*, for* $1 \leq i \leq n$.

Then E *is inconsistent.*[6]

Proof. By the Diagonal Lemma, we can find a λ such that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1(\overline{\ulcorner\mathbf{N}_2\urcorner}\mathbin{\dot{\frown}}\dot{\mathsf{q}}(\ulcorner\mathbf{N}_3(\overline{\ulcorner\mathbf{N}_4\urcorner}\mathbin{\dot{\frown}}\dot{\mathsf{q}}(\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner^{\cdots}}\urcorner}))\urcorner)) \tag{4.30}$$

As in the proof of the Naive Typing of Montague's Paradox, because $\dot{\mathsf{q}}$ represents the quotation function, from (4.30) we have that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1(\overline{\ulcorner\mathbf{N}_2\urcorner}\mathbin{\dot{\frown}}\ulcorner\ulcorner\mathbf{N}_3(\overline{\ulcorner\mathbf{N}_4\urcorner}\mathbin{\dot{\frown}}\dot{\mathsf{q}}(\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner^{\cdots}}\urcorner}))\urcorner\urcorner) \tag{4.31}$$

Because $\mathbin{\dot{\frown}}$ represents the binary concatenation function, from (4.31) we have that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\mathbf{N}_3(\overline{\ulcorner\mathbf{N}_4\urcorner}\mathbin{\dot{\frown}}\dot{\mathsf{q}}(\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner^{\cdots}}\urcorner}))\urcorner\urcorner \tag{4.32}$$

By alternating these two steps repeatedly, i.e. alternating the facts that $\dot{\mathsf{q}}$ and $\mathbin{\dot{\frown}}$ represent the quotation and binary concatenation functions respectively, from (4.32) we get that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner^{\cdots}}\urcorner\urcorner \tag{4.33}$$

[6]Similarly E is inconsistent for any other permutation of the $\mathbf{N}_i$ in (MT_n).

For the rest of this thesis I will not make these alternating steps explicit, and simply say that by the Diagonal Lemma we can e.g. find a λ such that we have (4.33); though technically it is more subtle than that.

We then have the following:

$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner}^{\cdots}\urcorner}\urcorner}$	(4.33)	(4.34)
$\mathsf{E} \vdash \neg\lambda \leftrightarrow \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner}^{\cdots}\urcorner}\urcorner}$	(4.34)	(4.35)
$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner}^{\cdots}\urcorner}\urcorner} \rightarrow \lambda$	MT_n	(4.36)
$\mathsf{E} \vdash \neg\lambda \rightarrow \lambda$	(4.35), (4.36)	(4.37)
$\mathsf{E} \vdash \lambda$	(4.37)	(4.38)
$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner}^{\cdots}\urcorner}\urcorner}$	$\mathsf{NEC}_{\mathbf{N}_i}$, for $1 \leq i \leq n$; (4.38)	(4.39)
$\mathsf{E} \vdash \neg\lambda$	(4.35), (4.39)	(4.40)

Lines (4.38) and (4.40) imply E is inconsistent. Q.E.D.

Montague's Paradox shows us that assuming unrestricted factivity and closure under the necessitation rule for a single unary modal syntactic predicate leads to paradox. The Bimodal Montague's Paradox shows us that assuming the combination of two unary modal syntactic predicates for unrestricted factivity and closure under the necessitation rules for both predicates leads to paradox. The Multimodal Montague's Paradox then shows us that this can be generalised to any finite combination of multiple unary modal syntactic predicates, assuming closure under the necessitation rules for all of the involved modal syntactic predicates. That is, given classical logic no finite combination of multiple unary modal syntactic predicates can be a sufficient condition for unrestricted factivity, and thus there is no hope in the multimodal setting to recover unrestricted factivity.

Furthermore, as Theorem 4.1.1 (Gödel-Löb) shows in the case of unary paradoxes, there are merging paradoxes unrelated to the merging paradoxes that generalise Montague's Paradox; either by naively typing in the multimodal setting as in the Naive Typing of Montague's Paradox; or by weakening factivity as in the Multimodal Montague's

Paradox. Consider for example the following paradox, which is a bimodal generalisation of Theorem 4.1.1 (Gödel-Löb).[7]

Theorem 4.1.5 (Bimodal Gödel-Löb)**.** *Let* E *be a theory that meets the following conditions:*

i. E *contains* $\mathsf{PA}_{\mathbf{N}_1,\,\mathbf{N}_2}$.

ii. E *contains the schemata*

$$\mathbf{N}_1\overline{\ulcorner\phi \rightarrow \psi\urcorner} \rightarrow (\mathbf{N}_1\overline{\ulcorner\phi\urcorner} \rightarrow \mathbf{N}_1\overline{\ulcorner\psi\urcorner}) \qquad (\mathsf{K}_{\mathbf{N}_1})$$

$$\mathbf{N}_2\overline{\ulcorner\phi \rightarrow \psi\urcorner} \rightarrow (\mathbf{N}_2\overline{\ulcorner\phi\urcorner} \rightarrow \mathbf{N}_2\overline{\ulcorner\psi\urcorner}) \qquad (\mathsf{K}_{\mathbf{N}_2})$$

where $\phi, \psi \in \mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\,\mathbf{N}_2})$.

iii. E *contains the schemata*

$$\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\phi\urcorner}\urcorner} \rightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\neg\phi\urcorner}\urcorner} \qquad (\mathsf{D}_{\langle\mathbf{N}_1,\mathbf{N}_2\rangle})$$

$$\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\phi\urcorner}\urcorner} \rightarrow \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\phi\urcorner}\urcorner}\urcorner}\urcorner} \qquad (4_{\langle\mathbf{N}_1,\mathbf{N}_2\rangle})$$

where $\phi, \psi \in \mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\,\mathbf{N}_2})$.[8]

iv. E *is closed under the necessitation rules*

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\phi\urcorner} \qquad (\mathsf{NEC}_{\mathbf{N}_1})$$

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}_2\overline{\ulcorner\phi\urcorner} \qquad (\mathsf{NEC}_{\mathbf{N}_2})$$

where $\phi \in \mathsf{Sent}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\,\mathbf{N}_2})$.

[7] Note, however, that these two paradoxes are, in a sense, genuinely distinct, even though they are structurally similar. In particular, if we look at the (bi)modal *operator* logic versions of the conditions from Theorems 4.1.1 (Gödel-Löb) and 4.1.5 (Bimodal Gödel-Löb), we see that they place different constraints independent of each other on the accessibility relations of possible-world models making these conditions true.

[8] The $\mathsf{D}_{\langle\mathbf{N}_1,\mathbf{N}_2\rangle}$ and $4_{\langle\mathbf{N}_1,\mathbf{N}_2\rangle}$-schemata are so-called because they are structurally similar to the $\mathsf{D}_{\mathbf{N}}$ and $4_{\mathbf{N}}$-schemata, except with '$\mathbf{N}$' replaced by the ordered combination of '$\mathbf{N}_1$' and '$\mathbf{N}_2$'.

Then E *is inconsistent.*[9] [10]

Proof. By the Diagonal Lemma there is a λ such that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad \text{DL} \quad (4.41)$$

$$\mathsf{E} \vdash \mathbf{N}_2\ulcorner\lambda \leftrightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad \mathsf{NEC}_{\mathbf{N}_2}, (4.41) \quad (4.42)$$

$$\mathsf{E} \vdash \mathbf{N}_2\overline{\ulcorner\lambda\urcorner} \leftrightarrow \mathbf{N}_2\ulcorner\neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad \mathsf{K}_{\mathbf{N}_2}, (4.42) \quad (4.43)$$

$$\mathsf{E} \vdash \mathbf{N}_1\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner} \leftrightarrow \mathbf{N}_2\ulcorner\neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad \mathsf{NEC}_{\mathbf{N}_1}, (4.43) \quad (4.44)$$

$$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \leftrightarrow \mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad \mathsf{K}_{\mathbf{N}_1}, (4.44) \quad (4.45)$$

$$\mathsf{E} \vdash \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \leftrightarrow \neg\mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad (4.45) \quad (4.46)$$

$$\mathsf{E} \vdash \mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \rightarrow \neg\mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad \mathsf{D}_{\langle\mathbf{N}_1,\mathbf{N}_2\rangle} \quad (4.47)$$

$$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \rightarrow \mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad \mathsf{4}_{\langle\mathbf{N}_1,\mathbf{N}_2\rangle} \quad (4.48)$$

$$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \rightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad (4.48), (4.47), (4.46) \quad (4.49)$$

$$\mathsf{E} \vdash \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad (4.49) \quad (4.50)$$

$$\mathsf{E} \vdash \lambda \qquad (4.49), (4.41) \quad (4.51)$$

$$\mathsf{E} \vdash \mathbf{N}_2\overline{\ulcorner\lambda\urcorner} \qquad \mathsf{NEC}_{\mathbf{N}_2}, (4.51) \quad (4.52)$$

$$\mathsf{E} \vdash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\lambda\urcorner}\urcorner} \qquad \mathsf{NEC}_{\mathbf{N}_1}, (4.52) \quad (4.53)$$

Lines (4.50) and (4.53) imply E is inconsistent. Q.E.D.

Finally, it is important to note that the previous five paradoxes are not exhaustive, with the mention predicate approach susceptible to many more. See for instance Friedman and Sheard 1987.

[9] Similarly E is inconsistent if $\mathsf{D}_{\langle\mathbf{N}_1,\mathbf{N}_2\rangle}$ and $\mathsf{4}_{\langle\mathbf{N}_1,\mathbf{N}_2\rangle}$ are replaced by

$$\mathbf{N}_2\overline{\ulcorner\mathbf{N}_1\overline{\ulcorner\phi\urcorner}\urcorner} \rightarrow \neg\mathbf{N}_2\overline{\ulcorner\mathbf{N}_1\overline{\ulcorner\neg\phi\urcorner}\urcorner} \qquad (\mathsf{D}_{\langle\mathbf{N}_2,\mathbf{N}_1\rangle})$$

$$\mathbf{N}_2\overline{\ulcorner\mathbf{N}_1\overline{\ulcorner\phi\urcorner}\urcorner} \rightarrow \mathbf{N}_2\ulcorner\mathbf{N}_1\ulcorner\mathbf{N}_2\overline{\ulcorner\mathbf{N}_1\overline{\ulcorner\phi\urcorner}\urcorner} \qquad (\mathsf{4}_{\langle\mathbf{N}_2,\mathbf{N}_1\rangle})$$

[10] Just like for the Bimodal Montague's Paradox, Theorem 4.1.5 (Bimodal Gödel-Löb) can be appropriately generalised to show for any $n \in \mathbb{N}$ that, assuming $\mathsf{K}_{\mathbf{N}_i}$ and $\mathsf{NEC}_{\mathbf{N}_i}$ for all $1 \leq i \leq n$, the combination of $\mathsf{D}_{\langle\mathbf{N}_1,\ldots,\mathbf{N}_n\rangle}$ and $\mathsf{4}_{\langle\mathbf{N}_1,\ldots,\mathbf{N}_n\rangle}$ leads to paradox.

Although giving up on (the problematic instances of) $\mathsf{T}_{\mathbf{N}}$ avoids Montague's Paradox; the previous five paradoxes show us that this does not avoid all of the paradoxes that the mention predicate approach is prone to. Theorem 4.1.1 (Gödel-Löb) shows us that there are other *unary paradoxes* not involving $\mathsf{T}_{\mathbf{N}}$ at all. The Naive Typing of Montague's Paradox shows us that giving up on the problematic instances of $\mathsf{T}_{\mathbf{N}}$ that avoid Montague's Paradox in the unimodal case still leads to paradox in the multimodal case; instead we need to restrict any condition to formulae containing *no* modal syntactic predicates, as opposed to just the modal syntactic predicates occurring in said condition. The Bimodal and Multimodal Montague's Paradoxes show us that, assuming closure under the relevant necessitation rules, weakening factivity to the combination of multiple unary modal syntactic predicates also still leads to paradox. Finally, Theorem 4.1.5 (Bimodal Gödel-Löb) shows us that there are other *merging paradoxes* not involving $\mathsf{T}_{\mathbf{N}_i}$ or MT_n at all, for any $i, n \in \mathbb{N}$. These five paradoxes, along with all the other paradoxes that the mention predicate approach is susceptible to, suggest that a more unified strategy is necessary for avoiding paradox. Despite the fact that in the previous chapter 3 I argued that giving up on the Diagonal Lemma is not a viable means of addressing the paradoxes of the mention predicate approach; it does at the very least have the spirit of providing a unified strategy for avoiding paradox.

Furthermore, the previous five paradoxes make clear that a necessary condition for providing a unified strategy for avoiding paradox, in particular for avoiding the unary and the merging paradoxes, is to first categorise when paradox arises, and when it does not. The rest of this chapter will explore this question of categorisation. To do this, we will take a step away from the proof-theoretic perspective we have been considering so far, and introduce a possible-world predicate-semantics for the mention predicate approach. In particular, we will generalise the classical possible-world P-semantics developed by Halbach and Leigh 2024 to multiple distinct unary modal syntactic predicates. We will then consider some examples of merging paradoxes in this generalised classical possible-world P-semantics. We will also show that when expanding the classical possible-world P-semantics to a classical two-dimensional possible-world P-semantics, we cannot have the expected truth-conditions for well known indexicals of two-dimensional semantics

conceived of as syntactic predicates. Finally, we will prove a limitative result which will function as a categorisation theorem indicating when paradox arises in this model theory.

4.2 Possible-world predicate-semantics

Recall the model-theoretic semantics of classical logic introduced in chapter 2, including the definitions of model, truth in a model, validity, and logical consequence. To construct the possible-world P-semantics, we will appropriately build on this model-theoretic semantics.

However, we must first provide the base theory that by means of soundness we will have in the background of the possible-world P-semantics.

4.2.1 The base theory: Modal Syntactic Predicate Logic

Definition 4.2.1 (The Language of Modal Syntactic Predicate Logic)**.** We expand the language $\mathcal{L}_{\mathsf{PA}}$ with n distinct unary predicates $\mathbf{N}_i$, for $1 \leq i \leq n$. We will treat the $\mathbf{N}_i$ as modal syntactic predicates. Furthermore, we expand $\mathcal{L}_{\mathsf{PA}}$ with the zero-place predicate symbol p. By abuse of notation, call this language $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$.

We add the predicate symbol p to act as a freely-interpretable contingent predicate. We add a single, zero-place predicate symbol as our contingent vocabulary to keep our account as simple as possible, and whatever results we prove will be as general as possible; they will also hold if we add more complex contingent vocabulary. In addition, as we shall see, adding p also avoids definability constraints of PA. We also let '$\mathbf{P}_i x$' be the dual of '$\mathbf{N}_i x$', and let it abbreviate '$\neg \mathbf{N}_i \overline{\ulcorner \neg \urcorner}^{\frown}_{\bullet} x$', for $1 \leq i \leq n$.

Definition 4.2.2 (Modal Syntactic Predicate Logic)**.** Let $(\mathsf{MSPL})^{-}$ denote the theory of:

i. The axioms of PA formulated in $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$.

ii. The axiom schemata

$$\mathbf{N}_i \overline{\ulcorner \phi \rightarrow \psi \urcorner} \rightarrow (\mathbf{N}_i \overline{\ulcorner \phi \urcorner} \rightarrow \mathbf{N}_i \overline{\ulcorner \psi \urcorner}) \qquad (\mathsf{K}_{\mathbf{N}_i})$$

where $\phi, \psi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$, for each $1 \leq i \leq n$.

Our base theory Modal Syntactic Predicate Logic, denoted MSPL, is the closure of $(\mathsf{MSPL})^-$ under the necessitation rules

$$\mathsf{MSPL} \vdash \phi \Rightarrow \mathsf{MSPL} \vdash \mathbf{N}_i\overline{\ulcorner\phi\urcorner} \tag{$\mathsf{NEC}_{\mathbf{N}_i}$}$$

where $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$, for each $1 \leq i \leq n$.[11] We define this closure of $(\mathsf{MSPL})^-$ inductively as follows:

$$(\mathsf{MSPL})_0^- := (\mathsf{MSPL})^- \tag{4.54}$$

$$(\mathsf{MSPL})_{k+1}^- := \mathsf{Th}[(\mathsf{MSPL})_k^- \cup \{\mathbf{N}_i\overline{\ulcorner\phi\urcorner} \mid (\mathsf{MSPL})_k^- \vdash \phi,\ 1 \leq i \leq n\}] \tag{4.55}$$

$$\mathsf{MSPL} := \bigcup_{k\in\mathbb{N}} (\mathsf{MSPL})_k^- \tag{4.56}$$

Furthermore, in the rest of this chapter we will abuse notation and say that a theory E extends MSPL iff $\mathsf{MSPL} \subseteq \mathsf{E}$ and E is closed under the $\mathsf{NEC}_{\mathbf{N}_i}$ rules. Therefore, when we talk about a theory E′ extending MSPL with some $S \subseteq \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$, we also assume that E′ is closed under the $\mathsf{NEC}_{\mathbf{N}_i}$ rules.

All the diagonalisation results from chapter 2 are provable in MSPL. Likewise, extensions of MSPL are susceptible to the paradoxes of the mention predicate approach; such as Montague's Paradox, and the paradoxes considered earlier in this chapter.

4.2.2 Possible-world predicate-models

We will now generalise the classical possible-world P-semantics developed by Halbach and Leigh 2024 to multiple distinct unary modal syntactic predicates. To construct the (multimodal) possible-world P-models, we must first provide the possible worlds themselves. For these, we restrict ourselves to $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$-models

$$\mathcal{M} = \langle \mathbb{N}, V, \mathbf{N}_1^{\mathcal{M}}, \ldots, \mathbf{N}_n^{\mathcal{M}} \rangle$$

of MSPL which expand the standard model $\mathbb{N}$ of PA, where; V is the interpretation of p and thus assigns either *0* or *1*; and $\mathbf{N}_i^{\mathcal{M}} \subseteq \omega$ is the extension of $\mathbf{N}_i$, for $1 \leq i \leq n$. We

[11] We add the axiom schemata $\mathsf{K}_{\mathbf{N}_i}$, and close MSPL under the $\mathsf{NEC}_{\mathbf{N}_i}$ rules, because as we shall see, these come out as being valid in the possible-world P-semantics we will construct.

restrict ourselves to models expanding $\mathbb{N}$ to give the arithmetic vocabulary its intended interpretation, to avoid questions regarding (the modal treatment of) non-standard syntax, and to keep our account simple. Call such models 'ordinary'. In particular, for any ordinary model $\mathcal{M}$, any $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$, and any $1 \leq i \leq n$ we have that

$$\mathcal{M} \vDash \mathbf{N}_i\overline{\ulcorner\phi\urcorner} \text{ iff } \ulcorner\phi\urcorner \in \mathbf{N}_i^{\mathcal{M}} \tag{4.57}$$

Furthermore, because the domain of any ordinary model $\mathcal{M}$ is ω, and every $n \in \omega$ can be named by the closed term $\overline{n}$, we have for any $\phi[x] \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$ and any variable assignments g, g' that

$$\mathrm{V}_{\langle\mathcal{M},g\rangle}(\phi[x]) = \mathrm{V}_{\langle\mathcal{M},g'\rangle}(\phi[\overline{g(x)}]) \tag{4.58}$$

By applying (4.58) k many times, for any k, we get that for any $\phi[x_1,\ldots,x_k] \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$ and any variable assignments g, g' that

$$\mathrm{V}_{\langle\mathcal{M},g\rangle}(\phi[x_1,\ldots,x_k]) = \mathrm{V}_{\langle\mathcal{M},g'\rangle}(\phi[\overline{g(x_1)},\ldots,\overline{g(x_k)}]) \tag{4.59}$$

Observe that $\phi[\overline{g(x_1)},\ldots,\overline{g(x_k)}]$ is a sentence, not a formula; and straightforwardly by induction it can be shown that the interpretation of sentences in a model is invariant to the variable assignment. Thus, for ordinary models, (4.59) shows us that we can drop variable assignments in our definitions of truth and validity. That is, $\phi[x_1,\ldots,x_k]$ is true in an ordinary model $\mathcal{M}$, denoted $\mathcal{M} \vDash \phi[x_1,\ldots,x_k]$, iff $\mathcal{M} \vDash \phi[\overline{n_1},\ldots,\overline{n_k}]$ for all $n_1,\ldots,n_k \in \omega$. We say that $\phi[x_1,\ldots,x_k]$ is valid in the class of ordinary models, denoted $\vDash_{\mathrm{Ord}} \phi[x_1,\ldots,x_k]$, iff $\mathcal{M} \vDash \phi[x_1,\ldots,x_k]$ for every ordinary model $\mathcal{M}$. In particular, we have that $\mathcal{M} \vDash \forall x_1 \ldots \forall x_k \phi[x_1,\ldots,x_k]$ iff $\mathcal{M} \vDash \phi[\overline{n_1},\ldots,\overline{n_k}]$ for all $n_1,\ldots,n_k \in \omega$.

Before moving on with the possible-world P-semantics, recall definition 3.1.7, which outlines the possible-world O-semantics for a single modal sentential operator. This semantics can be straightforwardly generalised for multiple distinct modal sentential operators. To briefly sketch this generalisation, the logical constant $\Box$ of $L_{\Box}$ is replaced with $\Box_i$, for $1 \leq i \leq n$. Call this $L_{\Box_1,\ldots,\Box_n}$. Frames are then expanded to $F_{\Box_1,\ldots,\Box_n} := \langle W, R_1,\ldots,R_n\rangle$, where each R_i is a binary relation on W. Lastly, the clause for $\Box\phi$ in the inductive definition of the valuation function is replaced with the clauses $V_I(w,\Box_i\phi) = 1$

iff $[I(v, \phi) = 1$ for all $v \in R_i(w)]$. The remaining notions are defined in a similar fashion as in the unimodal case. From hereon, possible-world O-semantics will refer to this generalised semantics for multiple distinct modal sentential operators.

Just like with the possible-world O-semantics, to construct the possible-world P-semantics we introduce the notion of a multimodal frame, which is an $n+1$-tuple $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$; where $\mathfrak{W}$ is a non-empty world-set, and each $\mathfrak{R}_{\mathbf{N}_i}$ is a binary relation on $\mathfrak{W}$, for $1 \leq i \leq n$. As before, we write $w\mathfrak{R}_{\mathbf{N}_i}v$ for $\langle w, v \rangle \in \mathfrak{R}_{\mathbf{N}_i}$, and define $\mathfrak{R}_{\mathbf{N}_i}(w) \coloneqq \{v \in \mathfrak{W} \mid \langle w, v \rangle \in \mathfrak{R}_{\mathbf{N}_i}\}$. We want that a possible-world P-model assigns an ordinary model to each $w \in \mathfrak{W}$. Moreover, we want all $\mathbf{N}_i$ to behave such that, for any $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$, we have that $\mathbf{N}_i\overline{\ulcorner \phi \urcorner}$ holds at a world w iff ϕ holds at all $v \in \mathfrak{R}_{\mathbf{N}_i}(w)$. However, unlike with the possible-world O-semantics, we cannot externally fix the interpretation of each $\mathbf{N}_i$, for $1 \leq i \leq n$, i.e. inductively define the interpretation of each $\mathbf{N}_i$. This is because, under standard complexity, sentential operators syntactically function like connectives in the inductive definition of well-formed formulae, outputting a more complex formula than the formula inputted. In contrast, first-order syntactic predicates are atomic, no matter the complexity of the formulae whose names they predicate over; for example, $\mathbf{N}_i\overline{\ulcorner p \wedge p \urcorner}$ is an atomic formula, whereas $p \wedge p$ is not atomic and more complex. As will become clear in the rest of this chapter, certain merging paradoxes precisely ensure that for certain multimodal frames we can never define suitable interpretations for certain $\mathbf{N}_i$. Thus, this more generally proves that there is no possible complexity measure of formulae upon which we could induct to define a suitable interpretation for any $\mathbf{N}_i$.

Nonetheless, we can still *internally* fix the interpretation of each $\mathbf{N}_i$, for $1 \leq i \leq n$, to behave as required as depending on accessible worlds, i.e. that $\mathbf{N}_i\overline{\ulcorner \phi \urcorner}$ holds at a world w iff ϕ holds at all $v \in \mathfrak{R}_{\mathbf{N}_i}(w)$. We ensure this by introducing a suitable condition that the interpretation of each $\mathbf{N}_i$ must satisfy in the definition of a possible-world P-model, or pw-model for short.

Definition 4.2.3 (Pw-model)**.** A pw-model is a tuple

$$\mathfrak{M} = \langle \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle, \mathbb{N}, \mathfrak{V}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n} \rangle$$

where:

- $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ is a multimodal frame;
- $\mathfrak{V}$ is the interpretation of p and a function from $\mathfrak{W}$ to $\{0, 1\}$;
- $\mathfrak{B}_{\mathbf{N}_i}$ is a function from $\mathfrak{W}$ to $\mathcal{P}(\ulcorner \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \urcorner) = \mathcal{P}(\omega)$, and we call it an $\mathbf{N}_i$-interpretation, for $1 \leq i \leq n$.
- We denote $\mathfrak{M}(w) := \langle \mathbb{N}, \mathfrak{V}(w), \mathfrak{B}_{\mathbf{N}_1}(w), \ldots, \mathfrak{B}_{\mathbf{N}_n}(w) \rangle$, for each $w \in \mathfrak{W}$. Then, for all $w \in \mathfrak{W}$ we require that all $\mathfrak{B}_{\mathbf{N}_i}$ satisfy the condition

$$\mathfrak{B}_{\mathbf{N}_i}(w) = \ulcorner \{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \mid \text{for each } v \in \mathfrak{R}_{\mathbf{N}_i}(w),\ \mathfrak{M}(v) \vDash \phi\} \urcorner$$

(Intended $\mathbf{N}_i$-Interpretation)

The condition Intended $\mathbf{N}_i$-Interpretation ensures that for each $\mathbf{N}_i$ we have that

$$\mathfrak{M}(w) \vDash \mathbf{N}_i \overline{\ulcorner \phi \urcorner} \text{ iff } \mathfrak{M}(v) \vDash \phi \text{ for each } v \in \mathfrak{R}_{\mathbf{N}_i}(w) \qquad (\mathbf{N}_i \text{ Truth Condition})$$

This is precisely what we wanted. We have internally fixed the interpretations of unary modal syntactic predicates to behave, paradox notwithstanding, just like modal sentential operators behave when their interpretations are externally fixed in the possible-world O-semantics; as depending on accessible worlds. However, as a consequence of needing to internally fix these interpretations, our definition of a pw-model does not yet ensure the existence of pw-models.

Observation 4.2.1. *There are at least two pw-models.*

Proof. Consider the multimodal frame DEF $:= \langle \{w\}, \emptyset, \ldots, \emptyset \rangle$. Given either $\mathfrak{V} = \{\langle w, 0 \rangle\}$ or $\mathfrak{V} = \{\langle w, 1 \rangle\}$ (which is given any $\mathfrak{V}$ defined on DEF), we have that

$$\langle \text{DEF}, \mathbb{N}, \mathfrak{V}, \{\langle w, \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \rangle\}, \ldots, \{\langle w, \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \rangle\} \rangle$$

is a pw-model. Firstly, one can check it is well defined. Secondly, the conditions Intended $\mathbf{N}_i$-Interpretation, for $1 \leq i \leq n$, are satisfied; 'for each $v \in \mathfrak{R}_{\mathbf{N}_i}(w)$, $\mathfrak{M}(v) \vDash \phi$' is vacuously true, and hence $\mathfrak{M}(w) \vDash \mathbf{N}_i \overline{\ulcorner \phi \urcorner}$ for all $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$. Thus, DEF admits a pw-model on every interpretation. Q.E.D.

Since we will be analysing (classes of) multimodal frames, and hence fixing them; and as hinted at, certain multimodal frames never admit suitable interpretations for certain $\mathbf{N}_i$; we introduce the following useful definition.

Definition 4.2.4. A multimodal frame $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ admits a pw-model iff for some interpretation $\mathfrak{V}'$ there are $\mathbf{N}_i$-interpretations $\mathfrak{B}'_{\mathbf{N}_i}$ such that

$$\langle \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle, \mathbb{N}, \mathfrak{V}', \mathfrak{B}'_{\mathbf{N}_1}, \ldots, \mathfrak{B}'_{\mathbf{N}_n} \rangle$$

is a pw-model. It admits a pw-model *on every interpretation* iff for every interpretation $\mathfrak{V}$ there are $\mathbf{N}_i$-interpretations $\mathfrak{B}_{\mathbf{N}_i}$ such that

$$\langle \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle, \mathbb{N}, \mathfrak{V}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n} \rangle$$

is a pw-model. We say that a multimodal frame $\mathfrak{F}$ is raised to a pw-model iff a pw-model $\mathfrak{M}$ is provided whose frame, i.e. first component, is $\mathfrak{F}$.

Furthermore, the semantics is defined similarly to the possible-world O-semantics.

Definition 4.2.5 (Truth and Validity)**.** Let $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$. Then:

i) ϕ is true at a world w in a pw-model $\mathfrak{M}$ iff $\mathfrak{M}(w) \vDash \phi$.

ii) ϕ is true in a pw-model $\mathfrak{M}$ iff $\mathfrak{M}(v) \vDash \phi$ for all v in $\mathfrak{M}$. We denote this by $\mathfrak{M} \vDash \phi$.

iii) ϕ is true in a multimodal frame $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ iff

   (a) $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ admits a pw-model on every interpretation.

   (b) For every $\mathfrak{V}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n}$, if $\mathfrak{M} = \langle \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle, \mathbb{N}, \mathfrak{V}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n} \rangle$ is a pw-model, then $\mathfrak{M} \vDash \phi$.

   We denote this by $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle \vDash \phi$.

iv) ϕ is valid in a class of multimodal frames $\mathfrak{F}_{\mathbf{N}_i}$ iff for every $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle \in \mathfrak{F}_{\mathbf{N}_i}$ we have that $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle \vDash \phi$.

v) ϕ is $\mathfrak{K}$-valid iff for every multimodal frame $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ we have that

   (a) either $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ does not admit a pw-model on every interpretation;

(b) or $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle \vDash \phi$.

We denote this by $\vDash_{\mathfrak{R}} \phi$.

We restrict to frames that admit a pw-model on every interpretation in definitions iii)-v) because we want to freely interpret the contingent vocabulary p; and it is an open question whether the class of frames that admit a pw-model on some interpretation is the same as the class of frames that admit a pw-model on every interpretation.

The possible-world P-semantics behaves in some respects like the possible-world O-semantics.

Lemma 4.2.1 ('Normality')**.** *Each* $\mathbf{N}_i$*, for* $1 \leq i \leq n$*, is 'normal', in a sense similar to the possible-world O-semantics. That is, we have that*

$$\mathfrak{M} \vDash \mathbf{N}_i\overline{\ulcorner \phi \rightarrow \psi \urcorner} \rightarrow (\mathbf{N}_i\overline{\ulcorner \phi \urcorner} \rightarrow \mathbf{N}_i\overline{\ulcorner \psi \urcorner}) \tag{$\mathsf{K}_{\mathbf{N}_i}$}$$

for any pw-model $\mathfrak{M}$*, any* $\phi, \psi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$*, and any* $1 \leq i \leq n$*. Moreover, the following rules hold*

$$\mathfrak{M} \vDash \phi \Rightarrow \mathfrak{M} \vDash \mathbf{N}_i\overline{\ulcorner \phi \urcorner} \tag{$\mathsf{NEC}_{\mathbf{N}_i}$}$$

for any pw-model $\mathfrak{M}$*, any* $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$*, and any* $1 \leq i \leq n$*.*

Proof. Let $\mathfrak{M} = \langle \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle, \mathbb{N}, \mathfrak{V}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n} \rangle$ be a pw-model, $w \in \mathfrak{W}$, and $\phi, \psi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$. Then:

- For each $\mathsf{K}_{\mathbf{N}_i}$, assume that $\mathfrak{M}(w) \vDash \mathbf{N}_i\overline{\ulcorner \phi \rightarrow \psi \urcorner}$ and $\mathfrak{M}(w) \vDash \mathbf{N}_i\overline{\ulcorner \phi \urcorner}$. If $\mathfrak{R}_{\mathbf{N}_i}(w) = \emptyset$, then vacuously $\mathfrak{M}(w) \vDash \mathbf{N}_i\overline{\ulcorner \psi \urcorner}$. If $\mathfrak{R}_{\mathbf{N}_i}(w) \neq \emptyset$, let $v \in \mathfrak{R}_{\mathbf{N}_i}(w)$ be arbitrary. Then by $\mathbf{N}_i$ Truth Condition, $\mathfrak{M}(v) \vDash (\phi \rightarrow \psi) \wedge \phi$ and hence $\mathfrak{M}(v) \vDash \psi$. Thus $\mathfrak{M}(w) \vDash \mathbf{N}_i\overline{\ulcorner \psi \urcorner}$.
- For each $\mathsf{NEC}_{\mathbf{N}_i}$, let $\mathfrak{M}(v) \vDash \phi$ for every $v \in \mathfrak{W}$. Then also $\mathfrak{M}(v') \vDash \phi$ for every $v' \in \mathfrak{R}_{\mathbf{N}_i}(w)$. Hence by $\mathbf{N}_i$ Truth Condition $\mathfrak{M}(w) \vDash \mathbf{N}_i\overline{\ulcorner \phi \urcorner}$.

Q.E.D.

Corollary 4.2.1. *Each* $\mathsf{K}_{\mathbf{N}_i}$*-schema is* $\mathfrak{R}$*-valid, and the rules '*$\vDash_{\mathfrak{R}} \phi \Rightarrow \vDash_{\mathfrak{R}} \mathbf{N}_i\overline{\ulcorner \phi \urcorner}$*' hold.*

Lemma 4.2.1 allows us to prove that MSPL is sound with respect to the possible-world P-semantics. This will be crucial when we use diagonalisation results to prove that for certain multimodal frames we can never define suitable interpretations for certain $\mathbf{N}_i$.

Lemma 4.2.2. *Let $\mathfrak{M}$ be a pw-model, $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$, and $m \in \omega$. Then we have that*

$$(\mathsf{MSPL})_m^- \vdash \phi \Rightarrow \mathfrak{M} \vDash \phi \tag{4.60}$$

Proof. We will prove this by induction on m. For the base case, consider $(\mathsf{MSPL})_0^-$. Since any ordinary model expands the standard model $\mathbb{N}$ of PA, we have that the axioms of PA formulated in $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$ are true at any world, in any pw-model. By Lemma 4.2.1, the $\mathsf{K}_{\mathbf{N}_i}$-schema, for $1 \leq i \leq n$, are also true at any world, in any pw-model. Thus, $(\mathsf{MSPL})_0^-$ is the theory of a set of formulae that are true at any world, in any pw-model. Now, recall that no proof theory is specified; instead, it is left up to the reader to use whichever logical calculus they are comfortable with, so long as it is sound with respect to the model-theoretic semantics outlined in section 2.1. Given that each world is assigned a model that is defined using this model-theoretic semantics (namely an ordinary model), by a standard inductive soundness argument we have that $(\mathsf{MSPL})_0^-$ itself is true at any world, in any pw-model. That is, for any pw-model $\mathfrak{M}$ and $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$, if $(\mathsf{MSPL})_0^- \vdash \phi$, then $\mathfrak{M} \vDash \phi$.

For the inductive step, assume that for any pw-model $\mathfrak{M}$ and $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$, if $(\mathsf{MSPL})_m^- \vdash \phi$, then $\mathfrak{M} \vDash \phi$. Now consider $(\mathsf{MSPL})_{m+1}^- \coloneqq \mathsf{Th}[(\mathsf{MSPL})_m^- \cup \{\mathbf{N}_i\overline{\ulcorner\phi\urcorner} \mid (\mathsf{MSPL})_m^- \vdash \phi,\ 1 \leq i \leq n\}]$. Since $(\mathsf{MSPL})_m^- \vdash (\mathsf{MSPL})_m^-$, by the inductive hypothesis we have that $(\mathsf{MSPL})_m^-$ is true in any pw-model. Moreover, by Lemma 4.2.1 we have that, if $\mathfrak{M} \vDash \phi$, then $\mathfrak{M} \vDash \mathbf{N}_i\overline{\ulcorner\phi\urcorner}$, for any pw-model $\mathfrak{M}$. Thus, also by the inductive hypothesis we have that $\{\mathbf{N}_i\overline{\ulcorner\phi\urcorner} \mid (\mathsf{MSPL})_m^- \vdash \phi,\ 1 \leq i \leq n\}$ is true in any pw-model. Therefore, $(\mathsf{MSPL})_{m+1}^-$ is the theory of a set of formulae that are true at any world, in any pw-model. By a standard inductive soundness argument we have that $(\mathsf{MSPL})_{m+1}^-$ itself is true at any world, in any pw-model. That is, for any pw-model $\mathfrak{M}$ and $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$, if $(\mathsf{MSPL})_{m+1}^- \vdash \phi$, then $\mathfrak{M} \vDash \phi$. Q.E.D.

Theorem 4.2.1 (Soundness)**.** *Let $\mathfrak{M}$ be a pw-model, and $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$. Then we have that*

$$\mathsf{MSPL} \vdash \phi \Rightarrow \mathfrak{M} \vDash \phi \tag{4.61}$$

Proof. Since proofs are finite; $(\mathsf{MSPL})^-_m \subseteq (\mathsf{MSPL})^-_{m+1}$ for each $m \in \omega$; and $\mathsf{MSPL} := \bigcup_{k\in\mathbb{N}}(\mathsf{MSPL})^-_k$; for some l we have that $(\mathsf{MSPL})^-_l \vdash \phi$. By Lemma 4.2.2, we have that $\mathfrak{M} \vDash \phi$. Q.E.D.

Corollary 4.2.2. *Let $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$ and $m \in \omega$. Then we have that*

$$(\mathsf{MSPL})^-_m \vdash \phi \Rightarrow \vDash_{\mathfrak{K}} \phi \tag{4.62}$$

and

$$\mathsf{MSPL} \vdash \phi \Rightarrow \vDash_{\mathfrak{K}} \phi \tag{4.63}$$

Corollary 4.2.3. *Let $\mathfrak{M}$ be a pw-model, $S \subseteq \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$ be such that $\mathfrak{M} \vDash S$, E be the theory extending MSPL with S, and $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$. Then we have that*

$$\mathsf{E} \vdash \phi \Rightarrow \mathfrak{M} \vDash \phi \tag{4.64}$$

In particular, if $\mathsf{MSPL} \subsetneq \mathsf{E}'$ is an inconsistent theory, then there is no pw-model that makes E' true.

Proof. The proof is a straightforward generalisation of the proof of Theorem 4.2.2. Q.E.D.

Let us now briefly introduce some terminology on binary relations that will be useful to distinguish different classes of multimodal frames.

Definition 4.2.6. Let $R \subseteq W \times W$ be a binary relation. We say that

- R is serial iff

$$\forall w \in W\ \exists v \in W\ \langle w, v\rangle \in R \tag{4.65}$$

- R is reflexive iff

$$\forall w \in W\ \langle w, w\rangle \in R \tag{4.66}$$

- R is symmetric iff

$$\forall w, v \in W\ (\langle w, v\rangle \in R \rightarrow \langle v, w\rangle \in R) \tag{4.67}$$

- R is transitive iff

$$\forall w, v, u \in W\ ((\langle w, v\rangle \in R \wedge \langle v, u\rangle \in R) \rightarrow \langle w, u\rangle \in R) \tag{4.68}$$

- R is an equivalence relation iff R is reflexive, symmetric, and transitive.
- R is converse well-founded iff

$$\forall S \subseteq W(S \neq \emptyset \rightarrow \exists w \in S\ \forall v \in S\ \langle w, v\rangle \notin R) \tag{4.69}$$

- R is converse ill-founded iff it is not converse well-founded.
- R is well-founded (ill-founded) iff R^{-1} is converse well-founded (converse ill-founded).

Definition 4.2.7. For any property X of binary relations, we say that a multimodal frame $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle$ is $\mathfrak{R}_{\mathbf{N}_i}$-X iff $\mathfrak{R}_{\mathbf{N}_i}$ is X, for $1 \leq i \leq n$. For example, $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle$ is $\mathfrak{R}_{\mathbf{N}_1}$-serial iff $\mathfrak{R}_{\mathbf{N}_1}$ is serial; or $\mathfrak{R}_{\mathbf{N}_n}$-reflexive iff $\mathfrak{R}_{\mathbf{N}_n}$ is reflexive; etc.

Just like with the possible-world O-semantics, we have the following result.

Lemma 4.2.3 (Transitivity)**.** *For any pw-model* $\mathfrak{M}$ *where* $\mathfrak{R}_{\mathbf{N}_i}$*-transitive for some* $1 \leq i \leq n$*, we have that*

$$\mathfrak{M} \vDash \mathbf{N}_i\overline{\ulcorner\phi\urcorner} \rightarrow \mathbf{N}_i\overline{\ulcorner\mathbf{N}_i\overline{\ulcorner\phi\urcorner}\urcorner} \tag{$4_{\mathbf{N}_i}$}$$

for any $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$.

Proof. Assume for contradiction that there is a model $\mathfrak{M}' = \langle\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle, \mathbb{N}, \mathfrak{V}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n}\rangle$ and $w \in \mathfrak{W}$ such that

(i) $\mathfrak{M}'(w) \vDash \mathbf{N}_i\overline{\ulcorner\phi\urcorner}$

(ii) $\mathfrak{M}'(w) \nvDash \mathbf{N}_i\overline{\ulcorner\mathbf{N}_i\overline{\ulcorner\phi\urcorner}\urcorner}$

By (ii) and $\mathbf{N}_i$ Truth Condition, there is some $v \in \mathfrak{R}_{\mathbf{N}_i}(w)$ such that $\mathfrak{M}'(v) \not\models \mathbf{N}_i\overline{\ulcorner\phi\urcorner}$. Again by $\mathbf{N}_i$ Truth Condition, there is some $u \in \mathfrak{R}_{\mathbf{N}_i}(v)$ such that $\mathfrak{M}'(u) \not\models \phi$. By $\mathfrak{R}_{\mathbf{N}_i}$-transitivity, $w\mathfrak{R}_{\mathbf{N}_i}u$. Hence, by (i) and $\mathbf{N}_i$ Truth Condition, $\mathfrak{M}'(u) \models \phi$ ※. Q.E.D.

However, in other important respects, the possible-world P-semantics we have constructed behaves very different to the possible-world O-semantics. For instance, as before, Lemma 4.2.3 does not yet ensure the existence of $\mathfrak{R}_{\mathbf{N}_i}$-transitive pw-models. Luckily, Lemma 4.2.3 does not hold vacuously; as we shall see, there are $\mathfrak{R}_{\mathbf{N}_i}$-transitive frames that admit a pw-model on every interpretation. Furthermore, we cannot however strengthen Lemma 4.2.3 to an 'if and only if' statement. If we are given a multimodal frame that is not $\mathfrak{R}_{\mathbf{N}_i}$-transitive, we might not be guaranteed that the counter-interpretation we construct for it can be extended to a pw-model. For instance, below we will consider multimodal frames that are not $\mathfrak{R}_{\mathbf{N}_i}$-transitive and do not admit pw-models on *any* interpretation. If we add to Lemma 4.2.3 the constraint that the frame admits a pw-model on every interpretation, then in the usual way we can construct a counter-interpretation to prove an 'if and only if': find the $\mathsf{w}, \mathsf{v}, \mathsf{u}$ such that $\mathsf{w}\mathfrak{R}_{\mathbf{N}_i}\mathsf{v}$ and $\mathsf{v}\mathfrak{R}_{\mathbf{N}_i}\mathsf{u}$, but not $\mathsf{w}\mathfrak{R}_{\mathbf{N}_i}\mathsf{u}$. Then interpret p such that $\mathfrak{V}(w) = 1$ for $w \in \mathfrak{R}_{\mathbf{N}_i}(\mathsf{w})$, and $\mathfrak{V}(w) = 0$ otherwise. Extend this to a model $\mathfrak{M}$. Then $\mathfrak{M}(\mathsf{w}) \not\models \mathbf{N}_i\overline{\ulcorner p\urcorner} \rightarrow \mathbf{N}_i\overline{\ulcorner\mathbf{N}_i\overline{\ulcorner p\urcorner}\urcorner}$.

More generally, as alluded to, certain merging paradoxes ensure that for certain multimodal frames we can never define suitable interpretations for certain $\mathbf{N}_i$, i.e. they do not admit pw-models on any interpretation. For example, the following result is the first instance of such a 'model-theoretic paradox' we will consider in this chapter. This result goes in exactly the same way as in the unimodal setting; c.f. Halbach and Leigh 2024, Example 7.7.

Lemma 4.2.4 (No reflexivity)**.** *Let* $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle$ *be a multimodal frame that is* $\mathfrak{R}_{\mathbf{N}_i}$*-reflexive for some* $1 \leq i \leq n$*. Then* $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle$ *does not admit a pw-model. As a consequence, no frame that is an* $\mathfrak{R}_{\mathbf{N}_i}$*-equivalence for any* $1 \leq i \leq n$ *admits a pw-model.*

Proof. The proof is a model-theoretic version of the proof of Montague's Paradox. Assume for contradiction that there are $\mathfrak{V}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n}$ such that

$$\mathfrak{M} = \langle\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle, \mathbb{N}, \mathfrak{V}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n}\rangle$$

is a pw-model. By the Diagonal Lemma, we can find a λ such that

$$\mathsf{MSPL} \vdash \lambda \leftrightarrow \neg \mathbf{N}_i \overline{\ulcorner \lambda \urcorner}$$

Since $\mathfrak{W} \neq \emptyset$, pick an arbitrary $w \in \mathfrak{W}$. By soundness,

$$\mathfrak{M}(w) \vDash \lambda \leftrightarrow \neg \mathbf{N}_i \overline{\ulcorner \lambda \urcorner}$$

If $\mathfrak{M}(w) \vDash \neg\lambda$, then $\mathfrak{M}(w) \vDash \mathbf{N}_i \overline{\ulcorner \lambda \urcorner}$. By $\mathfrak{R}_{\mathbf{N}_i}$-reflexivity we have that $w \in \mathfrak{R}_{\mathbf{N}_i}(w)$. By $\mathbf{N}_i$ Truth Condition we have that $\mathfrak{M}(w) \vDash \lambda$ ※.

Thus, $\mathfrak{M}(w) \vDash \lambda$ for all $w \in \mathfrak{W}$. Pick a $\mathsf{w} \in \mathfrak{W}$. Then, $\mathfrak{M}(\mathsf{w}) \vDash \lambda$. By the $\mathsf{NEC}_{\mathbf{N}_i}$ rule we have that $\mathfrak{M}(\mathsf{w}) \vDash \mathbf{N}_i \overline{\ulcorner \lambda \urcorner}$. Hence, using the biconditional, we have that $\mathfrak{M}(\mathsf{w}) \vDash \neg\lambda$ ※. Q.E.D.

In a very similar fashion, we have the following result.

Lemma 4.2.5 (No $\mathsf{T}_{\mathbf{N}_i}$)**.** *No schema $\mathsf{T}_{\mathbf{N}_i}$, for $1 \leq i \leq n$, is true in any pw-model. In particular, no property of $\mathfrak{R}_{\mathbf{N}_i}$ ensures $\mathsf{T}_{\mathbf{N}_i}$ is true in a pw-model; and no (class of) frame(s) makes any $\mathsf{T}_{\mathbf{N}_i}$ valid.*

Proof. The proof is the same as the proof of Lemma 4.2.4, except for that we replace reference to $\mathfrak{R}_{\mathbf{N}_i}$-reflexivity with the fact that $\mathsf{T}_{\mathbf{N}_i}$ is true in the pw-model we are considering for contradiction. This Lemma can also be proved from Montague's Paradox using Corollary 4.2.3. Q.E.D.

Furthermore, we will consider other model-theoretic paradoxes later in this chapter that similarly show that no $\mathfrak{R}_{\mathbf{N}_i}$-serial frame admits pw-models, and no $\mathfrak{R}_{\mathbf{N}_i}$-symmetric frame admits pw-models. Recall from chapter 3 what happens when diagonalisation is blocked in Stern's framework in (2014); (2015), in particular Theorem 3.1.3. This elucidates that

the difference in behaviour between the possible-world P-semantics and the possible-world O-semantics arises from the mention predicate approach having more expressive power than the operator approach; the Diagonal Lemma is what leads to the inconsistency in Lemma 4.2.4.

These observations illuminate the substantial differences between the possible-world P-semantics constructed here and the standard possible-world O-semantics, even though they are defined in seemingly analogous ways. In particular, where for the standard possible-world O-semantics every frame admits a model on every interpretation; for the possible-world P-semantics there are frames that admit no pw-models. Thus, three natural questions spring out:

1. Which multimodal frames admit no pw-models?
2. Which multimodal frames admit a pw-model on every interpretation?
3. Which multimodal frames admit a pw-model?
    3.1 Are there multimodal frames that admit a pw-model, but do not admit a pw-model on every interpretation?

Note that question 3 is answered if necessary and sufficient conditions are provided for question 1. Furthermore, if question 3.1 is answered in the affirmative, then question 3 is more general than question 2. Finally, note that without the contingent vocabulary p, question 3.1 disappears and questions 2 and 3 collapse to the same question of which multimodal frames admit a pw-model.

Question 2 is answered in the unimodal setting by Halbach and Leigh 2024, §7:

Theorem 4.2.2 (Strong Characterisation Theorem)**.** *A unimodal frame admits a pw-model on every interpretation iff the frame is converse well-founded.*

Questions 3 and 1 have been partially answered in the unimodal setting without contingent vocabulary by Halbach, Leitgeb, et al. 2003, and ends up being a technically difficult endeavour. Consider the unimodal language $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}}$; a unimodal possible-world P-semantics; a unimodal frame $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}} \rangle$; and the natural rank function associated with a well-order.

For any $w \in \mathfrak{W}$ we then say that w is converse well-founded iff $\mathfrak{R}_{\mathbf{N}}\big|_{\mathfrak{R}_{\mathbf{N}}(w)\times\mathfrak{R}_{\mathbf{N}}(w)}$ is converse well-founded; and converse ill-founded iff $\mathfrak{R}_{\mathbf{N}}\big|_{\mathfrak{R}_{\mathbf{N}}(w)\times\mathfrak{R}_{\mathbf{N}}(w)}$ is converse ill-founded. We also define the converse well-founded part of $\mathfrak{R}_{\mathbf{N}}(w)$ as the largest $\mathfrak{R}_{\mathbf{N}}(w) \subseteq \mathsf{CWFP}[\mathfrak{R}_{\mathbf{N}}(w)] \subseteq \mathfrak{W}$ upwards-closed under $\mathfrak{R}_{\mathbf{N}}$, i.e. such that $\forall x \in \mathsf{CWFP}[\mathfrak{R}_{\mathbf{N}}(w)]\forall y(y \in \mathfrak{R}_{\mathbf{N}}(x) \rightarrow y \in \mathsf{CWFP}[\mathfrak{R}_{\mathbf{N}}(w)])$, with $\mathfrak{R}_{\mathbf{N}}\big|_{\mathsf{CWFP}[\mathfrak{R}_{\mathbf{N}}(w)]\times\mathsf{CWFP}[\mathfrak{R}_{\mathbf{N}}(w)]}$ converse well-founded. We then define the depth of a world $w \in \mathfrak{W}$ as the rank of $(\mathfrak{R}_{\mathbf{N}}\big|_{\mathsf{CWFP}[\mathfrak{R}_{\mathbf{N}}(w)]\times\mathsf{CWFP}[\mathfrak{R}_{\mathbf{N}}(w)]})^{-1}$. Importantly, a world can be converse ill-founded and have depth. Furthermore, let L denote Gödel's constructible hierarchy (see Barwise 2017); let ω_1 denote the first non-recursive ordinal (see Rogers 1967); let κ be the least ordinal such that L_κ has a transitive Σ_1-end extension (see Barwise 2017); let ADM be the class of what are called admissible ordinals α such that $(L_\alpha, \in)$ is a model of Kripke–Platek Set Theory (see Barwise 2017); and let ADM^* be ADM with all of its limit points. Halbach, Leitgeb, et al. 2003 then provide sufficient conditions for a transitive unimodal frame $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}} \rangle$ to admit a pw-model:

- If every $w \in \mathfrak{W}$ has depth $\alpha_w < \omega_1$ then [$\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}} \rangle$ admits a pw-model iff $\mathfrak{R}_{\mathbf{N}}$ is converse well-founded].

- If every converse ill-founded world w has depth $\alpha_w \geq \kappa$ (with no restriction on $\mathfrak{R}_{\mathbf{N}}$ for the part containing those converse ill-founded worlds), then $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}} \rangle$ admits a pw-model.

Halbach, Leitgeb, et al. 2003 also provide a necessary condition for a unimodal frame $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}} \rangle$ to admit a pw-model:

- If $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}} \rangle$ admits a pw-model, then any converse ill-founded world w in the transitive closure $\langle \mathfrak{W}, \mathfrak{R}^*_{\mathbf{N}} \rangle$ of $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}} \rangle$ has depth α_w such that either $\alpha_w \in ADM^*$ or $\alpha_w \geq \kappa$.

Furthermore, Halbach, Leitgeb, et al. 2003, §12 discuss generalising their approach to include contingent vocabulary, and they claim that their results still hold. Given this, we see that Question 3.1 is answered in the affirmative; by Halbach, Leitgeb, et al. 2003, any transitive unimodal frame that has converse ill-founded worlds w, all of whose depth is

such that $\alpha_w \geq \kappa$, admits a pw-model; however, such a frame is converse ill-founded, and thus by Theorem 4.2.2 does not admit a pw-model on every interpretation.

In the rest of this chapter we will successively answer questions 1 and 2. In particular, question 1 will be partially answered by considering examples of model-theoretic merging paradoxes. Question 2 will be answered by generalising Theorem 4.2.2 to the multimodal setting. Our primary focus will be on question 2, because it is concerned with how the structure of a multimodal frame in question avoids paradox and contradiction. Consider for example the propositional modal operator logic Gödel-Löb, denoted *GL*. It is in the language $L_{\Box}$, and its axiom schemas are the propositional tautologies, $\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \psi)$, $\Box(\Box\phi \rightarrow \phi) \rightarrow \Box\phi$; and it is closed under modus ponens and necessitation for $\Box$. *GL* is consistent, and we can construct a possible-world O-model for it, known as the canonical model $M_{GL} \coloneqq \langle W_{GL}, R_{GL}, I_{GL}\rangle$ for GL. Let W_{GL} be the set of all maximally *GL*-consistent sets. For every $w, v \in W_{GL}$, let $wR_{GL}v$ hold iff for every sentence ϕ, if $\Box\phi \in w$, then $\phi \in v$. Finally, for every $w \in W_{GL}$ and every sentence letter $p \in SL$, let $I_{GL}(w, p) = 1$ iff $p \in w$. It can then be showed that $GL \vdash \phi$ iff $M_{GL} \vDash \phi$. For a proof see for example Boolos 1993, §6. However the canonical model for *GL* is not very interesting since it does not help us in providing a completeness result for *GL*. In particular, Boolos 1993, p. 90-91 remarks that the canonical model for GL is reflexive, and thus converse ill-founded. However, $\Box(\Box\phi \rightarrow \phi) \rightarrow \Box\phi$ looks like an induction schema, which is well-founded. And in fact, as Boolos 1993, Theorem 10, §4 proves, *GL* is sound and complete with respect to the class of transitive and converse well-founded frames.

In this thesis, we are not interested in which frames admit a cleverly constructed interpretation of the contingent vocabulary that makes the schema $\Box(\Box\phi \rightarrow \phi) \rightarrow \Box\phi$ true, like for M_{GL}. Rather, we are interested in the stricter constraint of what aspects of the structure of a frame ensure that the schema $\Box(\Box\phi \rightarrow \phi) \rightarrow \Box\phi$ is valid, no matter how we interpret the contingent vocabulary. Similarly, in the possible-world P-semantics we are interested in the stricter constraint of what aspects of the structure of a multimodal frame ensure that we always avoid paradox and contradiction, no matter how we interpret the contingent vocabulary; if paradox arises under a certain

interpretation of the contingent vocabulary, then this is in tension with the fact that the contingent vocabulary should precisely be freely interpretable. This is why the multimodal generalisation of Theorem 4.2.2 that we will prove at the end of this chapter will function as the promised categorisation theorem indicating when paradox arises in this model theory.

4.3 Merging paradoxes in the model-theoretic setting

In this section, we will consider examples of merging paradoxes in the possible-world P-semantics constructed in the previous section 4.2, proving that certain classes of multimodal frames admit no pw-models. Call these non-existence results, which will provide us with partial answers to question 1. We will also show that when expanding the classical possible-world P-semantics to a classical two-dimensional possible-world P-semantics, no syntactic predicate can have the same truth-conditions as the sentential operators @, $\times$, or F have in the two-dimensional possible-world O-semantics. Let us first consider examples of merging paradoxes. The possible-world P-semantics allows us to intuitively and suitably represent (i) the referential structures of merging paradoxes extending MSPL; and (ii) the interactions between the distinct modalities in merging paradoxes extending MSPL. As we shall see, proof-theoretic results proving inconsistencies of different extensions of MSPL can be adapted into non-existence results. In particular, diagonalisation lies at the heart of the merging paradoxes; and thus at the heart of the non-existence results. We can then visualise a merging paradox extending MSPL by means of the class of graphs associated to a class of frames for which said merging paradox proves said class of frames admits no pw-models. The classes of graphs then do the work of being an intuitive and suitable representation of merging paradoxes extending MSPL. Merging paradoxes have so far received little attention in the literature, and this analysis will further our understanding of them.

Considering first unimodal frames, observe that any unimodal frame can be formulated in the current multimodal framework. In particular, many unimodal non-existence results carry over to the current multimodal framework. For an analysis of unimodal paradoxes

and unimodal non-existence results see Halbach, Leitgeb, et al. 2003, §3 and Halbach and Leigh 2024, §7.2. In the current multimodal framework, their results show that e.g. multimodal frames do not admit pw-models if for any $1 \leq i \leq n$ they; are $\mathfrak{R}_{\mathbf{N}_i}$-reflexive (see also Lemma 4.2.4); form n-step $\mathfrak{R}_{\mathbf{N}_i}$-loops, for $n \in \omega$; are such that their domain is ω and $\mathfrak{R}_{\mathbf{N}_i}$ is the 'strictly less than' relation $<$; or are such that their domain is ω and $\mathfrak{R}_{\mathbf{N}_i}$ is the successor function S.

Moving on to multimodal frames, the first merging paradox we will consider is the Multimodal Postcard Paradox. This is a multimodal generalisation of Halbach and Leigh 2024, Example 7.8, and a model-theoretic multimodal version of the postcard paradox. Consider the graph depicted in figure 4.1 below which denotes the multimodal frame with world set $\mathfrak{W} = \{w_1, w_2, \ldots, w_{n-1}, w_n\}$; and relations $\mathfrak{R}_{\mathbf{N}_i} = \{\langle w_{[i]_n}, w_{[i+1]_n}\rangle\}$, where $j \in [k]_n$ iff $j \equiv k \mod n$.

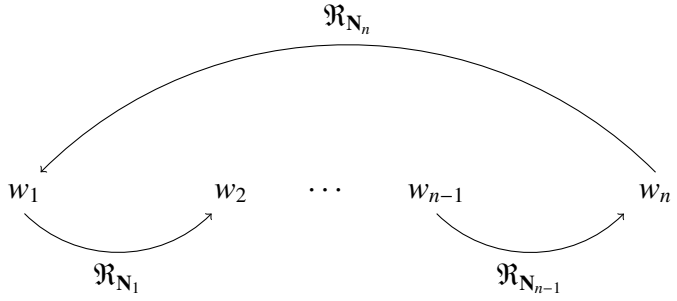

Figure 4.1: MPP

Theorem 4.3.1 (Multimodal Postcard Paradox)**.** *The multimodal frame* $\mathfrak{F}^1 := \langle \mathfrak{W}^1, \mathfrak{R}^1_{\mathbf{N}_1}, \ldots, \mathfrak{R}^1_{\mathbf{N}_n}\rangle$ *with world set* $\mathfrak{W}^1 = \{w_1, w_2, \ldots, w_{n-1}, w_n\}$*; and relations* $\mathfrak{R}^1_{\mathbf{N}_i} = \{\langle w_{[i]_n}, w_{[i+1]_n}\rangle\}$ *does not admit a pw-model. Similarly, any multimodal frame that is a relabelling of* $\mathfrak{F}^1$ *does not admit a pw-model.*

Proof. Assume for contradiction that $\mathfrak{M}^1$ does admit a pw-model, i.e. that there are some $\mathfrak{V}^1, \mathfrak{B}^1_{\mathbf{N}_1}, \ldots, \mathfrak{B}^1_{\mathbf{N}_n}$ such that

$$\mathfrak{M}^1 = \langle\langle \mathfrak{W}^1, \mathfrak{R}^1_{\mathbf{N}_1}, \ldots, \mathfrak{R}^1_{\mathbf{N}_n}\rangle, \mathbb{N}, \mathfrak{V}^1, \mathfrak{B}^1_{\mathbf{N}_1}, \ldots, \mathfrak{B}^1_{\mathbf{N}_n}\rangle$$

is a pw-model. By the Diagonal Lemma, we can find a λ such that

$$\mathsf{MSPL} \vdash \lambda \leftrightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner}^{\ldots}\urcorner}\urcorner} \tag{4.70}$$

By soundness we have that

$$\mathfrak{M}^1(w_1) \vDash \lambda \leftrightarrow \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner}^{\ldots}\urcorner}\urcorner}$$

Now either we have that $\mathfrak{M}^1(w_1) \vDash \lambda$, or $\mathfrak{M}^1(w_1) \vDash \neg\lambda$.

- If $\mathfrak{M}^1(w_1) \vDash \lambda$, then $\mathfrak{M}^1(w_1) \vDash \neg\mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner}^{\ldots}\urcorner}\urcorner}$. Thus, by applications of $\mathbf{N}_i$ Truth Condition and the fact that $\mathfrak{R}_{\mathbf{N}_i}(w_i) = \{w_{[i+1]_n}\}$, for $1 \leq i \leq n-1$, we have that $\mathfrak{M}^1(w_n) \vDash \neg\mathbf{N}_n\overline{\ulcorner\lambda\urcorner}$. One more application and the fact that $\mathfrak{R}_{\mathbf{N}_n}(w_n) = \{w_1\}$, implies that $\mathfrak{M}^1(w_1) \vDash \neg\lambda$ ※.
- If $\mathfrak{M}^1(w_1) \vDash \neg\lambda$, then $\mathfrak{M}^1(w_1) \vDash \mathbf{N}_1\overline{\ulcorner\mathbf{N}_2\overline{\ulcorner\ldots\mathbf{N}_n\overline{\ulcorner\lambda\urcorner}^{\ldots}\urcorner}\urcorner}$. Again, by applications of $\mathbf{N}_i$ Truth Condition and the fact that $\mathfrak{R}_{\mathbf{N}_i}(w_i) = \{w_{[i+1]_n}\}$, for $1 \leq i \leq n$, we have that $\mathfrak{M}^1(w_1) \vDash \lambda$ ※.

Both cases lead to contradiction. Hence, $\langle \mathfrak{W}^1, \mathfrak{R}^1_{\mathbf{N}_1}, \ldots, \mathfrak{R}^1_{\mathbf{N}_n}\rangle$ does not admit a pw-model.

Q.E.D.

We can generalise the Multimodal Postcard Paradox to the following merging paradox.

Theorem 4.3.2 (Model-Theoretic Multimodal Montague's Paradox)**.** *Let* $\langle \mathfrak{W}^2, \mathfrak{R}^2_{\mathbf{N}_1}, \ldots, \mathfrak{R}^2_{\mathbf{N}_n}\rangle$ *be a multimodal frame such that*

> *For some* $2 \leq j \leq n$*; for every* $w \in \mathfrak{W}^2$*, there is a tuple* $\langle w, v^w_2, \ldots, v^w_j\rangle$ *such that* $v^w_2 \in \mathfrak{R}_{\mathbf{N}_1}(w)$*;* $v^w_{i+1} \in \mathfrak{R}_{\mathbf{N}_i}(v^w_i)$*, for* $2 \leq i \leq j-1 \leq n-1$*; and* $w \in \mathfrak{R}_{\mathbf{N}_j}(v^w_j)$*.*

Then $\langle \mathfrak{W}^2, \mathfrak{R}^2_{\mathbf{N}_1}, \ldots, \mathfrak{R}^2_{\mathbf{N}_n}\rangle$ *does not admit a pw-model. Likewise, for any other permutation and repetitions of the* $\mathbf{N}_i$*, if for every* $w \in \mathfrak{W}^2$ *there is such a tuple, then* $\langle \mathfrak{W}^2, \mathfrak{R}^2_{\mathbf{N}_1}, \ldots, \mathfrak{R}^2_{\mathbf{N}_n}\rangle$ *does not admit a pw-model. In other words, if there is a loop of j many 'steps', where each 'step' is given by some accessibility relation* $\mathfrak{R}_{\mathbf{N}_i}$ *for some* $1 \leq i \leq n$*, such that every world in a given multimodal frame admits said loop; then said frame does not admit a pw-model.*

Proof. The proof is similarly a model-theoretic version of the proof of the Multimodal Montague's Paradox (Theorem 4.1.4). Assume for contradiction that there are $\mathfrak{V}^2, \mathfrak{B}^2_{\mathbf{N}_1}, \ldots, \mathfrak{B}^2_{\mathbf{N}_n}$ such that

$$\mathfrak{M}^2 = \langle\langle \mathfrak{W}^2, \mathfrak{R}^2_{\mathbf{N}_1}, \ldots, \mathfrak{R}^2_{\mathbf{N}_n}\rangle, \mathbb{N}, \mathfrak{V}^2, \mathfrak{B}^2_{\mathbf{N}_1}, \ldots, \mathfrak{B}^2_{\mathbf{N}_n}\rangle$$

is a pw-model. By the Diagonal Lemma, we can find a λ such that

$$\mathsf{MSPL} \vdash \lambda \leftrightarrow \neg\overline{\mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\ldots\mathbf{N}_j\overline{\ulcorner\lambda\urcorner}^{\cdots}\urcorner\urcorner}$$

Since $\mathfrak{W}^2 \neq \emptyset$, pick an arbitrary $w \in \mathfrak{W}^2$. By soundness,

$$\mathfrak{M}^2(w) \vDash \lambda \leftrightarrow \neg\overline{\mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\ldots\mathbf{N}_j\overline{\ulcorner\lambda\urcorner}^{\cdots}\urcorner\urcorner}$$

If $\mathfrak{M}^2(w) \vDash \neg\lambda$, then $\mathfrak{M}^2(w) \vDash \overline{\mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\ldots\mathbf{N}_j\overline{\ulcorner\lambda\urcorner}^{\cdots}\urcorner\urcorner}$. By assumption, there is a tuple $\langle v_2^w, \ldots, v_j^w\rangle$ such that $v_2^w \in \mathfrak{R}_{\mathbf{N}_1}(w)$; $v_{i+1}^w \in \mathfrak{R}_{\mathbf{N}_i}(v_i^w)$, for $2 \leq i \leq j-1 \leq n-1$; and $w \in \mathfrak{R}_{\mathbf{N}_j}(v_j^w)$. Then, as in the Multimodal Postcard Pardox, by applications of $\mathbf{N}_i$ Truth Condition, for $1 \leq i \leq n$, we have that $\mathfrak{M}^2(w) \vDash \lambda$ ※.

Thus, $\mathfrak{M}^2(w) \vDash \lambda$ for all $w \in \mathfrak{W}^2$. Pick a $\mathsf{w} \in \mathfrak{W}^2$. Then, $\mathfrak{M}^2(\mathsf{w}) \vDash \lambda$. By the $\mathsf{NEC}_{\mathbf{N}_i}$ rules, we have that $\mathfrak{M}^2(\mathsf{w}) \vDash \overline{\mathbf{N}_1\ulcorner\mathbf{N}_2\ulcorner\ldots\mathbf{N}_j\overline{\ulcorner\lambda\urcorner}^{\cdots}\urcorner\urcorner}$. Hence, using the biconditional, $\mathfrak{M}^2(\mathsf{w}) \vDash \neg\lambda$ ※. Q.E.D.

In the operator approach, a multimodal frame satisfying the condition from Theorem 4.3.2 makes the schema

$$\Box_1\Box_2\ldots\Box_j\phi \rightarrow \phi \tag{MT}$$

true. As a corollary, we also have the following.

Corollary 4.3.1. *The schema* MT_j *is not true in any pw-model, for any* $1 \leq j \leq n$*. In particular, no condition on the* $\mathfrak{R}_{\mathbf{N}_i}$*, for* $1 \leq i \leq j$*, ensures that* MT_j *is true in a pw-model; and no (class of) frame(s) makes* MT_j *valid. Similarly for any other permutation of the* $\mathbf{N}_i$*.*

Proof. This follows from Theorem 4.1.4 and Corollary 4.2.3. Just like with Lemma 4.2.5, this corollary can also be proved using the proof of Theorem 4.3.2, except for replacing reference to the relevant tuple with the fact that MT_j is true in the pw-model we are considering for contradiction. Q.E.D.

Theorem 4.3.2 can be generalised even further by easing the restriction that the loop structure, i.e. which accessibility relation is associated with each step in the loop, is fixed for each world. That is, the tuples of worlds and accessibility relations from Theorem 4.3.2 can allow for repetitions and can differ between worlds.

Theorem 4.3.3 (General Model-Theoretic Multimodal Montague's Paradox)**.** *For a sequence* a*, let* $|\ \mathsf{a}\ |$ *denote the 'length' of a sequence, i.e. the number of elements in* a*; and let* $A^{\leq m}$ *for* $m \in \omega \setminus \{0\}$ *denote the set of sequences* a *of elements in A such that* $|\ \mathsf{a}\ | \leq m$*. Now, let* $k \in \omega \setminus \{0\}$ *and let* $\langle \mathfrak{W}^3, \mathfrak{R}^3_{\mathbf{N}_1}, \ldots, \mathfrak{R}^3_{\mathbf{N}_n} \rangle$ *be a multimodal frame such that for all* $w \in \mathfrak{W}^3$*, there are sequences* $b^w \in (\mathfrak{W}^3)^{\leq k}$ *and* $s^w \in \{1, \ldots, n\}^{\leq k}$ *such that*

I. $|\ b^w\ | = |\ s^w\ | \leq k$

II. $b^w_1 = w$*;*

III. For all $1 \leq i \leq |\ b^w\ |$ *we have that* $b^w_{(i+1 \mod |b^w|)} \in \mathfrak{R}_{\mathbf{N}_{s^w_i}}(a^w_i)$*.*

Then $\langle \mathfrak{W}^3, \mathfrak{R}^3_{\mathbf{N}_1}, \ldots, \mathfrak{R}^3_{\mathbf{N}_n} \rangle$ *does not admit a pw-model. In other words, if there is a fixed natural number k such that each world in a given multimodal frame admits some of loop of length k or less (but not none); then said frame does not admit a pw-model.*

Proof. The idea of the proof is that, fixing a natural number k, there are only finitely many different loops of length less than or equal to k. Thus, we can form a finite disjunction of all of these loops, as represented by some sequence of compounded modal syntactic predicates. We can then use the Diagonal Lemma to prove a contradiction. More precisely, assume for contradiction that there are $\mathfrak{V}^3, \mathfrak{B}^3_{\mathbf{N}_1}, \ldots, \mathfrak{B}^3_{\mathbf{N}_n}$ such that

$$\mathfrak{M}^3 = \langle \langle \mathfrak{W}^3, \mathfrak{R}^3_{\mathbf{N}_1}, \ldots, \mathfrak{R}^3_{\mathbf{N}_n} \rangle, \mathbb{N}, \mathfrak{V}^3, \mathfrak{B}^3_{\mathbf{N}_1}, \ldots, \mathfrak{B}^3_{\mathbf{N}_n} \rangle$$

is a pw-model. Let $B^{\mathfrak{W}^3}$ and $S^{\mathfrak{W}^3}$ be the sets of all sequences b^w and s^w respectively for $w \in \mathfrak{W}^3$ satisfying constraints *I-III*. By the Diagonal Lemma, we can find a λ such that

$$\mathsf{MSPL} \vdash \lambda \leftrightarrow \bigvee_{s \in S^{\mathfrak{W}^3}} \neg \mathbf{N}_{s_1} \overline{\ulcorner \mathbf{N}_{s_2} \overline{\ulcorner \ldots \mathbf{N}_{s_{|s|}} \overline{\ulcorner \overline{\lambda} \urcorner^{\cdots}} \urcorner} \urcorner}$$

This is a well-formed formula, because $|\, S^{\mathfrak{W}^3} \,| \leq |\, \{1, \ldots, n\}^{\leq k} \,| < \omega$. Since $\mathfrak{W}^3 \neq \emptyset$, pick an arbitrary $w \in \mathfrak{W}^3$. By soundness,

$$\mathfrak{M}^3(w) \vDash \lambda \leftrightarrow \bigvee_{s \in S^{\mathfrak{W}^3}} \neg \mathbf{N}_{s_1} \overline{\ulcorner \mathbf{N}_{s_2} \overline{\ulcorner \ldots \mathbf{N}_{s_{|s|}} \overline{\ulcorner \lambda \urcorner} \ldots \urcorner} \urcorner} \tag{4.71}$$

If $\mathfrak{M}^3(w) \vDash \neg\lambda$, then

$$\mathfrak{M}^3(w) \vDash \bigwedge_{s \in S^{\mathfrak{W}^3}} \mathbf{N}_{s_1} \overline{\ulcorner \mathbf{N}_{s_2} \overline{\ulcorner \ldots \mathbf{N}_{s_{|s|}} \overline{\ulcorner \lambda \urcorner} \ldots \urcorner} \urcorner} \tag{4.72}$$

By assumption, there are $b^w \in B^{\mathfrak{W}^3}$ and $s^w \in S^{\mathfrak{W}^3}$ satisfying the constraints *I-III*. In particular, we have that

$$\mathfrak{M}^3(w) \vDash \mathbf{N}_{s_1^w} \overline{\ulcorner \mathbf{N}_{s_2^w} \overline{\ulcorner \ldots \mathbf{N}_{s_{|s|}^w} \overline{\ulcorner \lambda \urcorner} \ldots \urcorner} \urcorner} \tag{4.73}$$

By $|\, b^w \,|$-many applications of $\mathbf{N}_i$ Truth Condition along with conditions *II* and *III*, from (4.73) we have that $\mathfrak{M}^3(w) \vDash \lambda$ ※.

Thus, $\mathfrak{M}^3(w) \vDash \lambda$ for all $w \in \mathfrak{W}^3$. Pick a $\mathsf{w} \in \mathfrak{W}^3$. Then, $\mathfrak{M}^3(\mathsf{w}) \vDash \lambda$. By applying the $\mathsf{NEC}_{\mathbf{N}_i}$ rules to λ sequentially for all $s \in S^{\mathfrak{W}^3}$, we have that

$$\mathfrak{M}^3(w) \vDash \bigwedge_{s \in S^{\mathfrak{W}^3}} \mathbf{N}_{s_1} \overline{\ulcorner \mathbf{N}_{s_2} \overline{\ulcorner \ldots \mathbf{N}_{s_{|s|}} \overline{\ulcorner \lambda \urcorner} \ldots \urcorner} \urcorner} \tag{4.74}$$

Hence, from (4.71) we have that $\mathfrak{M}^3(\mathsf{w}) \vDash \neg\lambda$ ※. Q.E.D.

Corollary 4.3.2 (Paradox of Invertibility)**.** *Let* $\langle \mathfrak{W}^4, \mathfrak{R}^4_{\mathbf{N}_1}, \ldots, \mathfrak{R}^4_{\mathbf{N}_n} \rangle$ *be a multimodal frame such that for some* $1 \leq i \leq n$ *we have that*

$$(\mathfrak{R}^4_{\mathbf{N}_i})^{-1} \subseteq \bigcup_{1 \leq k \leq n} \mathfrak{R}^4_{\mathbf{N}_k} \tag{4.75}$$

$$\mathfrak{R}^4_{\mathbf{N}_i}(w) \neq \emptyset, \forall w \in \mathfrak{W}^4 \tag{4.76}$$

Then $\langle \mathfrak{W}^4, \mathfrak{R}^4_{\mathbf{N}_1}, \ldots, \mathfrak{R}^4_{\mathbf{N}_n} \rangle$ *does not admit a pw-model. In other words, if a given multimodal frame has an accessibility relation that is non-empty at each world, and its inverse can be recovered using all, or some, of the other accessibility relations; then said frame does not admit a pw-model.*

Proof. $\langle \mathfrak{W}^4, \mathfrak{R}^4_{\mathbf{N}_1}, \ldots, \mathfrak{R}^4_{\mathbf{N}_n} \rangle$ satisfies the constraints *I-III*, and thus by Theorem 4.3.3 does not admit a pw-model. Q.E.D.

For the following group of results, we will consider a bimodal temporal logic, where we will expand $\mathcal{L}_{\mathsf{PA}}$ with two unary modal syntactic predicates $\mathbf{H}$ and $\mathbf{G}$, and denote it $\mathcal{L}_{\mathsf{PA}}^{\mathbf{H},\mathbf{G}}$. We will read $\mathbf{H}$ and $\mathbf{G}$ as 'it has been the case that' and 'it is going to be the case that' respectively. We abbreviate their duals by $\mathbf{P}$ and $\mathbf{F}$ respectively. We denote temporal bimodal frames by $\langle \mathfrak{T}, \mathfrak{R}_{\mathbf{H}}, \mathfrak{R}_{\mathbf{G}} \rangle$, where we will read the set of 'worlds' $\mathfrak{T}$ as points in time. We then have the following proof-theoretic paradox.

Theorem 4.3.4 (Proof-Theoretic No Future / Past Paradox)**.** *This paradox is due to Horsten and Leitgeb 2001. For $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{H},\mathbf{G}})$, both of the pairs of schemata*

$$\mathbf{G}\overline{\ulcorner \phi \urcorner} \rightarrow \mathbf{F}\overline{\ulcorner \phi \urcorner} \tag{$\mathsf{D}_{\mathbf{G}}$}$$

$$\phi \rightarrow \mathbf{G}\overline{\ulcorner \mathbf{P}\overline{\ulcorner \phi \urcorner} \urcorner} \tag{$\mathsf{IN}_{\mathbf{GH}}$}$$

or

$$\mathbf{H}\overline{\ulcorner \phi \urcorner} \rightarrow \mathbf{P}\overline{\ulcorner \phi \urcorner} \tag{$\mathsf{D}_{\mathbf{H}}$}$$

$$\phi \rightarrow \mathbf{H}\overline{\ulcorner \mathbf{F}\overline{\ulcorner \phi \urcorner} \urcorner} \tag{$\mathsf{IN}_{\mathbf{HG}}$}$$

are inconsistent. Intuitively, $\mathsf{D}_{\mathbf{G}}$ *says that if something is going to be the case, then at some point in the future said thing is going to be the case; and* $\mathsf{IN}_{\mathbf{GH}}$ *says that if something is the case, then it is going to be the case that at some point in the past said thing was the case. Similarly for* $\mathsf{D}_{\mathbf{H}}$ *and* $\mathsf{IN}_{\mathbf{HG}}$*. The inconsistency of the former pair is called the No Future Paradox; and the inconsistency of the latter pair is called the No Past Paradox.*

Proof. We will only prove the first claim; the second is proved by a mirrored argument. We will follow the structure of the proof of Fischer and Stern 2015, Theorem 10. Firstly, denote by $(\mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}))^{-}$ the theory of the axioms of PA formulated in $\mathcal{L}_{\mathsf{PA}}^{\mathbf{H},\mathbf{G}}$ and the axiom schemata $(\mathsf{D}_{\mathbf{G}})$, $(\mathsf{IN}_{\mathbf{GH}})$,

$$\mathbf{G}\overline{\ulcorner \phi \rightarrow \psi \urcorner} \rightarrow (\mathbf{G}\overline{\ulcorner \phi \urcorner} \rightarrow \mathbf{G}\overline{\ulcorner \psi \urcorner}) \tag{$\mathsf{K}_{\mathbf{G}}$}$$

and

$$\mathbf{H}\overline{\ulcorner \phi \rightarrow \psi \urcorner} \rightarrow (\mathbf{H}\overline{\ulcorner \phi \urcorner} \rightarrow \mathbf{H}\overline{\ulcorner \psi \urcorner}) \tag{$\mathsf{K}_{\mathbf{H}}$}$$

where $\phi, \psi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{H},\mathbf{G}})$. Denote by $\mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}})$ the closure of $(\mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}))^{-}$ under the rules

$$\mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}) \vdash \phi \Rightarrow \mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}) \vdash \mathbf{G}\overline{\ulcorner\phi\urcorner} \tag{$\mathsf{NEC}_{\mathbf{G}}$}$$

and

$$\mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}) \vdash \phi \Rightarrow \mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}) \vdash \mathbf{H}\overline{\ulcorner\phi\urcorner} \tag{$\mathsf{NEC}_{\mathbf{H}}$}$$

where $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{H},\mathbf{G}})$. We then have the following.

$$\mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}) \vdash \mathbf{G}\overline{\ulcorner\mathbf{H}\overline{\ulcorner\phi\urcorner}\urcorner} \rightarrow \mathbf{F}\overline{\ulcorner\mathbf{H}\overline{\ulcorner\phi\urcorner}\urcorner} \qquad (\mathsf{D}_{\mathbf{G}}) \tag{4.77}$$

$$\mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}) \vdash \neg\phi \rightarrow \mathbf{G}\overline{\ulcorner\mathbf{P}\overline{\ulcorner\neg\phi\urcorner}\urcorner} \qquad (\mathsf{IN}_{\mathbf{GH}}) \tag{4.78}$$

$$\mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}) \vdash \neg\mathbf{G}\overline{\ulcorner\neg\mathbf{H}\overline{\ulcorner\neg\neg\phi\urcorner}\urcorner} \rightarrow \phi \qquad (4.78) \tag{4.79}$$

$$\mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}) \vdash \mathbf{F}\overline{\ulcorner\mathbf{H}\overline{\ulcorner\phi\urcorner}\urcorner} \rightarrow \phi \qquad \mathsf{NEC}_{\mathbf{G}}, \mathsf{NEC}_{\mathbf{H}}, \mathsf{K}_{\mathbf{G}}, \mathsf{K}_{\mathbf{H}} \tag{4.80}$$

$$\mathsf{PA}_{\mathbf{H},\mathbf{G}} + (\mathsf{D}_{\mathbf{G}}) + (\mathsf{IN}_{\mathbf{GH}}) \vdash \mathbf{G}\overline{\ulcorner\mathbf{H}\overline{\ulcorner\phi\urcorner}\urcorner} \rightarrow \phi \qquad (4.77), (4.80) \tag{4.81}$$

Line (4.81) leads to a contradiction by Theorem 4.1.3 once appropriately translated into $\mathcal{L}_{\mathsf{PA}}^{\mathbf{H},\mathbf{G}}$. Q.E.D.

We have the following corollary.

Corollary 4.3.3. *For $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{H},\mathbf{G}})$, neither of the pairs of schemata [$(\mathsf{D}_{\mathbf{G}})$ and $(\mathsf{IN}_{\mathbf{GH}})$] or [$(\mathsf{D}_{\mathbf{H}})$ and $(\mathsf{IN}_{\mathbf{HG}})$] are true in any pw-model. In particular, no condition on $\mathfrak{R}_{\mathbf{H}}$ and $\mathfrak{R}_{\mathbf{G}}$ ensures that either of these pairs are true in a pw-model; and no (class of) frame(s) makes either of these pairs valid.*

Proof. This follows from Theorem 4.3.4 and Corollary 4.2.3. Q.E.D.

As before, we also have a model-theoretic version of the paradox of Horsten and Leitgeb 2001.

Corollary 4.3.4 (Model-Theoretic No Future / Past Paradox)**.** *Let $\langle \mathfrak{T}^1, \mathfrak{R}^1_{\mathbf{H}}, \mathfrak{R}^1_{\mathbf{G}} \rangle$ be a temporal bimodal frame such that*

$$\mathfrak{R}^1_{\mathbf{H}} = (\mathfrak{R}^1_{\mathbf{G}})^{-1} \textit{ iff } (\mathfrak{R}^1_{\mathbf{H}})^{-1} = \mathfrak{R}^1_{\mathbf{G}} \tag{4.82}$$

and either (i) for all $t \in \mathfrak{T}$ there is a $t' \in \mathfrak{R}^1_{\mathbf{G}}(t)$; or (ii) for all $t \in \mathfrak{T}$ there is a $t' \in \mathfrak{R}^1_{\mathbf{H}}(t)$. Then $\langle \mathfrak{T}^1, \mathfrak{R}^1_{\mathbf{H}}, \mathfrak{R}^1_{\mathbf{G}} \rangle$ does not admit a pw-model.[12]

Proof. This is a bimodal instance of Corollary 4.3.2. Q.E.D.

Given the binary relation $\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i}$, let us now define

$$(\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})(w) := \{v \in \mathfrak{W} \mid \langle w, v \rangle \in \bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i}\} \tag{4.83}$$

The examples we have considered so far have involved loops in $\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i}$. However, circularity is not the only condition that ensures non-existence results.

Theorem 4.3.5 (Multimodal '≤' Paradox)**.** *Let $\langle \mathfrak{W}^5, \mathfrak{R}^5_{\mathbf{N}_1}, \ldots, \mathfrak{R}^5_{\mathbf{N}_n} \rangle$ be a multimodal frame such that $\bigcup_{1 \leq i \leq j} \mathfrak{R}^5_{\mathbf{N}_i}$ is transitive for some $j \leq n$; and for all $w \in \mathfrak{W}^5$ we have $\bigcup_{1 \leq i \leq j} \mathfrak{R}^5_{\mathbf{N}_i}(w) \neq \emptyset$. Then $\langle \mathfrak{W}^5, \mathfrak{R}^5_{\mathbf{N}_1}, \ldots, \mathfrak{R}^5_{\mathbf{N}_n} \rangle$ does not admit a pw-model. In other words, if a given multimodal frame has some number of accessibility relations that taken together are transitive, and such that every world sees another world with one of these accessibility relations; then said frame does not admit a pw-model.*

Proof. Assume for contradiction that there are $\mathfrak{V}^5, \mathfrak{B}^5_{\mathbf{N}_1}, \ldots, \mathfrak{B}^5_{\mathbf{N}_n}$ such that

$$\mathfrak{M}^5 = \langle \langle \mathfrak{W}^5, \mathfrak{R}^5_{\mathbf{N}_1}, \ldots, \mathfrak{R}^5_{\mathbf{N}_n} \rangle, \mathbb{N}, \mathfrak{V}^5, \mathfrak{B}^5_{\mathbf{N}_1}, \ldots, \mathfrak{B}^5_{\mathbf{N}_n} \rangle$$

is a pw-model. By the Diagonal Lemma, we can find a λ such that

$$\mathsf{MSPL} \vdash \lambda \leftrightarrow \bigvee_{1 \leq i \leq j} \neg \mathbf{N}_i \overline{\ulcorner \lambda \urcorner}$$

Pick an arbitrary $w \in \mathfrak{W}^5$. By soundness,

$$\mathfrak{M}^5(w) \vDash \lambda \leftrightarrow \bigvee_{1 \leq i \leq j} \neg \mathbf{N}_i \overline{\ulcorner \lambda \urcorner}$$

If $\mathfrak{M}^5(w) \vDash \lambda$, then $\mathfrak{M}^5(w) \vDash \bigvee_{1 \leq i \leq j} \neg \mathbf{N}_i \overline{\ulcorner \lambda \urcorner}$. Hence, there is some $v \in \bigcup_{1 \leq i \leq j} \mathfrak{R}^5_{\mathbf{N}_i}(w)$ and $\mathfrak{M}^5(v) \vDash \neg\lambda$. Or $\mathfrak{M}^5(w) \vDash \neg\lambda$. In either case, there is some $u \in \mathfrak{W}^5$ with $\mathfrak{M}^5(u) \vDash \neg\lambda$, i.e. $\mathfrak{M}^5(u) \vDash \bigwedge_{1 \leq i \leq j} \mathbf{N}_i \overline{\ulcorner \lambda \urcorner}$. Now find $u' \in \bigcup_{1 \leq i \leq j} \mathfrak{R}^5_{\mathbf{N}_i}(u) \neq \emptyset$. We then have that $\mathfrak{M}^5(u') \vDash$

[12] Constraints (4.82) and (i) ensure $\mathsf{D}_{\mathbf{G}}$ and $\mathsf{IN}_{\mathbf{GH}}$ are true, which is the No Future Paradox; and constraints (4.82) and (ii) ensure $\mathsf{D}_{\mathbf{H}}$ and $\mathsf{IN}_{\mathbf{HG}}$ are true, which is the No Past Paradox.

λ, i.e. $\bigvee_{1\leq i\leq j}\neg\mathbf{N}_i\overline{\ulcorner\lambda\urcorner}$. Hence there is some $u'' \in \bigcup_{1\leq i\leq j}\mathfrak{R}^5_{\mathbf{N}_i}(u')$, and $\mathfrak{M}^5(u'') \vDash \neg\lambda$. Because $\bigcup_{1\leq i\leq j}\mathfrak{R}^5_{\mathbf{N}_i}$ is transitive, we have that $u'' \in \bigcup_{1\leq i\leq j}\mathfrak{R}^5_{\mathbf{N}_i}(u)$, and hence $\mathfrak{M}^5(u'') \vDash \lambda$ ※. Q.E.D.

Theorem 4.3.5 shows that not only loops, but also infinite chains in $\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i}$, i.e. the fact that there are no 'dead end' worlds that 'see' nothing; ensure non-existence results, assuming $\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i}$ is transitive. For example, the frame $\langle\mathbb{N},\leq\rangle$, which does not contain any loops but does contain infinite ascending chains, does not admit a pw-model. Furthermore, transitivity is not particularly crucial for the philosophical import of the inconsistency result from theorem 4.3.5. This is because we can, in a sense, drop the transitivity constraint.

Theorem 4.3.6 (Multimodal 'No Dead-End' Paradox)**.** *Let* $\langle\mathfrak{W}^6,\mathfrak{R}^6_{\mathbf{N}_1},\ldots,\mathfrak{R}^6_{\mathbf{N}_n}\rangle$ *be a multimodal frame such that for some* $j \leq n$ *we have that* $\bigcup_{1\leq i\leq j}\mathfrak{R}^6_{\mathbf{N}_i}(w) \neq \emptyset$ *for all* $w \in \mathfrak{W}^6$*. Then* $\langle\mathfrak{W}^6,\mathfrak{R}^6_{\mathbf{N}_1},\ldots,\mathfrak{R}^6_{\mathbf{N}_n}\rangle$ *does not admit a pw-model. In other words, if a given multimodal frame has some number of accessibility relations such that every world sees another world with one of these accessibility relations; then said frame does not admit a pw-model.*

Proof. This proof is the model-theoretic version of the proof of McGee's Paradox (1985). Assume for contradiction that there are $\mathfrak{V}^6,\mathfrak{B}^6_{\mathbf{N}_1},\ldots,\mathfrak{B}^6_{\mathbf{N}_n}$ such that

$$\mathfrak{M}^6 = \langle\langle\mathfrak{W}^6,\mathfrak{R}^6_{\mathbf{N}_1},\ldots,\mathfrak{R}^6_{\mathbf{N}_n}\rangle,\mathbb{N},\mathfrak{V}^6,\mathfrak{B}^6_{\mathbf{N}_1},\ldots,\mathfrak{B}^6_{\mathbf{N}_n}\rangle$$

is a pw-model. Since pw-models are built up from $\mathbb{N}$, by Corollary 4.2.3 any pw-model is also sound with respect to the theory that extends MSPL with the schemata

$$\forall x\mathbf{N}_i\overline{\ulcorner\phi(\overline{\underset{\bullet}{x}})\urcorner} \rightarrow \mathbf{N}_i\overline{\ulcorner\forall x\phi(x)\urcorner} \tag{4.84}$$

where $\phi \in \mathsf{Form}(\mathcal{L}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}_{\mathsf{PA}})$, for all $1 \leq i \leq n$. Note that for $\phi \in \mathsf{Form}(\mathcal{L}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}_{\mathsf{PA}})$ this theory proves

$$\forall x \bigwedge_{1\leq i\leq j}\mathbf{N}_i\overline{\ulcorner\phi(\overline{\underset{\bullet}{x}})\urcorner} \rightarrow \bigwedge_{1\leq i\leq j}\mathbf{N}_i\overline{\ulcorner\forall x\phi(x)\urcorner} \tag{$\mathsf{FO}_{\mathbf{N}_j}$}$$

Call $\mathsf{PA}^{\omega}_{\mathsf{D}}$ the theory that extends MSPL with the axiom schema

$$\bigwedge_{1\leq i\leq j} \mathbf{N}_i\overline{\ulcorner\phi\urcorner} \rightarrow \neg \bigwedge_{1\leq i\leq j} \mathbf{N}_i\overline{\ulcorner\neg\phi\urcorner} \tag{$\mathsf{WD}_{\mathbf{N}_j}$}$$

where $\phi \in \mathsf{Form}(\mathcal{L}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}_{\mathsf{PA}})$, and the axiom schemata (4.84) for all $1 \leq i \leq n$; and is closed under the $\mathsf{NEC}_{\mathbf{N}_i}$ rules, for all $1 \leq i \leq n$. Since there is a $j \leq n$ such that $\bigcup_{1\leq i\leq j} \mathfrak{R}^6_{\mathbf{N}_i}(w) \neq \emptyset$ for all $w \in \mathfrak{W}^6$, by Corollary 4.2.3 we have that $\mathfrak{M}^6$ is sound with respect to $\mathsf{PA}^{\omega}_{\mathsf{D}}$.

We now follow the structure of part of the proof of McGee's Paradox. By an application of the Uniform Diagonal Lemma, let $F^*(x, y, z)$ be a formula such that

$$\begin{aligned}\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \forall x\forall y\forall z(F^*(x, y, z) \leftrightarrow \\ \exists m(x = \mathsf{S}(m) \wedge y = \overline{\ulcorner\forall y(F^*(\overline{\underset{\bullet}{m}}, y, \overline{\underset{\bullet}{z}}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y)\urcorner}) \\ \vee (x = \underline{0} \wedge y = z))\end{aligned} \tag{F^*}$$

By abuse of notation, $F^*(x, y, z)$ says that y codes the sentence which 'prefixes x many instances of '$\bigwedge_{1\leq i\leq j} \mathbf{N}_i$' to z' (along with the proper number of nested quotations). This is an abuse of notation, because '$\bigwedge_{1\leq i\leq j} \mathbf{N}_i$' is an abbreviation and not in the object language.

By the Diagonal Lemma, we can find a λ such that

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \lambda \leftrightarrow \neg\forall x\forall y(F^*(x, y, \overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y) \tag{4.85}$$

λ says that the result of prefixing '$\bigwedge_{1\leq i\leq j} \mathbf{N}_i$' some number of times to 'λ' does not hold. We then have the following:

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \lambda \leftrightarrow \neg\forall x\forall y(F^*(x, y, \overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y) \tag{4.86}$$

(4.85)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \neg\lambda \rightarrow \forall y(F^*(\underline{0}, y, \overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y) \tag{4.87}$$

(4.86)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \forall y(F^*(\underline{0}, y, \overline{\ulcorner\lambda\urcorner}) \leftrightarrow y = \overline{\ulcorner\lambda\urcorner}) \tag{4.88}$$

(F^*)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \neg\lambda \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i\overline{\ulcorner\lambda\urcorner} \tag{4.89}$$

(4.87), (4.88)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \lambda \rightarrow \neg\forall x\forall y(F^*(x,y,\overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y) \tag{4.90}$$

(4.86)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \bigwedge_{1\leq i\leq j} \mathbf{N}_i\overline{\ulcorner\lambda \rightarrow \neg\forall x\forall y(F^*(x,y,\overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y)\urcorner} \tag{4.91}$$

(4.90); $\mathsf{NEC}_{\mathbf{N}_i}$, $1 \leq i \leq j$

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \bigwedge_{1\leq i\leq j} \mathbf{N}_i\overline{\ulcorner\lambda\urcorner} \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i\overline{\ulcorner\neg\forall x\forall y(F^*(x,y,\overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y)\urcorner} \tag{4.92}$$

(4.91); $\mathsf{K}_{\mathbf{N}_i}$, $1 \leq i \leq j$

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \neg\lambda \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i\overline{\ulcorner\neg\forall x\forall y(F^*(x,y,\overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y)\urcorner} \tag{4.93}$$

(4.89), (4.92)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \neg \bigwedge_{1\leq i\leq j} \mathbf{N}_i\overline{\ulcorner\neg\forall x\forall y(F^*(x,y,\overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y)\urcorner} \rightarrow \lambda \tag{4.94}$$

(4.93)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \bigwedge_{1\leq i\leq j} \mathbf{N}_i\overline{\ulcorner\forall x\forall y(F^*(x,y,\overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y)\urcorner} \rightarrow \lambda \tag{4.95}$$

($\mathsf{WD}_{\mathbf{N}_j}$), (4.94)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \forall x \bigwedge_{1\leq i\leq j} \mathbf{N}_i\overline{\ulcorner\forall y(F^*(\overset{\cdot}{x},y,\overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y)\urcorner} \rightarrow \lambda \tag{4.96}$$

($\mathsf{FO}_{\mathbf{N}_j}$), (4.95)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \forall x\forall y(F^*(\mathsf{S}(x),y,\overline{\ulcorner\lambda\urcorner}) \leftrightarrow y = \overline{\ulcorner\forall y(F^*(\overset{\cdot}{x},y,\overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y)\urcorner}) \tag{4.97}$$

(F^*)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \forall x\forall y(F^*(\mathsf{S}(x),y,\overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y) \rightarrow \lambda \tag{4.98}$$

(4.96), (4.97)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \forall x\forall y(F^*(x,y,\overline{\ulcorner\lambda\urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y) \rightarrow \lambda \tag{4.99}$$

(4.98)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \neg\lambda \rightarrow \lambda \tag{4.100}$$

(4.86), (4.99)

$$\mathsf{PA}^{\omega}_{\mathsf{D}} \vdash \lambda \tag{4.101}$$

(4.100)

Since $\mathfrak{M}^6$ is sound with respect to $\mathsf{PA}^{\omega}_{\mathsf{D}}$, we have that $\mathfrak{M}^6(w) \vDash \lambda$, for all $w \in \mathfrak{W}^6$. By the $\mathsf{NEC}_{\mathbf{N}_i}$ rules, for $1 \leq i \leq j$, we have that $\mathfrak{M}^6(w) \vDash \bigwedge_{1\leq i\leq j} \mathbf{N}_i \overline{\ulcorner \lambda \urcorner}$, for all $w \in \mathfrak{W}^6$. And similarly by induction we can prove that for any $n \in \mathbb{N}$, again abusing notation, prefixing n many instances of '$\bigwedge_{1\leq i\leq j} \mathbf{N}_i$' to '$\lambda$' is true at all $w \in \mathfrak{W}^6$. That is, for any $n \in \mathbb{N}$ and any $w \in \mathfrak{W}^6$, we have that

$$\mathfrak{M}^6(w) \vDash \forall y(F^*(\overline{n}, y, \overline{\ulcorner \lambda \urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y) \tag{4.102}$$

Again, since every pw-model is built up from $\mathbb{N}$, for any $w \in \mathfrak{W}^6$ we have that

$$\mathfrak{M}^6(w) \vDash \forall x \forall y(F^*(x, y, \overline{\ulcorner \lambda \urcorner}) \rightarrow \bigwedge_{1\leq i\leq j} \mathbf{N}_i y) \tag{4.103}$$

By line (4.86), we see that, for any $w \in \mathfrak{W}^6$, we have that

$$\mathfrak{M}^6(w) \vDash \neg\lambda \tag{4.104}$$

This is a contradiction with the fact that $\mathfrak{M}^6(w) \vDash \lambda$, for any $w \in \mathfrak{W}^6$ ※. Q.E.D.

Crucial to the proof of Theorem 4.3.6 is the fact that we are assuming pw-models are built up from $\mathbb{N}$, i.e. the standard model of arithmetic. This is why we can drop the transitivity constraint of Theorem 4.3.5 in Theorem 4.3.6. In particular, McGee's Paradox is not an outright inconsistency, but an ω-inconsistency; where a theory E extending PA is ω-inconsistent iff there is a formula $\phi(x)$ with $\mathsf{E} \vdash \phi(\overline{n})$ for each $n \in \mathbb{N}$, yet $\mathsf{E} \vdash \neg\forall x\phi(x)$. Such a theory E can be consistent, but $\mathbb{N}$ is not one of its models, all of which contain non-standard natural numbers. There might be 'pw-models' built up from non-standard models of the natural numbers such that for all $w \in \mathfrak{W}^3$ we have $\bigcup_{1\leq i\leq j} \mathfrak{R}_{\mathbf{N}_i}(w) \neq \emptyset$. However, the problem with such models is that the arithmetic vocabulary cannot be given its intended interpretation $\mathbb{N}$.

Finally, the non-existence results we have considered so far can be generalised from results about frames to results about what we will call full subframes.

Definition 4.3.1. Let $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ be a multimodal frame. We say that $\langle \mathfrak{W}', \mathfrak{R}'_{\mathbf{N}_1}, \ldots, \mathfrak{R}'_{\mathbf{N}_n} \rangle$ is a full subframe of $\mathfrak{F}$ iff $\mathfrak{W}' \subseteq \mathfrak{W}$ and for every $1 \leq i \leq n$ and every $w \in \mathfrak{W}'$ we have that $\mathfrak{R}'_{\mathbf{N}_i}(w) = \mathfrak{R}_{\mathbf{N}_i}(w)$.

The idea of a full subframe is that, once you 'enter' a full subframe, there is no 'leaving' the full subframe.

Theorem 4.3.7 (Preservation of PW-Models Under Full Subframes)**.** *Let* $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ *be a multimodal frame,* $\mathfrak{F}' = \langle \mathfrak{W}', \mathfrak{R}'_{\mathbf{N}_1}, \ldots, \mathfrak{R}'_{\mathbf{N}_n} \rangle$ *be a full subframe of* $\mathfrak{F}$*, and* $\mathfrak{M} = \langle \mathfrak{F}, \mathbb{N}, \mathfrak{V}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n} \rangle$ *be a pw-model. Then* $\mathfrak{M}\big|_{\mathfrak{F}'} := \langle \mathfrak{F}', \mathbb{N}, \mathfrak{V}\big|_{\mathfrak{W}'}, \mathfrak{B}_{\mathbf{N}_1}\big|_{\mathfrak{W}'}, \ldots, \mathfrak{B}_{\mathbf{N}_n}\big|_{\mathfrak{W}'} \rangle$ *is also a pw-model.*

Proof. The proof idea is to exploit the fact that once you 'enter' a full subframe, there is no 'leaving' the full subframe; and the fact that the modal syntactic predicates can only 'look forward', and not 'backward'. Thus, a modal syntactic predicate cannot 'tell the difference' between a full subframe and the larger frame. More precisely, looking at Definition 4.2.3 the only thing that we have to check is that for every $w \in \mathfrak{W}'$ and for every $1 \leq i \leq n$ we have that $\mathfrak{B}_{\mathbf{N}_i}\big|_{\mathfrak{W}'}$ satisfies the condition

$$\mathfrak{B}_{\mathbf{N}_i}\big|_{\mathfrak{W}'}(w) = \ulcorner \{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \mid \text{for each } v \in \mathfrak{R}'_{\mathbf{N}_i}(w),\ \mathfrak{M}\big|_{\mathfrak{F}'}(v) \vDash \phi\} \urcorner \tag{4.105}$$

where we denote $\mathfrak{M}\big|_{\mathfrak{F}'}(v) := \langle \mathbb{N}, \mathfrak{V}\big|_{\mathfrak{W}'}(w), \mathfrak{B}_{\mathbf{N}_1}\big|_{\mathfrak{W}'}(w), \ldots, \mathfrak{B}_{\mathbf{N}_n}\big|_{\mathfrak{W}'}(w) \rangle$.

Fix some $w \in \mathfrak{W}'$ and $1 \leq i \leq n$. By definition, since $w \in \mathfrak{W}' \subseteq \mathfrak{W}$ we have that $\mathfrak{B}_{\mathbf{N}_i}\big|_{\mathfrak{W}'}(w) = \mathfrak{B}_{\mathbf{N}_i}(w)$. Now since $\mathfrak{B}_{\mathbf{N}_i}$ satisfies $\mathsf{Intended}\ \mathbf{N}_i\text{-}\mathsf{Interpretation}$ we have that

$$\mathfrak{B}_{\mathbf{N}_i}\big|_{\mathfrak{W}'}(w) = \ulcorner \{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \mid \text{for each } v \in \mathfrak{R}_{\mathbf{N}_i}(w),\ \mathfrak{M}(v) \vDash \phi\} \urcorner \tag{4.106}$$

Now since $\mathfrak{F}'$ is a full subframe of $\mathfrak{F}$ we have that $\mathfrak{R}'_{\mathbf{N}_i}(w) = \mathfrak{R}_{\mathbf{N}_i}(w) \subseteq \mathfrak{W}'$. Thus, we have that

$$\mathfrak{B}_{\mathbf{N}_i}\big|_{\mathfrak{W}'}(w) = \ulcorner \{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \mid \text{for each } v \in \mathfrak{R}'_{\mathbf{N}_i}(w),\ \mathfrak{M}(v) \vDash \phi\} \urcorner \tag{4.107}$$

Further, for every $v \in \mathfrak{R}'_{\mathbf{N}_i}(w) \subseteq \mathfrak{W}'$ and every $1 \leq i \leq n$ we have that $\mathfrak{B}_{\mathbf{N}_i}\big|_{\mathfrak{W}'}(v) = \mathfrak{B}_{\mathbf{N}_i}(v)$, and $\mathfrak{V}\big|_{\mathfrak{W}'}(v) = \mathfrak{V}(v)$. Therefore, observe that

$$\mathfrak{M}(v) = \langle \mathbb{N}, \mathfrak{V}(v), \mathfrak{B}_{\mathbf{N}_1}(v), \ldots, \mathfrak{B}_{\mathbf{N}_n}(v) \rangle$$

$$= \langle \mathbb{N}, \mathfrak{B}\big|_{\mathfrak{W}'}(v), \mathfrak{B}_{\mathbf{N}_1}\big|_{\mathfrak{W}'}(v), \ldots, \mathfrak{B}_{\mathbf{N}_n}\big|_{\mathfrak{W}'}(v)\rangle$$
$$= \mathfrak{M}\big|_{\mathfrak{F}'}(v) \tag{4.108}$$

Putting (4.107) and (4.108) together we get that

$$\mathfrak{B}_{\mathbf{N}_i}\big|_{\mathfrak{W}'}(w) = \ulcorner\{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \mid \text{for each } v \in \mathfrak{R}'_{\mathbf{N}_i}(w),\ \mathfrak{M}\big|_{\mathfrak{F}'}(v) \vDash \phi\}\urcorner \tag{4.109}$$

as required. Q.E.D.

We can think of Theorem 4.3.7 as stating that for any multimodal frame we can 'excise' in a meaningful sense any of its full subframes. In particular, we have the following corollary.

Corollary 4.3.5. *If a multimodal frame $\mathfrak{F}$ has a full subframe $\mathfrak{F}'$ that does not admit a pw-model, then neither does $\mathfrak{F}$. In particular, if a multimodal frame $\mathfrak{F}$ has a full subframe $\mathfrak{F}'$ that satisfies any of the conditions of the non-existence results discussed so far in this chapter, then $\mathfrak{F}$ does not admit a pw-model.*

For example, by Corollary 4.3.5 we see that the following multimodal frames do not admit a pw-model:

- A multimodal frame that contains a world that accesses, for any $\mathfrak{R}_{\mathbf{N}_i}$, no other world but itself.
- A multimodal frame that contains an n-loop where each world in the loop accesses, for any $\mathfrak{R}_{\mathbf{N}_i}$, no other world but the next world in the loop.
- A multimodal frame that contains an infinite chain where each world in the chain accesses, for any $\mathfrak{R}_{\mathbf{N}_i}$, no other world but the next world in the chain.

So far, we have seen that both circularity and infinite chains, i.e. converse non-well-foundedness can cause frames to not admit pw-models. We will make this precise in the next section 4.4.

However, before we do that, we will consider one last group of paradoxes which shows that when expanding the classical possible-world P-semantics to a classical two-dimensional

possible-world P-semantics, no syntactic predicate can have the same truth-conditions as the sentential operators @, ×, or F have in the two-dimensional possible-world O-semantics. However, these indexicals are widespread in contemporary philosophical discourse. See for instance Davies and Humberstone 1980, Evans 1979, and Stalnaker 1978.

Briefly, two-dimensional semantics expands the interpretation of any formula to a pair of worlds, as opposed to a single world. The first world in this pair is called the 'reference world'; and the second world in this pair is called the 'evaluation world'. As before, the evaluation world is the world we evaluate a formula from. What is new is the reference world, which acts as a temporary actual world. Using such pairs of worlds we can talk about actuality. Actuality at first may seem dispensable. 'Actually, I have light blond hair' (or 'It is actual that I have light blond hair') just is 'I have light blond hair'. However, actuality is indispensable when occurring in the complement of a modal statement. Consider for example the true modal statement 'It is necessary that if I have light blond hair, then I have light blond hair'. The meaning of this modal statement can change when we add actuality. Consider the modal statement 'It is necessary that if I have light blond hair, then I actually have light blond hair'. This is in fact false, since I do not have light blond hair.

In the two-dimensional possible-world O-semantics, three sentential operators are introduced to allow for talk about actuality; @, for actually itself; ×, which makes the evaluation world the actual world, thereby allowing us to change which world counts as the actual world; and F, which allows us to quantify over every world as the actual world. We will show that @, ×, and F when suitably introduced in the two-dimensional possible-world P-semantics lead to inconsistency. That is, two-dimensional modal logic works radically different in the possible-world P-semantics than in the possible-world O-semantics. These paradoxes are unlike the merging paradoxes we have considered so far. However, in a sense they do similarly involve converse non-well-founded conditions on frames. Furthermore, they highlight an interesting limitation of possible-world P-semantics.

To introduce @, $\times$, and F we must first expand the possible-world P-semantics we have constructed.

Definition 4.3.2 (2D pw-model)**.** A 2D pw-model is a tuple

$$\mathfrak{M} = \langle\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle, \mathbb{N}, \mathfrak{V}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n}\rangle$$

where:

- $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle$ is a multimodal frame;
- $\mathfrak{V}$ is the interpretation of p and a function from $\mathfrak{W} \times \mathfrak{W}$ to $\{0, 1\}$;
- $\mathfrak{B}_{\mathbf{N}_i}$ is a function from $\mathfrak{W} \times \mathfrak{W}$ to $\mathcal{P}(\ulcorner \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})\urcorner)$, and we call it a two-dimensional $\mathbf{N}_i$-interpretation, for $1 \leq i \leq n$.
- We denote $\mathfrak{M}(w, v) := \langle \mathbb{N}, \mathfrak{V}(w, v), \mathfrak{B}_{\mathbf{N}_1}(w, v), \ldots, \mathfrak{B}_{\mathbf{N}_n}(w, v)\rangle$, for each $w, v \in \mathfrak{W}$. We read w as the world of reference, and v as the world of evaluation. Then, for all $w, v \in \mathfrak{W}$ we require that all $\mathfrak{B}_{\mathbf{N}_i}$ satisfy the condition

$$\mathfrak{B}_{\mathbf{N}_i}(w, v) = \ulcorner\{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \mid \text{for each } u \in \mathfrak{R}_{\mathbf{N}_i}(v),\ \mathfrak{M}(w, u) \vDash \phi\}\urcorner$$

(Intended Two-Dimensional $\mathbf{N}_i$-Interpretation)

As before, we talk about multimodal frames admitting 2D pw-models (on every interpretation). However, so far nothing is introduced which interacts with the reference worlds, making 2D pw-models just pw-models with an empty parameter.

Let us first expand our language $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$ by adding a predicate '$\mathbf{A}$', calling this $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{A}}$. Furthermore, we will expand our 2D pw-models so that the 2D pw-models are defined on $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{A}}$; and to include an $\mathbf{A}$-interpretation $\mathfrak{B}_{\mathbf{A}}$. As with the $\mathbf{N}_i$-interpretations, we need to add a condition to $\mathfrak{B}_{\mathbf{A}}$, so that it behaves like @ in the two-dimensional possible-world O-semantics; the aim is to have $\mathbf{A}$ make the reference world the evaluation world. We achieve this as follows. Firstly, we require that $\mathfrak{B}_{\mathbf{A}}$ is a function from $\mathfrak{W} \times \mathfrak{W}$ to $\mathcal{P}(\mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{A}}))$. Secondly, for all $w, v \in \mathfrak{W}$, we require that $\mathfrak{B}_{\mathbf{A}}$ satisfies the condition

$$\mathfrak{B}_{\mathbf{A}}(w, v) = \{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{A}}) \mid \mathfrak{M}(w, w) \vDash \phi\} \tag{4.110}$$

We call an expansion of a 2D pw-model that includes an **A**-interpretation satisfying (4.110) a $2D_{\mathbf{A}}$ pw-model. We then have the following limitative result.

Theorem 4.3.8 (Paradox of Diagonal Reference)**.** *No multimodal frame admits a $2D_A$ pw-model.*

Proof. This is a model-theoretic version of the Liar Paradox, but for actuality. Let $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ be a multimodal frame. Assume for contradiction that it does admit a $2D_{\mathbf{A}}$ pw-model. Since $\mathfrak{W} \neq \emptyset$, pick some $\mathsf{w} \in \mathfrak{W}$. By the Diagonal Lemma, we can find a λ such that

$$\mathsf{MSPL} \vdash \lambda \leftrightarrow \mathbf{A}\overline{\ulcorner \neg \lambda \urcorner}$$

By soundness

$$\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \lambda \leftrightarrow \mathbf{A}\overline{\ulcorner \neg \lambda \urcorner}$$

Then $\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \lambda$ iff $\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \mathbf{A}\overline{\ulcorner \neg \lambda \urcorner}$ iff by condition (4.110) $\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \neg\lambda$. Since either $\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \lambda$ or $\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \neg\lambda$, we have a contradiction ※. Q.E.D.

Thus, in the two-dimensional possible-world P-semantics no syntactic predicate can have the same truth-conditions as the sentential operator @ has in the two-dimensional possible-world O-semantics. Let us now expand our language $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$ by adding a predicate '**C**', calling this $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{C}}$. As with $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{A}}$, we will expand our 2D pw-models so that the 2D pw-models are defined on $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{C}}$; and to include a **C**-interpretation $\mathfrak{B}_{\mathbf{C}}$. We need to add a condition to $\mathfrak{B}_{\mathbf{C}}$, so that it behaves like $\times$ in the two-dimensional possible-world O-semantics; the aim is to have **C** make the evaluation world the reference world. We achieve this as follows. Firstly, we require that $\mathfrak{B}_{\mathbf{C}}$ is a function from $\mathfrak{W} \times \mathfrak{W}$ to $\mathcal{P}(\mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{C}}))$. Secondly, for all $w, v \in \mathfrak{W}$, we require that $\mathfrak{B}_{\mathbf{C}}$ satisfies the condition

$$\mathfrak{B}_{\mathbf{C}}(w, v) = \{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{C}}) \mid \mathfrak{M}(v, v) \vDash \phi\}. \tag{4.111}$$

We call an expansion of a 2D pw-model that includes a **C**-interpretation satisfying (4.111) a $2D_{\mathbf{C}}$ pw-model. We then have the following limitative result.

Theorem 4.3.9 (Paradox of Diagonal Evaluation)**.** *No multimodal frame admits a 2D$_{\mathbf{C}}$ pw-model.*

Proof. The proof is analogous to the proof of Theorem 4.3.8, where '**A**' is substituted by '**C**'. More specifically, let $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ be a multimodal frame. Assume for contradiction that it does admit a 2D$_{\mathbf{C}}$ pw-model. Since $\mathfrak{W} \neq \emptyset$, pick some $\mathsf{w} \in \mathfrak{W}$. By the Diagonal Lemma, we can find a λ such that

$$\mathsf{MSPL} \vdash \lambda \leftrightarrow \mathbf{C}^{\overline{\ulcorner \neg \lambda \urcorner}}$$

By soundness

$$\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \lambda \leftrightarrow \mathbf{C}^{\overline{\ulcorner \neg \lambda \urcorner}}$$

Then $\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \lambda$ iff $\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \mathbf{C}^{\overline{\ulcorner \neg \lambda \urcorner}}$ iff by condition (4.111) $\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \neg\lambda$. Since either $\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \lambda$ or $\mathfrak{M}(\mathsf{w}, \mathsf{w}) \vDash \neg\lambda$, we have a contradiction ※. Q.E.D.

Thus, in the two-dimensional possible-world P-semantics no syntactic predicate can have the same truth-conditions as the sentential operator $\times$ has in the two-dimensional possible-world O-semantics. Let us finally expand our language $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$ by adding a predicate '**Fix**', calling this $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{Fix}}$. As with $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{A}}$ and $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{C}}$, we will expand our 2D pw-models so that the 2D pw-models are defined on $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{Fix}}$; and to include a **Fix**-interpretation $\mathfrak{B}_{\mathbf{Fix}}$. We need to add a condition to $\mathfrak{B}_{\mathbf{Fix}}$, so that it behaves like F in the two-dimensional possible-world O-semantics; the aim is to have **Fix** universally quantify over the reference worlds. We achieve this as follows. Firstly, we require that $\mathfrak{B}_{\mathbf{Fix}}$ is a function from $\mathfrak{W} \times \mathfrak{W}$ to $\mathcal{P}(\mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{Fix}}))$. Secondly, for all $w, v \in \mathfrak{W}$, we require that $\mathfrak{B}_{\mathbf{Fix}}$ satisfies the condition

$$\mathfrak{B}_{\mathbf{Fix}}(w, v) = \{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n,\mathbf{Fix}}) \mid \mathfrak{M}(w', v) \vDash \phi \text{ for each } w' \in \mathfrak{W}\}. \tag{4.112}$$

We call an expansion of a 2D pw-model that includes a **Fix**-interpretation satisfying (4.112) a 2D$_{\mathbf{Fix}}$ pw-model. We then have the following limitative result.

Theorem 4.3.10 (Paradox of Fixation)**.** *No multimodal frame admits a 2D$_{\mathbf{Fix}}$ pw-model.*

Proof. Let $\langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ be a multimodal frame. Assume for contradiction that it does admit a $2\mathrm{D}_{\mathbf{Fix}}$ pw-model. Since $\mathfrak{W} \neq \emptyset$, pick some $\mathsf{v} \in \mathfrak{W}$. By the Diagonal Lemma, we can find a λ such that

$$\mathsf{MSPL} \vdash \lambda \leftrightarrow \neg\mathbf{Fix}\overline{\ulcorner\lambda\urcorner}$$

By Soundness

$$\mathfrak{M}(\mathsf{v}, \mathsf{v}) \vDash \lambda \leftrightarrow \neg\mathbf{Fix}\overline{\ulcorner\lambda\urcorner}$$

If $\mathfrak{M}(\mathsf{v}, \mathsf{v}) \vDash \lambda$, then $\mathfrak{M}(\mathsf{v}, \mathsf{v}) \vDash \neg\mathbf{Fix}\overline{\ulcorner\lambda\urcorner}$. Hence by condition (4.112) there is some $\mathsf{w} \in \mathfrak{W}$ such that $\mathfrak{M}(\mathsf{w}, \mathsf{v}) \vDash \neg\lambda$, i.e. $\mathfrak{M}(\mathsf{w}, \mathsf{v}) \vDash \mathbf{Fix}\overline{\ulcorner\lambda\urcorner}$. Thus by condition (4.112) $\mathfrak{M}(\mathsf{w}, \mathsf{v}) \vDash \lambda$ ※.

If $\mathfrak{M}(\mathsf{v}, \mathsf{v}) \vDash \neg\lambda$, then $\mathfrak{M}(\mathsf{v}, \mathsf{v}) \vDash \mathbf{Fix}\overline{\ulcorner\lambda\urcorner}$. Thus by condition (4.112) $\mathfrak{M}(\mathsf{v}, \mathsf{v}) \vDash \lambda$ ※.

Q.E.D.

Thus, in the two-dimensional possible-world P-semantics no syntactic predicate can have the same truth-conditions as the sentential operator F has in the two-dimensional possible-world O-semantics. One might suggest weakening condition (4.112) so as to not quantify over all reference worlds, instead introducing an 'accessibility relation', and quantifying over all accessible worlds. In effect, this semantics is the same as for for the predicates $\mathbf{N}_i$, for $1 \leq i \leq n$. Thus, similarly, unimodal conditions on the accessibility relations $\mathfrak{R}_{\mathbf{N}_i}$ that lead to paradox will also lead to paradox for **Fix**.

Taken together, Theorems 4.3.8-4.3.10 show us that in the two-dimensional possible-world P-semantics no syntactic predicate can have the same truth-conditions as the sentential operators @, $\times$, or F have in the two-dimensional possible-world O-semantics.

Let us end this section by observing the limitations of the possible-world P-semantics we have constructed as a means of analysing merging paradoxes; we can only analyse merging paradoxes where all $\mathbf{N}_i$ are normal, i.e. extend MSPL. This is genuinely limiting, since there are examples of merging paradoxes that collapse to simpler ones, i.e. lose their subtlety or complexity, when all of the involved modalities are normal.

Theorem 4.3.11. *Let* $j_1 \leq n \geq j_2$. *Let* E *be a theory that contains the schemata*

$$\mathbf{N}_{j_1}\overline{\ulcorner\phi\urcorner} \rightarrow \phi \tag{4.113}$$

$$\phi \rightarrow \neg \mathbf{N}_{j_2} \overline{\ulcorner \neg \mathbf{N}_{j_1} \overline{\ulcorner \phi \urcorner} \urcorner} \tag{4.114}$$

where $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_{j_1}, \mathbf{N}_{j_2}})$*; and is closed under the rule*

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash \mathbf{N}_{j_2} \overline{\ulcorner \phi \urcorner} \tag{$\mathsf{NEC}_{\mathbf{N}_{j_2}}$}$$

where $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_{j_1}, \mathbf{N}_{j_2}})$*. Then* E *is inconsistent.*

For a proof see Halbach 2008. If both $\mathbf{N}_{j_1}$ and $\mathbf{N}_{j_2}$ were normal, then this inconsistency would lose its complexity and collapse into Montague's Paradox. To further highlight this limitation, Theorem 4.3.11 has, in a certain sense, little complexity; in a slightly different framework, Fischer and Stern 2015 argue that this paradox does not arise from the *interaction* between $\mathbf{N}_{j_1}$ and $\mathbf{N}_{j_2}$. In particular, they show that this paradox can be proved using the diagonal sentence $\lambda \leftrightarrow \neg \mathbf{N}_{j_1} \overline{\ulcorner \lambda \urcorner}$, which involves but a single syntactic modal predicate. Compare this with the No Future Paradox of Horsten and Leitgeb 2001 (Theorem 4.3.4), which Fischer and Stern 2015 prove can only be proved with a diagonal sentence involving both $\mathbf{N}_{j_1}$ and $\mathbf{N}_{j_2}$; there is no inconsistency if we have diagonal sentences involving only $\mathbf{N}_{j_1}$, and only $\mathbf{N}_{j_2}$.

4.4 Multimodal Strong Characterisation Theorem

In this section we will provide necessary and sufficient conditions for which multimodal frames admit a pw-model on every interpretation. We will do this by generalising the Strong Characterisation Theorem (4.2.2) to the multimodal setting. Recall that Theorem 4.2.2 states that a unimodal frame admits a pw-model on every interpretation iff the frame is converse well-founded. Note that this is a slightly different result to the merging paradoxes we have considered in the previous section 4.3; there we have shown that certain classes of multimodal frames do not admit pw-models on any interpretation.

Before we can generalise Theorem 4.2.2, we need a definition.

Definition 4.4.1. Let $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ be a multimodal frame. For any $1 \leq i \leq n$ we say that $\mathfrak{F}$ is $\mathfrak{R}_{\mathbf{N}_i}$-converse well-founded iff $\mathfrak{R}_{\mathbf{N}_i}$ is converse well-founded. We say that $\mathfrak{F}$ is converse well-founded iff $\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i}$ is converse well-founded.

We can now generalise Theorem 4.2.2.

Theorem 4.4.1 (MSCT)**.** *A multimodal frame $\mathfrak{F}$ admits a pw-model on every interpretation iff $\mathfrak{F}$ is converse well-founded.*

We will first prove the right to left direction.

Lemma 4.4.1. *Let $\mathfrak{F}$ be a multimodal frame that is converse well-founded. Then $\mathfrak{F}$ admits a pw-model on every interpretation $\mathfrak{V}$.*

Proof. The proof is a simple multimodal generalisation of Halbach and Leigh 2024, Lemma 7.16. Let $\mathfrak{F}$ be a multimodal frame that is converse well-founded, and let $\mathfrak{V}$ be an interpretation function. For each $1 \leq i \leq n$ we now have to construct an $\mathbf{N}_i$-interpretation function $\mathfrak{B}_{\mathbf{N}_i}$ that satisfies condition (Intended $\mathbf{N}_i$-Interpretation) for each $w \in \mathfrak{W}$ to provide a pw-model. Since $\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i}$ is converse well-founded, we have that $(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{-1}$ is well-founded. Hence, by transfinite recursion we can construct a function $\mathfrak{B} : \mathfrak{W} \to \prod_{1\leq i\leq n} \mathcal{P}(\ulcorner \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})\urcorner)$ which defines the $\mathfrak{B}_{\mathbf{N}_i}$ simultaneously, for $1 \leq i \leq n$.

We do this as follows. For each $w \in \mathfrak{W}$ let

$$\begin{aligned}\mathfrak{B}(w) \coloneqq\ & \langle \ulcorner \{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \mid \text{for each } v \in \mathfrak{R}_{\mathbf{N}_1}(w),\ \mathfrak{M}(v) \vDash \phi\} \urcorner, \\ & \ldots, \\ & \ulcorner \{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \mid \text{for each } v \in \mathfrak{R}_{\mathbf{N}_n}(w),\ \mathfrak{M}(v) \vDash \phi\} \urcorner \rangle\end{aligned}$$

Assume for contradiction that $\mathfrak{B}$ is not well-defined. Consider the set

$$M \coloneqq \{w \in \mathfrak{W} \mid \neg\exists x(\mathfrak{B}(w) = x) \vee \exists y_1, y_2(\mathfrak{B}(w) = y_1 \wedge \mathfrak{B}(w) = y_2 \wedge y_1 \neq y_2)\},$$

i.e. the set of all $w \in \mathfrak{W}$ for which $\mathfrak{B}(w)$ is not well-defined. By assumption $M \neq \emptyset$, and hence by well-foundedness there is some $(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{-1}$-minimal $\mathsf{w} \in M$. For any $w \in \mathfrak{W}$, we have that $(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})(w) = \emptyset$ implies that $\mathfrak{B}(w) = \langle \ulcorner \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})\urcorner, \ldots, \ulcorner \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})\urcorner \rangle$, i.e. $\mathfrak{B}(w)$ exists and is unique. Thus, $(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})(\mathsf{w}) \neq \emptyset$ and for all $v \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})(\mathsf{w})$ we have that $\mathfrak{B}(v)$ exists and is unique. Now observe via the projection mappings that $\mathfrak{B}_{\mathbf{N}_i} = \pi_i \circ \mathfrak{B}$, for $1 \leq i \leq n$. Then, for all $v \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})(\mathsf{w})$ and for all $1 \leq i \leq n$ we have that $\mathfrak{B}_{\mathbf{N}_i}(v)$ exists and is unique. Hence, for all $v \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})(\mathsf{w})$ we have

that $\ulcorner\{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \mid \mathfrak{M}(v) \vDash \phi\}\urcorner$ exists and is unique. In particular, we have that $\mathfrak{B}_{\mathbf{N}_i}(\mathsf{w}) = \ulcorner\{\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}) \mid \text{for each } v \in \mathfrak{R}_{\mathbf{N}_i}(w),\ \mathfrak{M}(v) \vDash \phi\}\urcorner$ exists and is unique, for each $1 \leq i \leq n$. However, by construction we then have that $\mathfrak{B}(\mathsf{w})$ exists and is unique ※.

Hence, $\mathfrak{B}$ is well-defined. In particular, we have obtained a pw-model, since by construction for each $1 \leq i \leq n$ and $w \in \mathfrak{W}$ we have that $\mathfrak{B}_{\mathbf{N}_i}(w) = \pi_i \circ \mathfrak{B}(w)$ satisfies condition (Intended $\mathbf{N}_i$-Interpretation). Q.E.D.

We will now prove the contraposition of the left to right direction.

Lemma 4.4.2. *Let $\mathfrak{F}$ be a multimodal frame that is converse ill-founded. Then $\mathfrak{F}$ does not admit a pw-model on every interpretation.*

The idea of the proof of Lemma 4.4.2 is as follows. Here too we follow the proof-strategy of Halbach and Leigh 2024, Theorem 7.17, with the results in the rest of this chapter (except for Löb's Theorem) being multimodal generalisations of their results in §7.4.

1) We will first define in the object-language $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$ the conjunction of the $\mathbf{N}_i$, for $1 \leq i \leq n$; and then take the transitive closure of this conjunction. We denote this transitive closure by $(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^\star$, and its associated relation on worlds by $(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$. The idea of $(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^\star$ is that it collapses the distinct modal syntactic predicates into one modality, and then takes its transitive closure.

2) We will then show that Löb's Theorem proves that the schema

$$(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^\star\overline{\ulcorner(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner}\rightarrow\phi\urcorner} \rightarrow (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner} \qquad (\mathsf{Löb}_{(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^\star})$$

where $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$ is true in every pw-model.

3) Subsequently, for any multimodal frame $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle$ we will prove that; if $(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$ is converse ill-founded; then there is an interpretation $\mathfrak{V}_{\mathfrak{F}}$ such that $\mathfrak{F}$ cannot be raised to a pw-model with $\mathfrak{V}_{\mathfrak{F}}$ as interpretation that makes $\mathsf{Löb}_{(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^\star}$ true. This is why we've added p to $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$; without it, restricted only to PA, we

would not be able to define the interpretation $\mathfrak{V}_{\mathfrak{F}}$. Thus, by 2) $\mathfrak{F}$ does not admit a pw-model on every interpretation.

4) Finally, we will prove that a multimodal frame $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ is converse well-founded iff $(\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$ is converse well-founded.

5) This concludes the proof; if $\mathfrak{F}$ is converse ill-founded, then by 4) $(\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$ is converse ill-founded; and thus by 3) does not admit a pw-model on every interpretation.

Thus, we first need some definitions:

Definition 4.4.2. The abbreviation $(\bigwedge_{1 \leq i \leq n} \mathbf{N}_i)^n \overline{\ulcorner \phi \urcorner}$ for any $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1, \ldots, \mathbf{N}_n})$ is inductively defined in the meta-language as follows:

$$(\bigwedge_{1 \leq i \leq n} \mathbf{N}_i)^0 \overline{\ulcorner \phi \urcorner} \text{ abbreviates } \bigwedge_{1 \leq i \leq n} \mathbf{N}_i \overline{\ulcorner \phi \urcorner}$$

$$(\bigwedge_{1 \leq i \leq n} \mathbf{N}_i)^{m+1} \overline{\ulcorner \phi \urcorner} \text{ abbreviates } \bigwedge_{1 \leq i \leq n} \mathbf{N}_i \overline{\ulcorner (\bigwedge_{1 \leq i \leq n} \mathbf{N}_i)^m \overline{\ulcorner \phi \urcorner} \urcorner}$$

Note that we cannot quantify over m in the object language in the expression $(\bigwedge_{1 \leq i \leq n} \mathbf{N}_i)^m \overline{\ulcorner \phi \urcorner}$, because this is an abbreviation defined in the meta-language. However, we can emulate this definition in the object language, and thus in a sense quantify over m in the object language, as follows. As in the proof of McGee's Paradox (1985) and Theorem 4.3.6, by an application of the Uniform Diagonal Lemma, let $F^*(x, y, z)$ be a formula such that

$$\begin{aligned} \mathsf{MSPL} \vdash \forall x \forall y \forall z (F^*(x, y, z) \leftrightarrow & \\ & \exists m (x = \mathsf{S}(m) \wedge y = \overline{\ulcorner \forall y (F^*(\overline{m}, y, \overline{z}) \rightarrow \bigwedge_{1 \leq i \leq n} \mathbf{N}_i y) \urcorner}) \\ & \vee (x = \underline{0} \wedge y = z)) \end{aligned} \tag{F^*}$$

Recall that, by abuse of notation, $F^*(x, y, z)$ says that y codes the sentence which 'prefixes x many instances of '$\bigwedge_{1 \leq i \leq n} \mathbf{N}_i$' to z' (along with the proper number of nested quotations). That is, for any $m \in \mathbb{N}$, $F^*(\mathsf{S}(\overline{m}), y, z)$ expresses that y codes the sentence abbreviated by $(\bigwedge_{1 \leq i \leq n} \mathbf{N}_i)^m z$.

Definition 4.4.3. We define the transitive closure of the conjunction $\bigwedge_{1\leq i\leq n}\mathbf{N}_i$ as

$$(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner}\ \text{ abbreviates }\ \forall x\forall y(F^*(x,y,\overline{\ulcorner\phi\urcorner})\rightarrow\bigwedge_{1\leq i\leq n}\mathbf{N}_i y)$$

This prefixes to z all permutations of finite length of the $\mathbf{N}_i$ for $1\leq i\leq n$, which can be made precise as follows.

Lemma 4.4.3. *Let* $\mathfrak{M}=\langle\langle\mathfrak{W},\mathfrak{R}_{\mathbf{N}_1},\ldots,\mathfrak{R}_{\mathbf{N}_n}\rangle,\mathbb{N},\mathfrak{V},\mathfrak{B}_{\mathbf{N}_1},\ldots,\mathfrak{B}_{\mathbf{N}_n}\rangle$ *be a pw-model;* $w\in\mathfrak{W}$*;* $\phi\in\mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$*; and* $(\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star$ *be the transitive closure of* $\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i}$*. Then* $\mathfrak{M}(w)\vDash(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner}$ *iff for all* $v\in(\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star(w)$ *we have that* $\mathfrak{M}(v)\vDash\phi$*.*

Proof.

$$\mathfrak{M}(w)\vDash(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner}\qquad\text{iff}\tag{4.115}$$

$$\mathfrak{M}(w)\vDash\forall x\forall y(F^*(x,y,\overline{\ulcorner\phi\urcorner})\rightarrow\bigwedge_{1\leq i\leq n}\mathbf{N}_i y)\qquad\text{iff}\tag{4.116}$$

$$\forall m\in\mathbb{N},\ \mathfrak{M}(w)\vDash\forall y(F^*(\overline{m},y,\overline{\ulcorner\phi\urcorner})\rightarrow\bigwedge_{1\leq i\leq n}\mathbf{N}_i y)\qquad\text{iff}\tag{4.117}$$

Line (4.116) follows by definition, and line (4.117) follows because all pw-models are built up from $\mathbb{N}$. We can now prove by induction that line (4.117) holds iff

$$\forall m\in\mathbb{N},\ \mathfrak{M}(w)\vDash(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^m\overline{\ulcorner\phi\urcorner}\tag{4.118}$$

For the base case that m is 0, given that one of the axioms of PA is that $\neg\exists x\,\underline{0}=\mathsf{S}(x)$, from (F^*) we have that

$$\mathfrak{M}(w)\vDash\forall y F^*(\underline{0},y,\overline{\ulcorner\phi\urcorner})\leftrightarrow y=\overline{\ulcorner\phi\urcorner}\tag{4.119}$$

This implies that

$$\mathfrak{M}(w)\vDash\forall y(F^*(\underline{0},y,\overline{\ulcorner\phi\urcorner})\rightarrow\bigwedge_{1\leq i\leq n}\mathbf{N}_i y)\qquad\text{iff}\tag{4.120}$$

$$\mathfrak{M}(w)\vDash\forall y(y=\overline{\ulcorner\phi\urcorner}\rightarrow\bigwedge_{1\leq i\leq n}\mathbf{N}_i y)\qquad\text{iff}\tag{4.121}$$

$$\mathfrak{M}(w)\vDash\bigwedge_{1\leq i\leq n}\mathbf{N}_i\overline{\ulcorner\phi\urcorner}\tag{4.122}$$

This proves the base case. For the inductive step, assume that for each $v \in \mathfrak{W}$

$$\mathfrak{M}(v) \vDash \forall y(F^*(\overline{k}, y, \overline{\ulcorner \phi \urcorner}) \rightarrow \bigwedge_{1 \leq i \leq n} \mathbf{N}_i y) \qquad \text{iff} \tag{4.123}$$

$$\mathfrak{M}(v) \vDash (\bigwedge_{1 \leq i \leq n} \mathbf{N}_i)^k \overline{\ulcorner \phi \urcorner} \tag{4.124}$$

Now consider $F^*(\mathsf{S}(\overline{k}), y, \overline{\ulcorner \phi \urcorner})$. Again given that one of the axioms of PA is that $\neg \exists x\, \underline{0} = \mathsf{S}(x)$, from (F^*) we have that

$$F^*(\mathsf{S}(\overline{k}), y, \overline{\ulcorner \phi \urcorner}) \leftrightarrow \exists m(\mathsf{S}(\overline{k}) = \mathsf{S}(m) \wedge y = \overline{\ulcorner \forall y(F^*(\overline{m}, y, \overline{\ulcorner \phi \urcorner}) \rightarrow \bigwedge_{1 \leq i \leq n} \mathbf{N}_i y) \urcorner}) \tag{4.125}$$

By the PA axiom $\forall x \forall y(\mathsf{S}(x) = \mathsf{S}(y) \rightarrow x = y)$, from line (4.125) we have that

$$F^*(\mathsf{S}(\overline{k}), y, \overline{\ulcorner \phi \urcorner}) \leftrightarrow y = \overline{\ulcorner \forall y(F^*(\overline{k}, y, \overline{\ulcorner \phi \urcorner}) \rightarrow \bigwedge_{1 \leq i \leq n} \mathbf{N}_i y) \urcorner} \tag{4.126}$$

This implies that

$$\mathfrak{M}(w) \vDash \forall y(F^*(\mathsf{S}(\overline{k}), y, \overline{\ulcorner \phi \urcorner}) \rightarrow \bigwedge_{1 \leq i \leq n} \mathbf{N}_i y) \qquad \text{iff} \tag{4.127}$$

$$\mathfrak{M}(w) \vDash \forall y(y = \overline{\ulcorner \forall y(F^*(\overline{k}, y, \overline{\ulcorner \phi \urcorner}) \rightarrow \bigwedge_{1 \leq i \leq n} \mathbf{N}_i y) \urcorner} \rightarrow \bigwedge_{1 \leq i \leq n} \mathbf{N}_i y) \qquad \text{iff} \tag{4.128}$$

$$\mathfrak{M}(w) \vDash \bigwedge_{1 \leq i \leq n} \mathbf{N}_i \overline{\ulcorner \forall y(F^*(\overline{k}, y, \overline{\ulcorner \phi \urcorner}) \rightarrow \bigwedge_{1 \leq i \leq n} \mathbf{N}_i y) \urcorner} \tag{4.129}$$

Now by $\mathbf{N}_i$ Truth Condition, line (4.129) holds iff for each $1 \leq i \leq n$; if $v \in \mathfrak{R}_{\mathbf{N}_i}(w)$, then

$$\mathfrak{M}(v) \vDash \forall y(F^*(\overline{k}, y, \overline{\ulcorner \phi \urcorner}) \rightarrow \bigwedge_{1 \leq i \leq n} \mathbf{N}_i y) \tag{4.130}$$

By the inductive hypothesis, line (4.130) holds iff for each $1 \leq i \leq n$; if $v \in \mathfrak{R}_{\mathbf{N}_i}(w)$, then

$$\mathfrak{M}(v) \vDash (\bigwedge_{1 \leq i \leq n} \mathbf{N}_i)^k \overline{\ulcorner \phi \urcorner} \tag{4.131}$$

By $\mathbf{N}_i$ Truth Condition, line (4.131) holds iff

$$\mathfrak{M}(w) \vDash \bigwedge_{1 \leq i \leq n} \mathbf{N}_i \overline{\ulcorner (\bigwedge_{1 \leq i \leq n} \mathbf{N}_i)^k \overline{\ulcorner \phi \urcorner} \urcorner} \tag{4.132}$$

This proves the inductive step, and hence the equivalence of lines (4.117) and (4.118).

Given the binary relation $\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i}$, recall that

$$(\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})(w) = \{v \in \mathfrak{W} \mid \langle w, v \rangle \in \bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i}\} \tag{4.133}$$

Now for any $m \in \mathbb{N}$ let

$$(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^m := \{\langle w, v\rangle \mid \text{there are } v_0, \ldots, v_{m+1} \in W \text{ with } v_0 = w, v_{m+1} = v, \text{ and } v_{k+1} \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})(v_k) \text{ for each } 0 \leq k < m\} \tag{4.134}$$

and let

$$(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^m(w) := \{v \in \mathfrak{W} \mid \langle w, v\rangle \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^m\} \tag{4.135}$$

By induction we can prove that

$$\forall m \in \mathbb{N},\ \mathfrak{M}(w) \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^m \overline{\ulcorner \phi \urcorner} \tag{4.136}$$

holds iff for all $m \in \mathbb{N}$ and for all $v \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^m(w)$, we have that $\mathfrak{M}(v) \vDash \phi$. For the base case that m is 0, we have that

$$\begin{aligned}(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^0 &= \{\langle w, v\rangle \mid \text{ there are } v_0, v_1 \in W \text{ with } v_0 = w, v_1 = v \text{ and } v_1 \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})(v_0)\} \\ &= \bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i}\end{aligned} \tag{4.137}$$

Since $(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^0 \overline{\ulcorner \phi \urcorner}$ abbreviates $\bigwedge_{1\leq i\leq n} \mathbf{N}_i \overline{\ulcorner \phi \urcorner}$, the claim follows from $\mathbf{N}_i$ Truth Condition. This proves the base case. The inductive step is proved in a similar way. Assume that for all $v \in W$

$$\mathfrak{M}(v) \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^k \overline{\ulcorner \phi \urcorner} \tag{4.138}$$

holds iff for all $u \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^k(v)$, we have that $\mathfrak{M}(u) \vDash \phi$. We then have that

$$\mathfrak{M}(w) \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{k+1} \overline{\ulcorner \phi \urcorner} \qquad \text{iff} \tag{4.139}$$

$$\mathfrak{M}(w) \vDash \bigwedge_{1\leq i\leq n} \mathbf{N}_i \overline{\ulcorner (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^k \overline{\ulcorner \phi \urcorner} \urcorner} \tag{4.140}$$

iff for all $v \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})(w)$, we have that

$$\mathfrak{M}(v) \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^k \overline{\ulcorner \phi \urcorner} \tag{4.141}$$

which by induction hypothesis holds iff for all $v \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})(w)$ and for all $u_v \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^k(v)$, we have that $\mathfrak{M}(u_v) \vDash \phi$. Observing that

$$(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{k+1}(w) = \bigcup_{v\in(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})(w)} (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^k(v) \tag{4.142}$$

we get that line (4.139) holds iff for all $v' \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{k+1}(w)$ we have that $\mathfrak{M}(v') \vDash \psi$. This proves the inductive step.

Further, observe the following equalities

$$(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{\star} = \bigcup_{m\in\mathbb{N}} (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{m} \tag{4.143}$$

and

$$(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{\star}(w) = \bigcup_{m\in\mathbb{N}} ((\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{m}(w)) \tag{4.144}$$

Thus

$$\forall m \in \mathbb{N},\ \mathfrak{M}(w) \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{m}\overline{\ulcorner\phi\urcorner} \tag{4.145}$$

holds; iff for all $m \in \mathbb{N}$ and for all $v \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{m}(w)$, we have that $\mathfrak{M}(v) \vDash \phi$; iff by line (4.144) for all $v \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{\star}(w)$, we have that $\mathfrak{M}(v) \vDash \phi$. This proves the lemma. Q.E.D.

We now have the following results regarding the behaviour of $(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}$.

Lemma 4.4.4. *Let $\mathfrak{M}$ be a pw-model; and $\phi \in \mathsf{Form}(\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n})$ $\ni \psi$. Then we have that:*

$$\mathfrak{M} \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\phi \to \psi\urcorner} \to ((\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\phi\urcorner} \to (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\psi\urcorner}) \qquad (\mathsf{K}_{(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}})$$

$$\mathfrak{M} \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\phi\urcorner} \to (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\phi\urcorner}\urcorner} \qquad (4_{(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}})$$

$$\mathfrak{M} \vDash \phi \Rightarrow \mathfrak{M} \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\phi\urcorner} \qquad (\mathsf{NEC}_{(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}})$$

Proof. We have that the schemata $\mathsf{K}_{(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}}$ and $4_{(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}}$, and the rule $\mathsf{NEC}_{(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}}$ follow from Lemma 4.4.3. The proof is similar to proofs we have already seen:

- For $\mathsf{K}_{(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}}$, assume for contradiction that

$$\mathfrak{M}(w) \nvDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\phi \to \psi\urcorner} \to ((\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\phi\urcorner} \to (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\psi\urcorner}) \tag{4.146}$$

  We then have that (1) $\mathfrak{M}(w) \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\phi \to \psi\urcorner}$; (2) $\mathfrak{M}(w) \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\phi\urcorner}$; and (3) $\mathfrak{M}(w) \nvDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner\psi\urcorner}$. By (3) and Lemma 4.4.3, there is some $v \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{\star}(w)$ such that (4) $\mathfrak{M}(v) \nvDash \psi$. By (2) and Lemma 4.4.3, we have that (5)

$\mathfrak{M}(v) \vDash \phi$. By (1) and Lemma 4.4.3, we have that (6) $\mathfrak{M}(v) \vDash \phi \rightarrow \psi$. By (5) and 6, we have that $\mathfrak{M}(v) \vDash \psi$, contradicting (4).

- For $\mathsf{4}_{(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star}$, assume for contradiction that

$$\mathfrak{M}(w) \nvDash (\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner} \rightarrow (\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner}\urcorner} \tag{4.147}$$

We then have that (1) $\mathfrak{M}(w) \vDash (\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner}$; and (2) $\mathfrak{M}(w) \nvDash (\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner}\urcorner}$. By (2) and Lemma 4.4.3, there is some $v \in (\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star(w)$ such that (3) $\mathfrak{M}(v) \nvDash (\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner}$. By (3) and Lemma 4.4.3, there is some $u \in (\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star(v)$ such that (4) $\mathfrak{M}(u) \nvDash \phi$. Since $(\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star$ is transitive, we have that (5) $u \in (\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star(w)$. By (1), (5), and Lemma 4.4.3, we have that $\mathfrak{M}(u) \vDash \phi$, contradicting (4).

- For $\mathsf{NEC}_{(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star}$, assume that $\mathfrak{M}(v) \vDash \phi$ for every $v \in \mathfrak{W}$. Since $(\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star(w) \subseteq \mathfrak{W}$, by Lemma 4.4.3 we have that $\mathfrak{M}(w) \vDash (\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner\phi\urcorner}$.

Q.E.D.

For the next result regarding the behaviour of $(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star$, we need the following theorem.

Theorem 4.4.2 (Löb's Theorem)**.** *Let $\mathcal{L}^X_{\mathsf{PA}}$ be any language that expands $\mathcal{L}_{\mathsf{PA}}$, and let X be a (complex) predicate of $\mathcal{L}^X_{\mathsf{PA}}$.*[13] *Furthermore, let PA_X denote the theory of the axioms of PA formulated in $\mathcal{L}^X_{\mathsf{PA}}$. Let E denote the theory of PA_X and the axiom schemata*

$$X\overline{\ulcorner\phi \rightarrow \psi\urcorner} \rightarrow (X\overline{\ulcorner\phi\urcorner} \rightarrow X\overline{\ulcorner\psi\urcorner}) \tag{K_X}$$

$$X\overline{\ulcorner\phi\urcorner} \rightarrow X\overline{\ulcorner X\overline{\ulcorner\phi\urcorner}\urcorner} \tag{$\mathsf{4}_X$}$$

where $\phi \in \mathsf{Form}(\mathcal{L}^X_{\mathsf{PA}})$; and which is closed under the rule

$$\mathsf{E} \vdash \phi \Rightarrow \mathsf{E} \vdash X\overline{\ulcorner\phi\urcorner} \tag{NEC_X}$$

[13] This theorem is by Löb 1955. In his original formulation of the theorem he proves this result for a provability predicate Pr (or Bew). Here I have generalised his result to any (complex) predicate X satisfying the required conditions.

Then E *contains the schema*

$$X^{\ulcorner}\overline{X^{\ulcorner}\overline{\phi}^{\urcorner}\rightarrow\phi}^{\urcorner}\rightarrow X^{\ulcorner}\overline{\phi}^{\urcorner} \tag{$\mathsf{L\ddot{o}b}_X$}$$

where $\phi\in\mathsf{Form}(\mathcal{L}^X_{\mathsf{PA}})$.

Proof. Let $\phi\in\mathsf{Form}(\mathcal{L}^X_{\mathsf{PA}})$. By the Diagonal Lemma there is a λ such that $\mathsf{E}\vdash\lambda\leftrightarrow(X^{\ulcorner}\overline{\lambda}^{\urcorner}\rightarrow\phi)$. We then have the following:

$\mathsf{E}\vdash\lambda\leftrightarrow(X^{\ulcorner}\overline{\lambda}^{\urcorner}\rightarrow\phi)$	DL	(4.148)
$\mathsf{E}\vdash\lambda\rightarrow(X^{\ulcorner}\overline{\lambda}^{\urcorner}\rightarrow\phi)$	(4.148)	(4.149)
$\mathsf{E}\vdash\overline{X^{\ulcorner}\lambda\rightarrow(X^{\ulcorner}\overline{\lambda}^{\urcorner}\rightarrow\phi)^{\urcorner}}$	NEC_X	(4.150)
$\mathsf{E}\vdash X^{\ulcorner}\overline{\lambda}^{\urcorner}\rightarrow(\overline{X^{\ulcorner}X^{\ulcorner}\overline{\lambda}^{\urcorner\urcorner}}\rightarrow X^{\ulcorner}\overline{\phi}^{\urcorner})$	K_X	(4.151)
$\mathsf{E}\vdash X^{\ulcorner}\overline{\lambda}^{\urcorner}\rightarrow\overline{X^{\ulcorner}X^{\ulcorner}\overline{\lambda}^{\urcorner\urcorner}}$	$\mathsf{4}_X$	(4.152)
$\mathsf{E}\vdash X^{\ulcorner}\overline{\lambda}^{\urcorner}\rightarrow X^{\ulcorner}\overline{\phi}^{\urcorner}$	(4.152), (4.151)	(4.153)
$\mathsf{E}\vdash(X^{\ulcorner}\overline{\phi}^{\urcorner}\rightarrow\phi)\rightarrow(X^{\ulcorner}\overline{\lambda}^{\urcorner}\rightarrow\phi)$	(4.153)	(4.154)
$\mathsf{E}\vdash(X^{\ulcorner}\overline{\phi}^{\urcorner}\rightarrow\phi)\rightarrow\lambda$	(4.148), (4.154)	(4.155)
$\mathsf{E}\vdash\overline{X^{\ulcorner}(X^{\ulcorner}\overline{\phi}^{\urcorner}\rightarrow\phi)^{\urcorner}}\rightarrow X^{\ulcorner}\overline{\lambda}^{\urcorner}$	NEC_X, K_X	(4.156)
$\mathsf{E}\vdash\overline{X^{\ulcorner}(X^{\ulcorner}\overline{\phi}^{\urcorner}\rightarrow\phi)^{\urcorner}}\rightarrow X^{\ulcorner}\overline{\phi}^{\urcorner}$	(4.153), (4.156)	(4.157)

Q.E.D.

Corollary 4.4.1. *Let* $\mathfrak{M}$ *be a pw-model; and* $\phi\in\mathsf{Form}(\mathcal{L}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}_{\mathsf{PA}})\ni\psi$. *Then we have that:*

$$\mathfrak{M}\vDash(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star\ulcorner}\overline{(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star\ulcorner}\overline{\phi}^{\urcorner}\rightarrow\phi}^{\urcorner}\rightarrow(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star\ulcorner}\overline{\phi}^{\urcorner} \tag{$\mathsf{L\ddot{o}b}_{(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star}}$}$$

Proof. Since by Lemma 4.4.4 the schemata $\mathsf{K}_{(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star}}$ and $\mathsf{4}_{(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star}}$ are true in $\mathfrak{M}$, and the fact that the rule $\mathsf{NEC}_{(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star}}$ holds for $\mathfrak{M}$; by Löb's Theorem, taking $(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star}$ for X, and soundness we have that

$$\mathfrak{M}\vDash(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star\ulcorner}\overline{(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star\ulcorner}\overline{\phi}^{\urcorner}\rightarrow\phi}^{\urcorner}\rightarrow(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^{\star\ulcorner}\overline{\phi}^{\urcorner} \tag{4.158}$$

Q.E.D.

Recall the idea of the proof of Lemma 4.4.2, which is the left to right direction of Theorem 4.4.1. So far we have detailed the first two steps; defining the transitive closure of the conjunction of the $\mathbf{N}_i$, for $1 \leq i \leq n$; and proving that the schema $\mathsf{Löb}_{(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star}$ is true in any pw-model. We will now detail the third step:

Lemma 4.4.5. *Let $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle$ be a multimodal frame. If $(\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star$ is converse ill-founded, then there is an interpretation $\mathfrak{V}_\mathfrak{F}$ such that $\mathfrak{F}$ cannot be raised to a pw-model with $\mathfrak{V}_\mathfrak{F}$ as interpretation that makes $\mathsf{Löb}_{(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star}$ true.*

Proof. The proof is a possible-worlds P-semantics version of part of the proof of Boolos 1993, Theorem 10. Let $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n}\rangle$ be a multimodal frame such that $(\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star$ is converse ill-founded. We will pick some $v \in \mathfrak{W}$ and construct a valuation $\mathfrak{V}_\mathfrak{F}$ such that; assuming there is a pw-model $\mathfrak{M}_{\mathfrak{V}_\mathfrak{F}} = \langle \mathfrak{F}, \mathbb{N}, \mathfrak{V}_\mathfrak{F}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n}\rangle$; then

$$\mathfrak{M}_{\mathfrak{V}_\mathfrak{F}}(v) \not\Vdash (\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner(\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner p\urcorner}\rightarrow p\urcorner} \rightarrow (\bigwedge_{1\leq i\leq n}\mathbf{N}_i)^\star\overline{\ulcorner p\urcorner}$$

Recall that p is the zero-place contingent predicate in the language $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$.

Since $(\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star$ is converse ill-founded, (*) there is some $\emptyset \neq W_\mathfrak{F} \subseteq \mathfrak{W}$ such that for any $w \in W_\mathfrak{F}$, there is another $w' \in W_\mathfrak{F}$ with $\langle w, w'\rangle \in (\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star$. We will now define $\mathfrak{V}_\mathfrak{F}$ as follows:

$$\mathfrak{V}_\mathfrak{F}(w)(p) := \begin{cases} 1 & \text{for} \quad w \notin W_\mathfrak{F} \\ 0 & \text{for} \quad w \in W_\mathfrak{F} \end{cases} \tag{4.159}$$

In effect, $\mathfrak{V}_\mathfrak{F}$ allows us to pick out the subset $W_\mathfrak{F}$.[14]

Now, since $\emptyset \neq W_\mathfrak{F}$, pick some $\mathsf{v} \in W_\mathfrak{F}$. Furthermore, assume that there are $\mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n}$ such that there is a pw-model $\mathfrak{M}_{\mathfrak{V}_\mathfrak{F}} = \langle \mathfrak{F}, \mathbb{N}, \mathfrak{V}_\mathfrak{F}, \mathfrak{B}_{\mathbf{N}_1}, \ldots, \mathfrak{B}_{\mathbf{N}_n}\rangle$. Let $u \in (\bigcup_{1\leq i\leq n}\mathfrak{R}_{\mathbf{N}_i})^\star(\mathsf{v})$. Then we have the following:

[14] Without p in $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$, if we are restricted to the definability of PA, then we would not be able to define a valuation that picks out arbitrary $W_\mathfrak{F}$. See the remarks on pages 87-88 regarding results by Halbach, Leitgeb, et al. 2003. In particular, recall that without contingent vocabulary Halbach, Leitgeb, et al. 2003 show that any transitive unimodal frame that has converse ill-founded worlds w, all of whose depth is such that $\alpha_w \geq \kappa$, admits a pw-model; however, such a frame is converse ill-founded, and thus by Theorem 4.2.2 does not admit a pw-model on every interpretation.

A. If $u \in W_{\mathfrak{F}}$, by (*) there is some $u' \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{\star}(u) \cap W_{\mathfrak{F}}$. Since $\mathfrak{V}_{\mathfrak{F}}(u')(p) = 0$, we have that $\mathfrak{M}_{\mathfrak{V}_{\mathfrak{F}}}(u') \nvDash p$. Thus, we have that $\mathfrak{M}_{\mathfrak{V}_{\mathfrak{F}}}(u) \nvDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner p \urcorner}$, and thus also that $\mathfrak{M}_{\mathfrak{V}_{\mathfrak{F}}}(u) \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner p \urcorner} \rightarrow p$.

B. If $u \notin W_{\mathfrak{F}}$, we have that $\mathfrak{M}_{\mathfrak{V}_{\mathfrak{F}}}(u) \vDash p$, and thus also that $\mathfrak{M}_{\mathfrak{V}_{\mathfrak{F}}}(u) \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner p \urcorner} \rightarrow p$.

Since $u \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{\star}(\mathsf{v})$ was arbitrary, by Lemma 4.4.3 we have that

$$\mathfrak{M}_{\mathfrak{V}_{\mathfrak{F}}}(\mathsf{v}) \vDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner p \urcorner} \rightarrow p \urcorner} \tag{4.160}$$

Furthermore, since $\mathsf{v} \in W_{\mathfrak{F}}$, by (*) there is some $v' \in (\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{\star}(\mathsf{v}) \cap W_{\mathfrak{F}}$. As before, since $\mathfrak{V}_{\mathfrak{F}}(v')(p) = 0$, we have that

$$\mathfrak{M}_{\mathfrak{V}_{\mathfrak{F}}}(\mathsf{v}) \nvDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner p \urcorner} \tag{4.161}$$

Taking (4.160) and (4.161) together, we have that

$$\mathfrak{M}_{\mathfrak{V}_{\mathfrak{F}}}(\mathsf{v}) \nvDash (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner p \urcorner} \rightarrow p \urcorner} \rightarrow (\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^{\star}\overline{\ulcorner p \urcorner} \tag{4.162}$$

Q.E.D.

Corollary 4.4.2. *Let $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ be a multimodal frame. If $(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{\star}$ is converse ill-founded, then there is an interpretation $\mathfrak{V}_{\mathfrak{F}}$ such that there is no pw-model with $\mathfrak{F}$ as frame, and $\mathfrak{V}_{\mathfrak{F}}$ as interpretation. That is, $\mathfrak{F}$ does not admit a pw-model on every interpretation.*

Proof. This follows from Lemmas 4.4.4 and 4.4.5. Q.E.D.

Recalling the idea of the proof of Lemma 4.4.2, we must now detail the final, fourth step:

Lemma 4.4.6. *Let $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ be a multimodal frame. Then $\mathfrak{F}$ is converse well-founded iff $(\bigcup_{1\leq i\leq n} \mathfrak{R}_{\mathbf{N}_i})^{\star}$ is converse well-founded.*

Proof. This is a well-known piece of folklore regarding relations and their transitive closures. Let $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ be a multimodal frame. Assume that $\mathfrak{F}$ is converse well-founded, i.e. that $\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i}$ is converse well-founded. Let $W \subseteq \mathfrak{W}$. Now consider the set

$$\begin{aligned} W' := \{w \in \mathfrak{W} \mid \text{there are } v_1, v_2 \in W \text{ with } \langle v_1, w \rangle \in (\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^m \\ \text{and } \langle w, v_2 \rangle \in (\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^n \text{ for some } m, n \in \mathbb{N}\} \end{aligned} \tag{4.163}$$

That is, W' is obtained by closing W under all finite paths in $\mathfrak{W}$ between elements of W. Now since $W' \subseteq \mathfrak{W}$, and $\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i}$ is converse well-founded, (**) there must be some $\mathsf{w} \in W'$ such that there is no $u \in W'$ with $\langle \mathsf{w}, u \rangle \in (\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})$. In particular, by construction of W' we have that $\mathsf{w} \in W$. Now assume that there is some $\mathsf{u} \in W$ with $\langle \mathsf{w}, \mathsf{u} \rangle \in (\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$. Then by construction of $(\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$, there is some $k \in \mathbb{N}$ with $\langle \mathsf{w}, \mathsf{u} \rangle \in (\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^k$. In particular, by construction of W', there is some $\mathsf{w}' \in W'$ with $\langle \mathsf{w}, \mathsf{w}' \rangle \in (\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})$. This contradicts (**). Hence, there is no $u \in W$ with $\langle \mathsf{w}, u \rangle \in (\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$. Since W is arbitrary, this implies that $(\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$ is converse well-founded.

For the other direction, assume that $(\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$ is converse well-founded. Let $W \subseteq \mathfrak{W}$. Then, there is some $\mathsf{w} \in W$ such that there is no $u \in W$ with $\langle \mathsf{w}, u \rangle \in (\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$. In particular, there is no $u \in W$ with $\langle \mathsf{w}, u \rangle \in \bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i}$. Since W is arbitrary, this implies that $\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i}$ is converse well-founded. Q.E.D.

We can now finally prove Lemma 4.4.2:

Proof. Let a multimodal frame $\mathfrak{F} = \langle \mathfrak{W}, \mathfrak{R}_{\mathbf{N}_1}, \ldots, \mathfrak{R}_{\mathbf{N}_n} \rangle$ be converse ill-founded. Then by Lemma 4.4.6 we have that $(\bigcup_{1 \leq i \leq n} \mathfrak{R}_{\mathbf{N}_i})^\star$ is converse ill-founded. Finally, by corollary 4.4.2 $\mathfrak{F}$ does not admit a pw-model on every interpretation. Q.E.D.

Putting the Lemmas 4.4.1 and 4.4.2 together provides the proof of Theorem 4.4.1, which states that a multimodal frame $\mathfrak{F}$ admits a pw-model on every interpretation iff $\mathfrak{F}$ is converse well-founded.

Before concluding this chapter, let us briefly summarise the results we proved in sections 4.3 and 4.4. Starting with section 4.3, the first major result we proved was Theorem 4.3.3, regarding loops in a multimodal frame. This theorem states that, if there is a fixed natural number k such that each world in a given multimodal frame admits some of loop of length k or less (but not none); then said frame does not admit a pw-model. The second major result we proved was Theorem 4.3.6, regarding infinite ascending chains. This theorem states that, if a given multimodal frame has some number of accessibility relations such that every world sees another world with one of these accessibility relations; then said frame does not admit a pw-model. The third major result we proved was Theorem 4.3.7, ensuring that if a multimodal frame has a full subframe that does not admit a pw-model, then neither does said initial frame; thereby widening the scope of the previous two major results. The final group of results we proved were Theorems 4.3.8-4.3.10, which showed us that in the two-dimensional possible-world P-semantics no syntactic predicate can have the same truth-conditions as the sentential operators @, $\times$, or F have in the two-dimensional possible-world O-semantics.

Moving on to section 4.4, the first major result we proved was Lemma 4.4.1, which states that a multimodal frame $\mathfrak{F}$ that is converse well-founded admits a pw-model on every interpretation $\mathfrak{V}$. The idea behind the proof of this theorem was to simply use transfinite recursion to define the $\mathbf{N}_i$-interpretation functions $\mathfrak{B}_{\mathbf{N}_i}$. The second, and final major result we proved was Lemma 4.4.2, which states that a multimodal frame $\mathfrak{F}$ that is converse ill-founded does not admit a pw-model on every interpretation. The idea behind the proof of this theorem was a bit more complex. First, we defined in the object-language $\mathcal{L}_{\mathsf{PA}}^{\mathbf{N}_1,\ldots,\mathbf{N}_n}$ the conjunction of the $\mathbf{N}_i$, for $1 \leq i \leq n$; and then took the transitive closure of this conjunction. We denoted this $(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^\star$, which collapses the distinct modal syntactic predicates into one modality, and then takes its transitive closure. We then showed that $(\bigwedge_{1\leq i\leq n} \mathbf{N}_i)^\star$ satisfies Löb's Theorem. Finally, we showed that if a frame is converse ill-founded, then we can construct an interpretation that contradicts Löb's Theorem.

To sum this chapter up, we provided a characterisation theorem for the paradoxes of the mention predicate approach. We first considered five more unary and merging

paradoxes besides Montague's Paradox that involve conditions other than $\mathsf{T}_\mathbf{N}$ or $\mathsf{NEC}_\mathbf{N}$. The paradoxes that the mention predicate approach is susceptible to then suggested that a more unified strategy is necessary for avoiding paradox. Furthermore, they made clear that a necessary condition for providing such a strategy is to first categorise when paradox arises, and when it does not. To explore this question of categorisation, we introduced a possible-world predicate-semantics for the mention predicate approach. In particular, we generalised the classical possible-world P-semantics developed by Halbach and Leigh 2024 to multiple distinct unary modal syntactic predicates. We then considered some examples of merging paradoxes in this generalised classical possible-world P-semantics, proving that certain classes of multimodal frames admit no pw-models. We also showed that when expanding the classical possible-world P-semantics to a classical two-dimensional possible-world P-semantics, no syntactic predicate can have the same truth-conditions as the sentential operators @, $\times$, or F have in the two-dimensional possible-world O-semantics. Finally, we proved that a multimodal frame $\mathfrak{F}$ admits a pw-model on every interpretation iff $\mathfrak{F}$ is converse well-founded. This result functions as a categorisation theorem indicating when paradox arises in this generalised classical possible-world P-semantics.

5

Applying The Mention Predicate Approach To Modal Potentialism

Contents

So far in this thesis our main focus has been on the semantics of the mention predicate approach. We considered a host of proof-theoretic paradoxes that the mention predicate approach is prone to. We also considered the model-theoretic semantics of Stern 2014c; Stern 2015 that blocks diagonalisation, and subsequently criticised it for not providing a viable conception of the mention predicate approach. Instead, we constructed a model-theoretic semantics that embraces diagonalisation, and considered a host of model-theoretic paradoxes of this semantics.

Furthermore, recall the methodological approach of sharpening philosophical notions by formalising them (in logic). Such a sharpening of philosophical notions requires us to 'successfully' / 'most appropriately' formalise these notions; which has as a necessary condition that the philosophical and formal frameworks agree in their relevant structural features. In the case of the mention predicate, use predicate, and operator approaches, the question is whether it is closest to the intended interpretation of the modality being formalised that it be formalised as using, or mentioning the complements of modal statements it occurs in. This would provide us with a methodological reason for not abandoning the mention predicate approach as a whole as a response to paradox, *provided* that there are modalities whose intended interpretations are best formalised by the mention predicate approach.

In this chapter, we will precisely provide such an example of a modality in the philosophical literature that is most appropriately formalised using the mention predicate approach; the modality employed by Øystein Linnebo in his account of modal potentialism in his book *Thin Objects* (2018). Modal potentialism has recently received a lot of attention in the literature on the philosophy of mathematics and the philosophy of logic. In this literature, potentialism is the view that mathematical objects are generated successively, and that certain generative processes are incompletable. Thus, certain totalities of mathematical objects are indefinitely extensible; for any such totality, new objects can be generated. Modal potentialism explicates potentialism using quantified modal logic, taking a modal object language as primitive for mathematics. Many modal potentialists deny that the intended interpretation of the primitive modality is absolute, metaphysical, or temporal.[1][2] Modal potentialism has many purported benefits, including giving a natural

[1]Most modal potentialists use temporal language when defending their view. Those that deny that the intended interpretation of the primitive modality is absolute, metaphysical, or temporal, will often clarify that their temporal language is merely a metaphor: 'Safe in the knowledge that it's dispensable, however, it's sometimes helpful to continue to talk in terms of a temporal ordering of stages as a vivid way of describing the usual ordering of ordinals: speaking loosely, we use 'stage', 'earlier', and so on, as suggestive alternatives to 'ordinal', 'less than', and the like. The tensed utterance 'set a will be formed at stage α' is then simply a colourful way of communicating the tenseless claim that a is a member of $V_\alpha(U)$. To guard against potential misunderstanding, we may simply disavow the literal, face-value interpretation' (Studd 2019, p. 50).

[2]Not all modal potentialists deny this. See for instance Scambler 2021, whose intended interpretation of the primitive modality is that of absolute modality.

philosophical solution to the set-theoretic paradoxes and motivation for the axioms of ZF set theory. Proponents of modal potentialism include Berry 2024, Fine 2005, Hellman 1993, Linnebo 2018, Scambler 2021, and Studd 2019.

In this chapter we will critically assess the account of modal potentialism developed by Linnebo 2018, in particular focussing on the primitive modality this account employs. We will do this as follows. In section 5.1, we will carefully expound Linnebo's framework in (2018), in particular the core notion of Fregean abstraction. In section 5.2, we will show that it is closest to Linnebo's intended interpretation of the primitive modality that it be formalised as mentioning, as opposed to using, the complements of modal statements it occurs in. Furthermore, we will show that this intended interpretation commits Linnebo to the legitimacy of introducing a primitive modal syntactic predicate to his object language, besides the primitive modal sentential operator he introduces. However, as we have seen, such a modal syntactic predicate is prone to paradox. In particular, in section 5.3 we will show that principles for this modal syntactic predicate that are fundamental to modal potentialism are inconsistent; it cannot behave anything like the modal sentential operator Linnebo introduces. In section 5.4 we will conclude that in the account of Linnebo 2018 the intended interpretation of the primitive modality and the formal framework do not match up; no modal syntactic predicate can underpin an account of modal potentialism. That is, Linnebo provides an account of modal potentialism that is incoherent, because he is committed to a philosophically motivated extension of his formal framework which is inconsistent. We will put forward some follow-up options available to Linnebo, requiring that he further develops or changes his account in (2018). Moreover, other approaches to modal potentialism are susceptible to a similar challenge; we will end this section by tentatively claiming that James Studd's approach in his book *Everything, more or less: A defence of generality relativism* (2019) is such an approach. Further work is needed to show whether other approaches to modal potentialism are committed to the legitimacy of introducing a primitive modal syntactic predicate to their object language, thus susceptible to the challenges set in this chapter; and what the ramifications are. At the very least, modal potentialists must have a response to these challenges.

More generally, we will see that this example provides yet another warning of the risks of paradox, in particular what goes wrong if one does not use the most appropriate formalisation when formalising philosophical notions.

5.1 Framework

We will now carefully expound Linnebo's framework in (2018). Particular attention will be paid to the intended interpretation of the primitive modality. The aim of this careful exposition is to straightforwardly read off in section 5.2 that it is closest to Linnebo's intended interpretation of the primitive modality that it be formalised as mentioning, as opposed to using, the complements of modal statements it occurs in. Furthermore, the aim is to show that Linnebo did not merely make some superficial mistake in (2018) that is easily corrected; rather, his approach in (2018) is a very reasonable attempt at a platonist account of modal potentialism based on the core notion of Fregean abstraction.

In his book *Thin Objects* (2018), Linnebo develops an account of modal potentialism centred around the ontological notion of 'thinness'. He argues that the referents of the singular terms in a certain kind of Fregean abstraction principle (defined below) are abstract objects, and do in fact exist. Moreover, he argues that these referents are 'thin'; their 'existence does not (loosely speaking) make a substantial demand on the world' (Linnebo 2018, p. 3). Linnebo defends the view that mathematical objects are 'generated' 'successively' by iterated applications of such abstraction principles. He argues that this process enriches the semantic interpretations of the formal language. He understands the primitive modality of modal potentialism as 'interpretational', since it pertains to how the language is interpreted, and argues that it is not 'circumstantial', i.e. does not bear on how reality is; mathematical objects understood as such are metaphysically necessary and eternal existents. The formal framework of Linnebo's account of modal potentialism is a modal first-order theory with plural quantifiers, which he calls MPFO. He then develops a modal picture of the nature of sets, which motivates the addition of axioms describing said picture to MPFO. He considers a handful of theories based on these axioms, the

strongest of which he calls MS, and offers a mutual interpretability result between MS and ZF.

5.1.1 Dynamic abstraction

At the core of Linnebo's framework lie Fregean abstraction principles. Following Linnebo 2018, the general form of an abstraction principle is given by

$$\S\alpha = \S\beta \leftrightarrow \alpha \sim \beta; \tag{AP}$$

where 'α' and 'β' are first- or higher-order variables; '$\sim$' abbreviates a complex expression expressing an equivalence relation; and § is a term operator producing singular terms. Call the entities 'α' and 'β' range over 'specifications', and call the entities '$\S\alpha$' and '$\S\beta$' range over 'specified objects'. 'Inheritance principles' can be adopted that allow for reasoning about properties of the specified objects that are inherited from the specifications. Consider some formula ϕ that is invariant under '$\sim$', i.e. the universal closure of

$$\alpha_1 \sim \beta_1 \wedge \cdots \wedge \alpha_n \sim \beta_n \rightarrow (\phi(\alpha_1, \ldots, \alpha_n) \leftrightarrow \phi(\beta_1, \ldots, \beta_n))$$

holds in the object language.[3] Then a predicate ϕ^* can be introduced, governed in the object language by

$$\phi^*(\S\alpha_1, \ldots, \S\alpha_n) \leftrightarrow \phi(\alpha_1, \ldots, \alpha_n), \tag{IP}$$

defined on the specified objects.

Acceptable abstraction principles are surrounded by unacceptable ones, or 'bad companions'; some are inconsistent, while others are more subtly problematic. For example, consider Hume's Principle

$$\#F = \#G \leftrightarrow F \approx G, \tag{HP}$$

where 'F' and 'G' are second-order variables, '$\#F$' is to be read as 'the number of Fs', and '$\approx$' abbreviates an expression defining the equinumerosity of Fs and Gs. Compare this with Frege's similar Basic Law V

$$\varepsilon F = \varepsilon G \leftrightarrow \forall x(Fx \leftrightarrow Gx), \tag{BLV}$$

[3] Formulae in section 5.1 should be read as their universal closures.

where 'εF' denotes the extension of F. HP is consistent (Geach 1976, p. 446-447), whereas Basic Law V (BLV) is inconsistent[4]; following a Russell-style reasoning, by second-order comprehension let R be such that

$$\forall x(Rx \leftrightarrow \neg\exists H(Hx \wedge x = \varepsilon H)). \tag{Rus}$$

Now let $r = \varepsilon R$. Then by Rus and BLV, $\neg Rr \leftrightarrow Rr$. Separating acceptable from unacceptable abstraction principles is called the bad company problem (BCP).[5] However, according to Linnebo 2018, p. 55 'after decades of work by a number of capable researchers, there is no consensus in sight; on the contrary, new problems keep emerging [c.f. Studd 2016 and Cook and Linnebo 2018]'.

Rather than what Linnebo 2018, p. 52 calls the 'static' approach to abstraction that 'operates with a single fixed domain in which one or more abstraction principles are taken to be true'; Linnebo 2018, p. 55 advocates for a 'dynamic' approach to abstraction 'on which abstraction may result in "new" objects that lie beyond the "old" domain with which we began'. For example, the inconsistency of BLV merely shows that the Russell object r cannot be 'old'.

Linnebo calls an abstraction principle AP predicative if no specified objects (particular to said abstraction principle) figure in the range of any quantifier occurring in the expression '$\sim$' abbreviates. Moreover, Linnebo 2018, p. 56 demarcates the difference between a collection that is specified by its extension, 'say by means of a comprehensive list'; and a collection that is specified by its intension, 'say by means of [a] concept'. Linnebo shows that if domains can expand, then any abstraction on extensionally specified collections where $\sim$ is unaffected by domain expansions can be reformulated to become predicative; and argues this solves the BCP.

An oft-made objection against dynamic abstraction, is that individual instances of it are weaker than individual instances of static abstraction. Linnebo 2018, p. 60-61 for example notes that static HP implies the existence of infinitely many numbers by Frege's

[4]Russell pointed this out to Frege in a 1902 letter (2002).
[5]For an overview see Linnebo 2009.

'bootstrapping argument', whereas finite applications of dynamic HP cannot imply the existence of an infinite domain from a finite domain. Linnebo thinks this weakness can be overcome, by arguing that dynamic abstraction can be iterated indefinitely into the transfinite: dynamic abstraction applied to a base domain can result in an expanded domain, upon which dynamic abstraction can be applied again to potentially result in an even further expanded domain, and into the transfinite. Thus, dynamic abstraction can give rise to indefinitely extensible domains. Applying dynamic HP infinitely many times does imply the existence of infinitely many numbers.

In endorsing dynamic abstraction Linnebo has the burden of explaining how he understands the 'new' objects. He wants to satisfy the platonist thesis that mathematical objects exist independently of 'intelligent agents' (Linnebo 2018, p. 189), which he calls the independence thesis. Hence, Linnebo does not view the 'new' objects as 'created', only coming into existence when an intelligent agent introduces them in an expanded domain. Rather, he argues that dynamic abstraction is a picture of generative reference; all $\S\alpha$ introduced by dynamic abstraction refer post-abstraction, where they did not refer pre-abstraction. These referents then exist, since he endorses the Fregean conception of an object as the possible referent of a singular term. Moreover, some of these referents could not be referred to, and hence quantified over, pre-abstraction. Linnebo therefore argues that dynamic abstraction expands the *semantic interpretation* of the formal language qua uninterpreted syntactic strings (Linnebo 2018, §8). Furthermore, Linnebo also argues that his account of dynamic abstraction satisfies the independence thesis (Linnebo 2018, §11), by arguing that his account ensures the counterfactual independence of pure mathematics: '*however things had been*, pure mathematical objects would have existed and been related to one another just as they in fact are' (Linnebo 2018, p. 189, italics in original).

More specifically, consider some predicative abstraction principle R. Let $\mathcal{L}_0$ be a 'base' formal language that does not contain the singular terms $\S_R\alpha$, which is considered 'unproblematic' in that there is a semantic interpretation $\mathcal{I}_0$ of $\mathcal{L}_0$; and let D_0 be the base domain consisting of all and only the objects the terms of $\mathcal{L}_0$ refer to under $\mathcal{I}_0$. Moreover, let $\mathcal{L}_1$ be the uninterpreted expanded language that adds the term operator $\S_R$

to $\mathcal{L}_0$. Linnebo then argues that the singular terms $\S_R\alpha$ refer using what he calls Frege's generalised context principle, where 'it suffices for an expression to have reference that all appropriate contexts in which the expression occurs also refer' (Linnebo 2018, p. 127). Linnebo takes the appropriate contexts of an expression of $\mathcal{L}_1$ to be given by reducing it to a relevant expression of, or in Frege's terms 'recarving' in $\mathcal{L}_0$.

Linnebo subsequently provides 'assertibility conditions' of the syntactic strings that are well-formed formulae of $\mathcal{L}_1$, formulated in the meta-language using their 'recarving' (Linnebo 2018, §8-9), which 'govern the use of the [expanded] language' (Linnebo 2018, p. 138). He explicitly states that the assertibility conditions themselves are not a semantic interpretation; they rather express when a language community regards 'the assertoric utterance of a formula as correct' (Linnebo 2018, p. 138). Instead, Linnebo argues that the assertibility conditions can be used as 'linguistic data' by an interpreter to determine the semantic interpretation of (the uninterpreted syntactic strings that are well-formed formulae of) $\mathcal{L}_1$. He bases his theory on 'the idea that the use of a language is prior to semantic theorizing about it' (Linnebo 2018, p. 148).

If $\mathcal{L}_1$ is a two-sorted language, with a sort for 'old' and a sort for 'new' objects, then Linnebo makes Frege's generalised context principle precise by defining a translation τ from $\mathcal{L}_1$ to $\mathcal{L}_0$ (Linnebo 2018, §6); for example, '$\S\alpha$' is mapped to 'α', and '$\S\alpha = \S\beta$' is mapped to '$\alpha \sim \beta$'. Linnebo shows that only predicative abstraction guarantees the existence of τ (Linnebo 2018, §6). Linnebo 2018, p. 153-154 then uses τ to formulate the aforementioned assertibility conditions. For example, where σ is a variable assignment mapping α and β to objects a and b respectively, '$\S\alpha = \S\beta$' is assertible of a and b in D_0 iff '$\alpha \sim \beta$' is true under $\mathcal{I}_0$ relative to σ.

Linnebo then argues for the so-called 'non-reductionist' semantic interpretation of $\mathcal{L}_1$, denote it $\mathcal{I}_1$, which interprets the singular terms $\S\alpha$ as actually referring to abstract objects; as opposed to the so-called 'reductionist' interpretation where the singular terms $\S\alpha$ and α have the same reference. He shows that $\mathcal{I}_1$ expands $\mathcal{I}_0$, and moreover, if appropriate, $\mathcal{I}_1$ can satisfy inheritance principles, and *reinterpret* predicates from $\mathcal{L}_0$ to include 'new' objects. Linnebo 2018, p. 153 makes clear that he does endorse a form

of *meta-semantic* reductionism: '[w]hile I hold that the [expanded] language genuinely refers to [abstract objects], these truths about reference are grounded in some simpler truths about the use of the language, which make no mention of [said abstract objects]'. He argues that this meta-semantic reductionism ensures predicative abstraction gives rise to 'undemanding' reference; and thus undemanding, or thin, objects.

Moreover, let D_1 be the expanded domain consisting of all and only the objects the terms of $\mathcal{L}_1$ refer to under $\mathcal{I}_1$. If R is iterated, $\mathcal{L}_1$ is not expanded, since it already contains the term operator $\S_R$. Instead, only $\mathcal{I}_1$ is extended to a certain semantic interpretation $\mathcal{I}_2$. Further iterating R similarly does not expand $\mathcal{L}_1$. In addition, $\mathcal{I}_0$ can be raised to a semantic interpretation of $\mathcal{L}_1$ where no $\S_R\alpha$ refers. Thus, iterating the same dynamic abstraction principle gives rise to successively extended semantic interpretations of the *same* formal language, encoding ever-extended domains of the objects the interpreted terms of the formal language refer to.[6]

5.1.2 Mathematical objects

Recall that for Linnebo extensionally specified collections play an important role in ensuring dynamic abstraction. He captures these with pluralities, and develops a theory of set abstraction where abstracting on pluralities results in sets. Formally, he supplements the language of standard first-order logic with plural variables 'xx' and quantifiers binding them, along with binary predicates $\prec$ and SET; '$\exists xx$' stands for 'there are some things xx'; '$u \prec xx$' stands for 'u is one of xx'; '$SET(xx, v)$' stands for 'xx form the set y'; and '$u \in v$' abbreviates '$\exists xx(u \prec xx \wedge SET(xx, v))$'. Let $\mathcal{L}_{\mathsf{PFO}}$ denote this plural language. Besides the axioms of first-order logic, Linnebo introduces inference rules for the plural quantifiers that are just like those for the singular quantifiers; and the plural comprehension scheme

$$\exists xx \forall u(u \prec xx \leftrightarrow \phi(u)) \qquad \text{(P-Comp)}$$

[6] Linnebo does not address the point that iterating R only expands $\mathcal{L}_0$ to $\mathcal{L}_1$, and no further. However, he is aware of this, as he talks about iterative dynamic abstraction giving rise to transitions 'from one interpretation of a language to a more inclusive one' (Linnebo 2018, p. 61).

where $\phi(u) \in \mathsf{Form}(\mathcal{L}_{\mathsf{PFO}})$ does not contain 'xx' free. Linnebo allows for an empty plurality, and P-Comp not to be predicative (Linnebo 2018, p. 67, p. 208). Linnebo 2018, p. 210 calls this plural first-order logic (PFO).

Furthermore, Linnebo wants to say that every plurality forms a set. Thus, Linnebo 2018, p. 58-60 introduces a naive set theory where sets are obtained through the abstraction principle

$$\{xx\} = \{yy\} \leftrightarrow \forall u(u \prec xx \leftrightarrow u \prec yy), \tag{PLV}$$

which he calls 'Plural Law V'. However, due to Russell-style reasoning PLV is inconsistent. To avoid this, Linnebo makes PLV predicative. First, he expands $\mathcal{L}_{\mathsf{PFO}}$ to a two-sorted language, adding singular variables and terms of a new sort. He then implements PLV in the expanded language; '$\forall u$' ranges over the first sort, representing the 'old' domain; and $\{\cdot\}$ is an operator mapping plural terms of the first sort to singular terms of the second sort, representing the 'new' domain. This separates 'old' and 'new' objects, blocking Russell-style reasoning. Subsequently, Linnebo collapses the two-sorted language to the single-sorted language that adds to $\mathcal{L}_{\mathsf{PFO}}$ the predicate 'OLD', which is true in the meta-language of all and only objects from the 'old' domain. Linnebo 2018, p. 59 spells out this single-sorted language as follows. First, he splits PLV into two principles whose conjunction he claims is mutually interpretable with PLV; an existence criterion

$$\forall xx \exists v SET(xx, v) \tag{COLLAPSE}$$

and an identity criterion

$$SET(xx, v_1) \wedge SET(yy, v_2) \rightarrow (v_1 = v_2 \leftrightarrow \forall u(u \prec xx \leftrightarrow u \prec yy)). \tag{EXT}$$

He then restricts COLLAPSE to

$$\forall u(u \prec xx \rightarrow OLD(u)) \rightarrow \exists v SET(xx, v). \tag{COLLAPSE$^-$}$$

Finally, he takes the conjunction of COLLAPSE$^-$ and EXT for his theory of set abstraction. Linnebo 2018, p. 56 argues that extensionally specified collections, and thus pluralities stay fixed when enlarging domains, in contrast with intensionally specified

collections. In particular, he takes $\prec$ to be unaffected by domain expansions, ensuring COLLAPSE^{-} and EXT are well-behaved.

Linnebo 2018, p. 61 notes that the restriction of COLLAPSE to COLLAPSE^{-} 'becomes impractical when the abstraction step is iterated many times'. Instead, he suggests that:

> A more elegant option uses modal resources to represent the transition from one interpretation of a language to a more inclusive one. Let us add to our language the modal sentential operators $\Box$ and $\Diamond$. We may think of '$\Box\phi$' as meaning 'no matter what abstraction steps we carry out, it will remain the case that ϕ', and '$\Diamond\phi$' as 'we can abstract so as to make it the case that ϕ'. Obviously, this interpretation of the modal sentential operators is different from the more familiar one in terms of metaphysical modality. In the useful terminology of [Fine 2006], the present interpretation is "interpretational" rather than "circumstantial"; that is, it is concerned with how the language is interpreted, not with how reality is. In particular, every interpretational possibility is compatible with the metaphysically actual world. (Linnebo 2018, p. 61-62)

Thus, Linnebo supplements $\mathcal{L}_{\mathsf{PFO}}$ with modal sentential operators $\Box$ and $\Diamond$, where '$\Diamond\phi$' abbreviates '$\neg\Box\neg\phi$'. He subsequently develops a theory of modal set theory, based on the previously defined theory of non-modal set abstraction. The introduction of $\Box$ ensures the predicate $OLD(\cdot)$ is not needed.

To capture the behaviour of the primitive $\Box$, Linnebo 2018, p. 206-208 uses a possible-world semantics analogy to motivate certain proof-theoretic modal axioms. Recall that iterative applications of non-modal set abstraction do not expand $\mathcal{L}_{\mathsf{PFO}}$, instead giving rise to successively extended semantic interpretations $\mathcal{I}$ of the theory PFO, which encode ever-extended domains $D_{\mathcal{I}}$ of the objects the terms of $\mathcal{L}_{\mathsf{PFO}}$ refer to under $\mathcal{I}$. Linnebo then understands the 'possible worlds' as follows:

> Each stage of the process of abstraction can be regarded as a possible world. The ontology of each possible world consists of the objects and concepts that have been introduced thus far. The possible world also specifies how these entities are related and in this way settles all questions about which entities at this world satisfy the various atomic predicates. (Linnebo 2018, p. 206)

By the previous subsection 5.1.1, it becomes clear that 'stages of the process of abstraction' are precisely successively extended semantic interpretations of PFO. In particular, for

some semantic interpretation $\mathcal{J}$ of PFO, and for a non-modal formula ϕ of $\mathcal{L}_{\mathsf{PFO}}$; relative to some variable assignment, ϕ is to 'hold at the stage $\mathcal{J}$, or possible world $w_{\mathcal{J}}$', iff the expression ϕ is true under $\mathcal{J}$. It is important to emphasise that Linnebo cannot be providing a full-blown possible-world semantics, on pain of reducing the interpretational modality to quantification over semantic interpretations. This is not available to him, since he precisely takes semantic interpretations to be indefinitely extensible.[7] Hence, Linnebo uses a possible-world semantics *analogy*.

For any worlds $w_{\mathcal{I}_1}$ and $w_{\mathcal{I}_2}$ Linnebo takes the accessibility relation $w_{\mathcal{I}_1} \leq w_{\mathcal{I}_2}$ to mean that 'we can get from $[w_{\mathcal{I}_1}]$ to $[w_{\mathcal{I}_2}]$ by some permissible introduction of more entities' (Linnebo 2018, p. 206), implicitly meaning that $\mathcal{I}_1$ can be extended by non-modal set abstraction to $\mathcal{I}_2$. Though $\leq$ strictly speaking does not have a single intended interpretation since it is indefinitely extensible, Linnebo argues that $\leq$ should have all of the properties of a partial order (reflexive, transitive, and anti-symmetric); be convergent, i.e. $w \leq v \wedge w \leq u \rightarrow \exists w'(v \leq w' \wedge u \leq w')$; and be well-founded.[8] The modal logic S4.2 axiomatised by the schemata

$$\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi) \tag{K}$$

$$\Box\phi \rightarrow \phi \tag{T}$$

$$\Box\phi \rightarrow \Box\Box\phi \tag{4}$$

$$\Diamond\Box\phi \rightarrow \Box\Diamond\phi \tag{G}$$

and adding the necessitation rule NEC 'if $\vdash \phi$ then $\vdash \Box\phi$', where for each $\phi, \psi \in$ Form($\mathcal{L}_{\mathsf{PFO}}$), is sound and complete with respect to possible-world frames (K and NEC) with a reflexive (T), transitive (4), and convergent (G) accessibility relation. Since well-foundedness is not first-order definable, Linnebo considers S4.2 as the appropriate modal logic for a minimal conception of iterated non-modal set abstraction. Importantly, it is a necessary condition for indefinite extensibility that every possible world access another,

[7] For more on this see Studd 2019, p. 171-172.

[8] Linnebo does not provide a definition of well-foundedness here. He cannot mean that every non-empty set of worlds has a $\leq$-minimal element, i.e. $\forall S \subseteq W(S \neq \emptyset \rightarrow \exists w \in S\ \forall v \in S\ v \not\leq w)$, as that contradicts with the reflexivity of $\leq$. After asking Linnebo what he does mean, he clarified that he understands well-foundedness here as that every non-empty set of worlds has a $\leq$-*least* member, i.e. $\forall S \subseteq W(S \neq \emptyset \rightarrow \exists w \in S\ \forall v \in S(v \leq w \rightarrow v = w))$.

distinct possible world. Hence, it is a necessary condition that the appropriate modal logic prove (every instance of) the D schema $\Box\phi \rightarrow \Diamond\phi$. Indeed, S4.2 does. Linnebo takes the use of possible worlds as a 'motivational ladder that can now be kicked away' (Linnebo 2018, p. 208).

Linnebo formalises the fixity of pluralities by stipulating that $\prec$ is 'stable', with the two conditions

$$u \prec xx \rightarrow \Box(u \prec xx) \qquad (\text{STB}^{+}\text{-}\prec)$$

$$u \not\prec xx \rightarrow \Box(u \not\prec xx) \qquad (\text{STB}^{-}\text{-}\prec)$$

and 'inextensible', with the schema

$$\forall u(u \prec xx \rightarrow \Box\theta) \rightarrow \Box\forall u(u \prec xx \rightarrow \theta) \qquad (\text{INEXT-}\prec)$$

where $\theta \in \mathsf{Form}(\mathcal{L}_{\mathsf{PFO}})$. Linnebo 2018, p. 210 defines MPFO as the system expanding PFO with the axioms STB^{+}-$\prec$, STB^{-}-$\prec$, INEXT-$\prec$, $\forall x\forall y(x \neq y \rightarrow \Box x \neq y)$, and S4.2. Linnebo 2018, p. 64 also recursively defines the potentialist translation $\phi^{\Diamond}$ of a formula ϕ, where '$\forall\psi$' is mapped to '$\Box\forall\psi$' and '$\exists\psi$' to '$\Diamond\exists\psi$'; calling '$\forall$' and '$\exists$' the non-modalised quantifiers, and '$\Box\forall$' and '$\Diamond\exists$' the modalised quantifiers. Importantly, Linnebo's account of modal potentialism in (2018) is a relativist position about the non-modalised quantifiers. In contrast, it is an absolutist position about the modalised quantifiers, i.e. absolute generality is retrieved through '$\Box\forall$' (Linnebo 2018, §3.6).

Linnebo 2018, p. 74 proves a result crucial to his development of modal set theory, which he calls the Mirroring Theorem:

Theorem 5.1.1 (Mirroring Theorem)**.** *Let $\mathcal{L}$ be any non-modal first-order language, and let $\vdash$ be the relation of classical deducibility in $\mathcal{L}$. Let $\mathcal{L}^{\Diamond}$ be the modal first-order language adding $\Box$ and $\Diamond$ to $\mathcal{L}$, and let $\vdash^{\Diamond}$ be the relation of deducibility in $\mathcal{L}^{\Diamond}$ by $\vdash$,* S4.2*, and axioms asserting that all atomic predicates of $\mathcal{L}$ are stable. Then:*

$$\phi_1, \ldots, \phi_n \vdash \psi \textit{ iff } \phi_1^{\Diamond}, \ldots, \phi_n^{\Diamond} \vdash^{\Diamond} \psi^{\Diamond}$$

Linnebo 2018, p. 62 lets modal set abstraction be given by

$$\Box\forall xx\Diamond\exists v SET(xx, v) \qquad \text{(COLLAPSE}^{\Diamond}\text{)}$$

and

$$SET(xx, v_1) \wedge SET(yy, v_2) \rightarrow (v_1 = v_2 \leftrightarrow \Box\forall u(u \prec xx \leftrightarrow u \prec yy)). \qquad \text{(EXT}^{\Diamond}\text{)}$$

He subsequently develops a modal picture of the nature of sets, which motivates the addition of axioms describing said picture to MPFO (Linnebo 2018, §12). He considers a handful of modal set theories based on these axioms, and calls the strongest MS. Using the Mirroring Theorem, he proves that MS and ZF are mutually interpretable, ensuring the mathematical strength of his formal framework.

5.2 Interpretational modal syntactic predicate

As the previous section 5.1 has shown, Linnebo's framework carries its interpretational nature on its sleeve. Semantic interpretations are a keystone in his philosophical framework of dynamic abstraction; without them his philosophical framework would be radically different. Linnebo supplements $\mathcal{L}_{\mathsf{PFO}}$ with primitive modal sentential operators $\Box$ and $\Diamond$ to streamline iterated non-modal set abstraction. However, does this appropriately capture the interpretational nature of his framework?

5.2.1 Intended interpretation of interpretational necessity

For ϕ a non-modal sentence of $\mathcal{L}_{\mathsf{PFO}}$, recall that Linnebo understands '$\Box\phi$' in his metalanguage as follows:

> We may think of '$\Box\phi$' as meaning 'no matter what abstraction steps we carry out, it will remain the case that ϕ', and '$\Diamond\phi$' as 'we can abstract so as to make it the case that ϕ' ... In the useful terminology of [Fine 2006], the present interpretation is "interpretational" rather than "circumstantial"; that is, it is concerned with how the language is interpreted, not with how reality is. In particular, every interpretational possibility is compatible with the metaphysically actual world. (Linnebo 2018, p. 61-62)

Thus, Linnebo cannot be making a counterfactual claim with interpretational necessity, as that contradicts non-circumstantiality. Instead, the abstraction steps are to be understood as giving rise to successively extended semantic interpretations $\mathcal{I}$ of the theory PFO. Thus, Linnebo implicitly understands interpretational necessity of ϕ in his meta-language as

(1) No matter how the semantic interpretation of PFO is extended, it will remain the case that ϕ.

Since ϕ is a non-modal sentence of $\mathcal{L}_{\mathsf{PFO}}$; and since successively extended semantic interpretations of PFO are in the subject matter of Linnebo's interpretational necessity; as previously shown, if some semantic interpretation $\mathcal{J}$ is fixed, ϕ is the case iff 'ϕ holds at $\mathcal{J}$' iff the expression ϕ is true under $\mathcal{J}$. Hence, where $\overline{\phi}$ is the syntactic quotation of the value of the meta-variable ϕ (recall that semantic interpretations as understood by Linnebo give meaning to uninterpreted syntactic strings), Linnebo implicitly understands interpretational necessity of ϕ in his meta-language as

(2) No matter how the semantic interpretation of PFO is extended, $\overline{\phi}$ is true.

Given (2), it is clear that Linnebo's intended interpretation of interpretational necessity mentions, as opposed to uses, the complement ϕ. His intended interpretation is best formalised using the mention predicate approach; (2) does *not* concern whatever aspect of reality the interpreted expression ϕ concerns, since the expression is precisely being re-interpreted by extended semantic interpretations. Rather, (2) concerns the expression ϕ, qua re-interpretable syntactic entity. To reiterate a portion of the previous quote, Linnebo's intended interpretation of interpretational necessity 'is concerned with how the language is interpreted, not with how reality is. In particular, every interpretational possibility is compatible with the metaphysically actual world' (Linnebo 2018, p. 61-62). If his intended interpretation of interpretational necessity uses the complement ϕ, however, then it *is* concerned with how reality is. Thus, (2) mentions, and does not use, the complement ϕ. Furthermore, (2) is *not* properly formulated by a modal sentential operator, or a higher-order predicate of propositions, precisely because they use, as opposed to mention, the complement ϕ. Therefore, to best formalise Linnebo's intended interpretation of

interpretational necessity, (2) should be formulated by introducing a first-order unary modal syntactic predicate **N**, besides the modal sentential operator □.

One might observe that (2) alone does not settle whether '**N**' can be nested within the scope of '**N**'; the semantic interpretations $\mathcal{I}$ in (2) are of the non-modal language of $\mathcal{L}_{\mathsf{PFO}}$, which does not contain '**N**'. One could for example use (2) to argue for a typed hierarchy of interpretational modal syntactic predicates $\mathbf{N}_l$, for $l \in \mathbb{N}$, only applying to formulae containing '$\mathbf{N}_k$' for $k < l$.[9] Going back to Linnebo's framework however, this would seemingly defeat his purpose of introducing '□' in the first place; to have an elegant formalisation of his interpretational modality. Moreover, he neither types in his meta-language, nor in his object-language, yet he nests both interpretational necessity, and □ (consider for example the axiom schema $\Box\phi \rightarrow \Box\Box\phi$ (4)). Therefore, in line with Linnebo's framework the predicate **N** is understood as being type free, also applying to formulae containing '**N**'. Thus, Linnebo's justification for introducing the modal sentential operator □ also commits him to the legitimacy of introducing the type-free first-order unary modal syntactic predicate **N**, couched in some theory of syntax.

5.2.2 Replies and objections

Before exploring the consequences of Linnebo's intended interpretation of interpretational necessity being best formalised using the mention predicate approach, which as we have seen is prone to paradox; and whether the predicate **N** can give rise to a formalisation of modal potentialism; it is worth responding to some replies and objections.

Firstly, three preliminary remarks are in order. It is important to emphasise that this chapter, and this thesis in general, is neutral on the relativism/absolutism debate about quantifiers. The arguments developed in this chapter are specific to Linnebo's account of modal potentialism, and point out that more work needs to be done for it to be a viable account of (modal) potentialism. Furthermore, Linnebo's account of modal potentialism is already committed to the possibility, in the modal sentential operator sense ◊, of a

[9] (2) could give rise to a typed hierarchy by expanding the semantic interpretations $\mathcal{I}$ to interpret '**N**'; introducing a modal syntactic predicate $\mathbf{N}_2$ via an analogue of (2); again expanding the semantic interpretations to interpret '$\mathbf{N}_2$'; and further recursively introducing all $\mathbf{N}_l$.

theory of syntax; he provides a version of set-theoretic rank potentialism where V_α is possible ($\Diamond$) for any ordinal α; whereas V itself is not possible ($\neg\Diamond$). In particular, V_ω is possible, i.e. PA. Thus, Linnebo cannot object to introducing the predicate **N** by objecting to couching the predicate **N** in a theory of syntax.

The final remark involves Lewy's argument. Recall from chapter 1 that Lewy's argument is an objection to the mention predicate approach. In particular, the worry is that "2+2 = 4' is necessary' is itself not necessary, since '2 + 2 = 4' for example could have meant 3 + 3 = 5. In particular, this is because either '2 + 2 = 4' could have referred to the sentence '3 + 3 = 5'; or the meanings of our terms do not stay fixed. This objection was addressed in the introduction by ensuring that the names of sentences rigidly refer; and that the sentences they refer to mean what they actually do. In the case of Linnebo's intended interpretation of interpretational necessity, one might now wonder what happens when we assess whether "2 + 2 = 4' is interpretationally necessary' is interpretationally necessary. Recall that Linnebo emphasises that he does not understand interpretational necessity as circumstantial. Rather, he implicitly understands interpretational necessity of ϕ in his meta-language as

(2) No matter how the semantic interpretation of PFO is extended, $\overline{\phi}$ is true.

In particular, we see that shifting from a semantic interpretation of the formal language to an expanded one does not change the circumstances of what '2 + 2 = 4' could refer to. That is, there is no way '2 + 2 = 4' could have referred to the sentence '3 + 3 = 5'. Thus, this avoids the first half of Lewy's argument. Furthermore, for the second half of Lewy's argument, there is no need to worry about keeping the meanings of our terms fixed; precisely because interpretational necessity is interested in what happens when the semantic interpretations of the formal language change. What is needed is that we cannot take just any semantic interpretation of the formal language, but rather that we shift to an expanded semantic interpretation; which is precisely ensured by (2). Therefore, Lewy's argument does not apply to Linnebo's intended interpretation of interpretational necessity.

Secondly, one might worry that the argument used to introduce the predicate **N** is so general, that in fact it is an argument that possible-world semantics in general, even when constructed for the use predicate and operator approaches to modality, is best formalised using the mention predicate approach. For, given any logical language $\mathcal{L}$ that includes a modal sentential operator / higher-order modal predicate $\Box$ and for $\phi \in \mathsf{Sent}(\mathcal{L})$; in the meta-language of possible-world semantics $\Box\phi$ is standardly understood in introductory textbooks as

(2′) For all semantic interpretations $\mathcal{I}$ of $\mathcal{L}$ whose associated possible worlds $w_{\mathcal{I}}$ are accessible, $\overline{\phi}$ is true under $\mathcal{I}$.

Using the same reasoning as with (2), is it then also not the case that the necessity of ϕ mentions, as opposed to uses, the complement ϕ; and is thus best formalised using the mention predicate approach? Fortunately, this worry can be diffused, and the reasoning regarding (2) is not so sweeping. This is because the meta-linguistic nature of (2)′ does not stem from the intended interpretation of possible-world semantics. Rather, it stems from the meta-linguistic nature of providing a semantics for a language qua syntactic structure. Subsequently, if the intended interpretation of $\Box$ is concerned with how reality is, then $\Box$ is tracking ϕ *as interpreted at other worlds*, and not as uninterpreted syntactic entity. It is for example also not the case that the intended interpretation of the negation connective is meta-linguistic, even though a similar worry to the worry in question can be voiced regarding the negation sentential operator. The modal realism of D. Lewis 1986, for example, is even more explicit about this. According to D. Lewis 1986, §1.1, a possible world is a total way things could have been, being so varied that 'absolutely *every* way that a world could possibly be is a way that some world *is*' (D. Lewis 1986, p. 2, italics in original). In particular, a possible world is not meta-linguistic, nor seen as being associated to a semantic interpretation of some logical language. D. Lewis 1986, p. 20 then holds that 'modal sentential operators are quantifiers over possible worlds; that very often they are restricted; and that the applicable restriction may be different from the standpoint of different worlds, and so may be given by a relation of "accessibility"'. There is nothing meta-linguistic about such an intended interpretation

of possible-world semantics and understanding of $\Box\phi$.[10] Compare this with how the interpretational necessity of ϕ, as implicitly understood by Linnebo, *is* meta-linguistic; as shown above, it is precisely tracking ϕ as re-interpretable syntactic entity, and no such modal realism about possible worlds is available to Linnebo.

Thirdly, one might also wonder just how much of the interpretational nature of Linnebo's framework should actually be captured, given that thinking in terms of possible worlds is meant to be merely an analogy. Hence, does the syntactic predicate **N** not actually show that Linnebo cannot be taking his account of how the primitive modality is supposed to be understood, namely as interpretational necessity, that seriously? And does this not fit with the fact that modal potentialists take, and insist on taking the relevant modality as primitive? Reasoning along these lines stems from two related underlying objections. The first objection agrees that there has to be a link between a formalisation of modal potentialism and an intended interpretation of the primitive modality; however, it disagrees that in the case of Linnebo's framework this link has to be so close as to justify the introduction of the syntactic predicate **N**. The second objection goes even further and denies that there has to be any such link at all; no account of how the primitive modality is supposed to be understood needs to be given.

To address the second objection first, it is clear Linnebo is not thinking along these lines, for otherwise he would not dedicate a whole book to developing an account of how the primitive modality is supposed to be understood. More generally, however, without any account of the intended interpretation of the primitive modality we have no pre-theoretic intuitions about it, and thus it is unclear what is supposed to motivate axioms that determine its behaviour. Why pick precisely those? For example, why is the primitive modality convergent in Linnebo's framework; why does it satisfy the G axiom? Without any account of the intended interpretation of the primitive modality, it is unclear why any account of modal potentialism, and Linnnebo's in particular, is a position in the philosophy of mathematics; above and beyond providing interesting mathematical results, such as providing links between certain modal and non-modal set theories (for

[10] See also Plantinga 1978, p. 126-128 for such a modal realist interpretation of the modal sentential operators.

Linnebo in particular, these results are the Mirroring Theorem (MT); and the mutual interpretability result between MS and ZF). Thus, there has to be some kind of account of the intended interpretation of the primitive modality, and a link between such an account and a formalisation of modal potentialism.

To now address the first objection, if one is going to give an account of the intended interpretation of the primitive modality using dynamic abstraction and generative reference, as Linnebo's framework does; then syntactic entities will always be mentioned by the primitive modality, no matter how one spells this picture out. Hence, syntactic entities and the syntactic predicate **N** must figure in the formalisation of Linnebo's account of modal pontentialism. To deny **N**, one would need a completely different account of the intended interpretation of the primitive modality to the one that Linnebo provides. That is, there is no link between Linnebo's intended interpretation of the primitive modality and the formalisation of modal potentialism that does not also justify the introduction of the predicate **N**.

Fourthly, one might have a worry about the de dicto / de re distinction. As it was made clear in chapter 1, a first-order unary modal syntactic predicate, such as **N**, is most appropriate for formalising de dicto modal statements. However, de re modal statements are best formalised using a binary satisfaction predicate. Furthermore, Linnebo's formalisation of modal potentialism, in particular MPFO, makes crucial reference to de re modal statements. For example, recall the schema

$$\forall u(u \prec xx \rightarrow \Box\theta) \rightarrow \Box\forall u(u \prec xx \rightarrow \theta). \qquad \text{(INEXT-}\prec\text{)}$$

Does this then not show that **N** itself is already ill-suited to give rise to a formalisation of modal potentialism? Should Linnebo not be committed to the legitimacy of introducing a binary modal satisfaction predicate Sat that predicates on pairs consisting of names of formulae of $\mathcal{L}_{\mathsf{PFO}}$ and names of variable assignments? Indeed, the closest reading of Linnebo's framework should introduce Sat. However, if paradox arises when an attempt is made for **N** to give rise to a formalisation of modal potentialism; then paradox will also arise when such an attempt is made for Sat; precisely because Sat captures **N**. Thus,

a paradox involving **N** is more general than a paradox involving Sat, and in particular dispels any worry that paradox could be avoided if modal potentialism is formalised with the simpler **N** rather than the more complex Sat. Therefore, I choose to introduce **N** as opposed to Sat, even though it is not the closest reading of Linnebo's framework.

Finally, one might agree that Linnebo should introduce an interpretational necessity predicate **N**, besides an interpretational necessity operator □; however, one might object that **N** should behave the same as, i.e. satisfy analogues axioms as for □. For example, it is not clear that the failure of the unrestricted T-schema for a truth predicate of sentences should be reason to deny the Triv axiom for a truth operator.[11] To address this objection, we again turn to intended interpretations. Truth as an operator, or truth as a predicate of names of sentences, must have underpinning them two very different intended interpretations of truth. These differences can already be seen in what truth acts on; the first takes truth to use sentences, i.e. operate on already interpreted sentences; whereas the second takes truth to mention sentences, i.e. predicate on the pre-interpreted syntactic strings that form sentences. Linnebo might deny that □ and **N** should behave the same. But then it is unclear what the intended interpretation of □ is, and as we concluded when addressing the second objection above, there has to be one. For **N** was precisely introduced as a result of sharpening Linnebo's implicit intended interpretation of interpretational necessity. This sharpening made clear that his implicit intended interpretation of interpretational necessity is not most appropriately formalised using either the operator or use predicate approaches; but rather using the mention predicate approach. Thus, it is worth exploring the behaviour of **N**; and in particular, to see whether **N** can give rise to a formalisation of modal potentialism.

5.2.3 Introducing the predicate N

Michael Dummett (1978) argued that the correct logic for indefinite extensibility is Intuitionistic Logic, not Classical Logic (See also Linnebo and Shapiro 2019). Linnebo's formal framework is couched in Classical Logic. However, Linnebo 2018, p. 74 also hints at a potential departure to what he calls Semi-Intuitionistic Logic; Intuitionistic

[11] I would like to thank James Studd for pointing out this objection to me.

Logic where bounded quantification $\forall x(x \in y \rightarrow \phi)$ behaves Classically. Thus, for comprehensive inconsistency results the background logic of the rest of this chapter will be weakened from Classical Logic to Intuitionistic Logic.[12]

To achieve the necessary syntactical machinery required to introduce the type-free first-order unary modal syntactic predicate **N** besides the modal sentential operator $\Box$, Peano Arithmetic will be added to MPFO. From hereon we will denote Intuitionistic Peano Arithmetic, also called Heyting Arithmetic, by HA, to emphasise that in the background we are working in Intuitionistic Logic.[13] Formally, the language of MPFO is expanded with the predicate symbol **N** and the language of HA (as introduced in chapter 2), and is denoted $\mathcal{L}^{\mathbf{N}}_{\mathsf{MPFO}^+}$.[14] We let '$\mathbf{P}x$' be the dual of '$\mathbf{N}x$', and let it abbreviate '$\neg\mathbf{N}\overline{\ulcorner\neg\urcorner}\dot{\frown}x$'. Let MPFO^+ denote the Intuitionistic theory of the axioms of HA and MPFO in $\mathcal{L}^{\mathbf{N}}_{\mathsf{MPFO}^+}$.[15]

5.3 Inconsistency

We will now explore the behaviour of the type-free first-order unary modal syntactic predicate **N**; and whether **N** can give rise to a formalisation of modal potentialism. In particular, in a strikingly similar fashion to the semantic paradoxes it will be shown that natural principles for '**N**' are inconsistent.

First, consider a plausible predicate analogue to S4.2:

[12] Linnebo and Shapiro 2019 prove an Intuitionistic Mirroring Theorem, where $\vdash^{\Diamond}_{int}$ consists in the conditions for the classical Mirroring Theorem, along with the decidability of all atomic formulae of $\mathcal{L}$, i.e. the universal closures of $\phi_{atom} \vee \neg\phi_{atom}$.

[13] For an introduction to HA see Troelstra and van Dalen 1988. Gödel coding and diagonalisation for PA are purely constructive (see e.g. Smullyan 1992), and thus hold for HA.

[14] Here, '**N**' is added on top of '$\Box$'. However, '**N**' could also have replaced '$\Box$'. This will be further discussed at the end section 5.3, and is of no consequence for the inconsistency results set forth earlier in that section.

[15] As discussed in subsection 5.2.2, Linnebo is committed to the possibility of PA, and thus HA; and a possible inconsistency leads to an outright inconsistency in Intuitionistic Logic. This is to be contrasted with the strictest form of potentialism called Aristotelian potentialism, which takes that it is not even possible to complete $\aleph_0$. It seems the Aristotelian potentialist could avoid the inconsistency results from section 5.3 by denying the introduction of HA. An interesting avenue for further research would be to see what the status is of PA/HA as syntax theories, of Gödel's Incompleteness Theorems, and of the inconsistency results from section 5.3, under the potentialist translations of the axioms of PA/HA (or suitable modifications thereof that different brands of Aristotelian potentialism would accept).

Definition 5.3.1. Let $\mathsf{MPFO}^+ \cup \mathsf{S4.2_N}$ denote the theory that extends MPFO^+ with the axiom schemata

$$\mathbf{N}\overline{\ulcorner \phi \rightarrow \psi \urcorner} \rightarrow (\mathbf{N}\overline{\ulcorner \phi \urcorner} \rightarrow \mathbf{N}\overline{\ulcorner \psi \urcorner}) \tag{$\mathsf{K_N}$}$$

$$\mathbf{N}\overline{\ulcorner \phi \urcorner} \rightarrow \phi \tag{$\mathsf{T_N}$}$$

$$\mathbf{N}\overline{\ulcorner \phi \urcorner} \rightarrow \mathbf{N}\overline{\ulcorner \mathbf{N}\overline{\ulcorner \phi \urcorner} \urcorner} \tag{$\mathsf{4_N}$}$$

$$\mathbf{P}\overline{\ulcorner \mathbf{N}\overline{\ulcorner \phi \urcorner} \urcorner} \rightarrow \mathbf{N}\overline{\ulcorner \mathbf{P}\overline{\ulcorner \phi \urcorner} \urcorner} \tag{$\mathsf{G_N}$}$$

and is closed under the necessitation rule

$$\text{if } \vdash \phi \text{ then } \vdash \mathbf{N}\overline{\ulcorner \phi \urcorner} \tag{$\mathsf{Nec_N}$}$$

where $\phi, \psi \in \mathsf{Form}(\mathcal{L}^{\mathbf{N}}_{\mathsf{MPFO}^+})$.[16]

Could Linnebo use $\mathsf{MPFO}^+ \cup \mathsf{S4.2_N}$ to give rise to a formalisation of modal potentialism? As we have seen in section 2.3, by Montague's Paradox this leads to inconsistency with Classical Logic in the background. The same can be said if we have Intuitionistic Logic in the background; if we consider the proof of Montague's Paradox we see that the proof holds in Intuitionistic Logic as well. For a comprehensive result we will reproduce the theorem and proof in the current setting:

Theorem 5.3.1 (Montague's Paradox)**.** *If a theory* E *contains* $\mathsf{HA} \cup \mathsf{T_N}$ *and is closed under* $\mathsf{Nec_N}$*, then it is inconsistent (Montague 1963). Hence, so is* $\mathsf{MPFO}^+ \cup \mathsf{S4.2_N}$.

Proof. By the Diagonal Lemma there is a λ such that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg \mathbf{N}\overline{\ulcorner \lambda \urcorner} \tag{5.1}$$

We then have the following:

$\mathsf{E} \vdash \lambda \leftrightarrow \neg \mathbf{N}\overline{\ulcorner \lambda \urcorner}$	(5.1)	(5.2)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner \lambda \urcorner} \rightarrow \lambda$	$\mathsf{T_N}$	(5.3)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner \lambda \urcorner} \rightarrow \neg \mathbf{N}\overline{\ulcorner \lambda \urcorner}$	(5.3), (5.2)	(5.4)

[16] $\mathsf{K_N}$, $\mathsf{4_N}$, and $\mathsf{G_N}$ suitably adapted can also be stated as axioms quantifying over formulae, and not as axiom schemata.

$\mathsf{E} \vdash \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}$	(5.4)	(5.5)
$\mathsf{E} \vdash \lambda$	(5.5), (5.2)	(5.6)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner}$	$\mathsf{Nec}_{\mathbf{N}}$, (5.6)	(5.7)

Lines (5.5) and (5.7) imply E is inconsistent. Q.E.D.

Thus, Linnebo cannot use $\mathsf{MPFO}^{+} \cup \mathsf{S4.2}_{\mathbf{N}}$ to give rise to a formalisation of modal potentialism. Linnebo, or a modal potentialist endorsing his approach, could respond by weakening the axiom schema $\mathsf{T}_{\mathbf{N}}$ to $\mathbf{N}\overline{\ulcorner\phi\urcorner} \rightarrow \mathbf{P}\overline{\ulcorner\phi\urcorner}$ ($\mathsf{D}_{\mathbf{N}}$), for $\phi \in \mathsf{Form}(\mathcal{L}^{\mathbf{N}}_{\mathsf{MPFO}^{+}})$; the modal sentential operator analogue D is an axiom in the formalisation of modal potentialism developed by Studd 2013, instead of T. As pointed out in subsection 5.1.2 regarding D, $\mathsf{D}_{\mathbf{N}}$ intuitively captures indefinite extensibility, a necessary condition for Linnebo's approach, better than $\mathsf{T}_{\mathbf{N}}$. This seems a feasible response, since Studd 2013 also has a mutual interpretability result between the modal set theory he develops and ZF. Thus, consider the following weakening of $\mathsf{MPFO}^{+} \cup \mathsf{S4.2}_{\mathbf{N}}$:

Definition 5.3.2. Let $(\mathsf{MPFO}^{+} \cup \mathsf{S4.2}_{\mathbf{N}})^{-}$ denote the theory that extends MPFO^{+} with the axiom schemata $\mathsf{K}_{\mathbf{N}}$, $\mathsf{D}_{\mathbf{N}}$, $\mathsf{4}_{\mathbf{N}}$, and $\mathsf{G}_{\mathbf{N}}$; and is closed under the rule $\mathsf{Nec}_{\mathbf{N}}$.

Could Linnebo use $(\mathsf{MPFO}^{+} \cup \mathsf{S4.2}_{\mathbf{N}})^{-}$ to give rise to a formalisation of modal potentialism? As we have seen in section 4.1, by the Paradox of Transitive Seriality this leads to inconsistency with Classical Logic in the background. As with Montague's Paradox, the same can be said if we have Intuitionistic Logic in the background; if we consider the proof of Paradox of Transitive Seriality we see that the proof holds in Intuitionistic Logic as well. Again, for a comprehensive result we will reproduce the theorem and proof in the current setting:

Theorem 5.3.2. *If a theory* E *contains* $\mathsf{HA} \cup \mathsf{K}_{\mathbf{N}} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{4}_{\mathbf{N}}$ *and is closed under* $\mathsf{Nec}_{\mathbf{N}}$*, then it is inconsistent (Friedman and Sheard 1987, p. 14). Hence, so is* $(\mathsf{MPFO}^{+} \cup \mathsf{S4.2}_{\mathbf{N}})^{-}$.

Proof. By the Diagonal Lemma there is a λ such that

$$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner} \tag{5.8}$$

We then have the following:

$\mathsf{E} \vdash \lambda \leftrightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}$	(5.8)	(5.9)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda \leftrightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner}$	$\mathsf{Nec}_{\mathbf{N}}$, (5.9)	(5.10)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner} \leftrightarrow \mathbf{N}\overline{\ulcorner\neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner}$	$\mathsf{K}_{\mathbf{N}}$, (5.10)	(5.11)
$\mathsf{E} \vdash \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner} \leftrightarrow \neg\mathbf{N}\overline{\ulcorner\neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner}$	(5.11)	(5.12)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner} \rightarrow \neg\mathbf{N}\overline{\ulcorner\neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner}$	$\mathsf{D}_{\mathbf{N}}$	(5.13)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner} \rightarrow \mathbf{N}\overline{\ulcorner\mathbf{N}\overline{\ulcorner\lambda\urcorner}\urcorner}$	$\mathsf{4}_{\mathbf{N}}$	(5.14)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner} \rightarrow \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}$	(5.14), (5.13), (5.12)	(5.15)
$\mathsf{E} \vdash \neg\mathbf{N}\overline{\ulcorner\lambda\urcorner}$	(5.15)	(5.16)
$\mathsf{E} \vdash \lambda$	(5.16), (5.9)	(5.17)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\lambda\urcorner}$	(5.17)	(5.18)

Lines (5.16) and (5.18) imply E is inconsistent. Q.E.D.

Thus, Linnebo cannot use $(\mathsf{MPFO}^{+} \cup \mathsf{S4.2}_{\mathbf{N}})^{-}$ to give rise to a formalisation of modal potentialism. Furthermore, Linnebo's mutual interpretability result between MS and ZF correlates 'possible worlds' with V_γ for every ordinal γ, and $\leq$ with $\subseteq$ on the ordinals. Thus, if Linnebo were to attempt to develop a formalisation of modal potentialism using $\mathbf{N}$, in denying any of $\mathsf{D}_{\mathbf{N}}$, $\mathsf{4}_{\mathbf{N}}$, $\mathsf{K}_{\mathbf{N}}$, $\mathsf{G}_{\mathbf{N}}$, or $\mathsf{Nec}_{\mathbf{N}}$, he would need a completely different mutual interpretability result, if that is at all possible.

Though it is very unlikely that Linnebo would deny $\mathsf{4}_{\mathbf{N}}$, since it is a necessary condition for the iterated nature of his account of modal potentialism, there is reason to believe $\mathsf{D}_{\mathbf{N}}$ is the 'sole' culprit of the above inconsistencies. We add to $\mathcal{L}^{\mathbf{N}}_{\mathsf{MPFO}^{+}}$ the predicate Nat, with the set of natural numbers as its intended extension. Consider then the schema

$$\forall x(\mathsf{Nat}(x) \rightarrow \mathbf{N}\overline{\ulcorner\phi(\overline{\underset{\bullet}{x}})\urcorner}) \rightarrow \mathbf{N}\overline{\ulcorner\forall x(\mathsf{Nat}(x) \rightarrow \phi(x))\urcorner} \qquad (\mathsf{FO}_{\mathbf{N}})$$

where $\phi(x) \in \mathsf{Form}(\mathcal{L}^{\mathbf{N}}_{\mathsf{MPFO}^{+}} \cup \{\mathsf{Nat}\})$ contains at most 'x' free. This is a formalised version of the ω-rule, and is the Barcan Formula for natural numbers. One of the axioms of MS is

$$\forall x \exists yy \Box \forall u(u \prec yy \leftrightarrow u \in x) \qquad (\text{ED-}\in)$$

which Linnebo calls the 'extensional definiteness' of $\in$. Due to the axiom INEXT-$\prec$, ED-$\in$ implies the inextensibility of $\in$ (Linnebo 2018, Lemma 12.1); and hence implies the modal sentential operator version of the schema $\mathsf{FO_N}$

$$\forall x(x \in \mathbb{N} \rightarrow \Box\phi(x)) \rightarrow \Box\forall x(x \in \mathbb{N} \rightarrow \phi(x)) \tag{FO}$$

where $\phi(x) \in \mathsf{Form}(\mathcal{L}_{\mathsf{PFO}})$. Thus, consider the following theory:

Definition 5.3.3. Let $\mathsf{MPFO}^+ \cup \mathsf{D_N} \cup \mathsf{FO_N}$ denote the theory that extends MPFO^+ with the axiom schemata $\mathsf{K_N}$, $\mathsf{D_N}$, $\mathsf{G_N}$, and $\mathsf{FO_N}$; and is closed under the rule $\mathsf{Nec_N}$.

Could Linnebo use $\mathsf{MPFO}^+ \cup \mathsf{D_N} \cup \mathsf{FO_N}$ to give rise to a formalisation of modal potentialism? Given the Classical Logic framework of the possible-world P-semantics constructed in chapter 4, recall that in section 4.3 we proved the Multimodal 'No Dead-End' Paradox by providing a model-theoretic version of the proof of McGee's Paradox (1985); we provided part of the proof-theoretic proof of McGee's Paradox, and subsequently used soundness. Crucial to our proof is the fact that pw-models are built up from the standard model $\mathbb{N}$ of PA. In particular, recall that McGee's Paradox is not an outright inconsistency, but an ω-inconsistency. As with the two previous paradoxes considered in this section, McGee's Paradox still holds if we have Intuitionistic Logic in the background. As before, we will formulate the Intuitionistic version of McGee's Paradox in the current setting:

Theorem 5.3.3 (Intuitionistic McGee's Paradox)**.** *If a theory* E *contains* $\mathsf{HA} \cup \mathsf{K_N} \cup \mathsf{D_N} \cup \mathsf{FO_N}$ *and is closed under* $\mathsf{Nec_N}$*, then it is* ω*-inconsistent, due to an Intuitionistic modification of McGee 1985; there is a formula* $\phi(x)$ *with* $\mathsf{E} \vdash \phi(\overline{n})$ *for each* $n \in \mathbb{N}$*, yet* $\mathsf{E} \vdash \neg\forall x(\mathsf{Nat}(x) \rightarrow \phi(x))$*. Hence, so is* $\mathsf{MPFO}^+ \cup \mathsf{D_N} \cup \mathsf{FO_N}$.

Proof. By an application of the Uniform Diagonal Lemma, let $F(x, y, z)$ be a formula such that

$$\begin{aligned} \mathsf{E} \vdash \forall x\forall y\forall z(F(x,y,z) \leftrightarrow & \\ & \exists n(\mathsf{Nat}(n) \wedge x = \mathsf{S}(n) \wedge y = \overline{\ulcorner\forall y(F(\dot{\overline{n}}, y, \dot{\overline{z}}) \rightarrow \mathbf{N}y)\urcorner}) \\ & \vee (x = \underline{0} \wedge y = z)) \end{aligned} \tag{F}$$

$F(x, y, z)$ says that y codes the sentence which 'prefixes x many instances of '$\mathbf{N}$' to z' (along with the proper number of nested quotations). Then $\forall x(\mathsf{Nat}(x) \to \forall y F(x, y, z))$ expresses $\mathbf{N}^\star z$, where $\mathbf{N}^\star$ is the transitive closure of $\mathbf{N}$. By the Diagonal Lemma, we can find a λ such that

$$\mathsf{E} \vdash \lambda \leftrightarrow \forall x(\mathsf{Nat}(x) \to \forall y(F(x, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y)) \tag{5.19}$$

λ says that the result of prefixing '$\mathbf{N}$' any number of times to '$\neg\lambda$' holds. We then have the following Intuitionistically acceptable steps:

$\mathsf{E} \vdash \lambda \leftrightarrow \forall x(\mathsf{Nat}(x) \to \forall y(F(x, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y))$	(5.19)	(5.20)
$\mathsf{E} \vdash \lambda \to \forall y(F(\underline{0}, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y)$	(5.20)	(5.21)
$\mathsf{E} \vdash \forall y(F(\underline{0}, y, \overline{\ulcorner \neg \lambda \urcorner}) \leftrightarrow y = \overline{\ulcorner \neg \lambda \urcorner})$	F	(5.22)
$\mathsf{E} \vdash \lambda \to \mathbf{N}\overline{\ulcorner \neg \lambda \urcorner}$	(5.21), (5.22)	(5.23)
$\mathsf{E} \vdash \neg\lambda \to \neg\forall x(\mathsf{Nat}(x) \to \forall y(F(x, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y))$	(5.20)	(5.24)
$\mathsf{E} \vdash \overline{\mathbf{N}\ulcorner \neg\lambda \to \neg\forall x(\mathsf{Nat}(x) \to \forall y(F(x, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y))\urcorner}$	$\mathsf{Nec}_{\mathbf{N}}$, (5.24)	(5.25)
$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner \neg \lambda \urcorner} \to \overline{\mathbf{N}\ulcorner \neg\forall x(\mathsf{Nat}(x) \to \forall y(F(x, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y))\urcorner}$	$\mathsf{K}_{\mathbf{N}}$, (5.25)	(5.26)
$\mathsf{E} \vdash \lambda \to \overline{\mathbf{N}\ulcorner \neg\forall x(\mathsf{Nat}(x) \to \forall y(F(x, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y))\urcorner}$	(5.23), (5.26)	(5.27)
$\mathsf{E} \vdash \neg\overline{\mathbf{N}\ulcorner \neg\forall x(\mathsf{Nat}(x) \to \forall y(F(x, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y))\urcorner} \to \neg\lambda$	(5.27)	(5.28)
$\mathsf{E} \vdash \overline{\mathbf{N}\ulcorner \forall x(\mathsf{Nat}(x) \to \forall y(F(x, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y))\urcorner} \to \neg\lambda$	$\mathsf{D}_{\mathbf{N}}$, (5.28)	(5.29)
$\mathsf{E} \vdash \forall x(\mathsf{Nat}(x) \to \overline{\mathbf{N}\ulcorner \forall y(F(\overline{\underset{\bullet}{x}}, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y)\urcorner}) \to \neg\lambda$	$\mathsf{FO}_{\mathbf{N}}$, (5.29)	(5.30)
$\mathsf{E} \vdash \forall x(\mathsf{Nat}(x) \to$ $\forall y(F(\mathsf{S}(x), y, \overline{\ulcorner \neg \lambda \urcorner}) \leftrightarrow y = \overline{\ulcorner \forall y(F(\overline{\underset{\bullet}{x}}, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y)\urcorner}))$	F	(5.31)
$\mathsf{E} \vdash \forall x(\mathsf{Nat}(x) \to \forall y(F(\mathsf{S}(x), y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y)) \to \neg\lambda$	(5.31), (5.30)	(5.32)
$\mathsf{E} \vdash \forall x(\mathsf{Nat}(x) \to \forall y(F(x, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y)) \to \neg\lambda$	(5.32)	(5.33)
$\mathsf{E} \vdash \lambda \to \neg\lambda$	(5.20), (5.33)	(5.34)
$\mathsf{E} \vdash \neg\lambda$	(5.34)	(5.35)

Then by induction $\mathsf{E} \vdash \forall y(F(\overline{n}, y, \overline{\ulcorner \neg \lambda \urcorner}) \to \mathbf{N}y)$, for each $n \in \mathbb{N}$.

Base case:

$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner \neg \lambda \urcorner}$	$\mathsf{Nec}_{\mathbf{N}}$, (5.35)	(5.36)

$$\mathsf{E} \vdash \forall y(F(\underline{0}, y, \overline{\ulcorner \neg \lambda \urcorner}) \leftrightarrow y = \overline{\ulcorner \neg \lambda \urcorner}) \qquad F \qquad (5.37)$$

$$\mathsf{E} \vdash \forall y(F(\underline{0}, y, \overline{\ulcorner \neg \lambda \urcorner}) \rightarrow \mathbf{N}y) \qquad (5.37), (5.36) \qquad (5.38)$$

Inductive step: Assume $n \in \mathbb{N}$ and

$$\mathsf{E} \vdash \forall y(F(\overline{n}, y, \overline{\ulcorner \neg \lambda \urcorner}) \rightarrow \mathbf{N}y) \qquad (5.39)$$

Then:

$$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner \forall y(F(\overline{n}, y, \overline{\ulcorner \neg \lambda \urcorner}) \rightarrow \mathbf{N}y) \urcorner} \qquad \mathsf{Nec}_{\mathbf{N}}, (5.39) \qquad (5.40)$$

$$\mathsf{E} \vdash \forall y(F(\mathsf{S}(\overline{n}), y, \overline{\ulcorner \neg \lambda \urcorner}) \leftrightarrow y = \overline{\ulcorner \forall y(F(\overline{n}, y, \overline{\ulcorner \neg \lambda \urcorner}) \rightarrow \mathbf{N}y) \urcorner}) \qquad F \qquad (5.41)$$

$$\mathsf{E} \vdash \forall y(F(\mathsf{S}(\overline{n}), y, \overline{\ulcorner \neg \lambda \urcorner}) \rightarrow \mathbf{N}y) \qquad (5.41), (5.40) \qquad (5.42)$$

Yet:

$$\mathsf{E} \vdash \neg \forall x(\mathsf{Nat}(x) \rightarrow \forall y(F(x, y, \overline{\ulcorner \neg \lambda \urcorner}) \rightarrow \mathbf{N}y)) \qquad (5.35), (5.24) \qquad (5.43)$$

It is important to verify that these steps are Intuitionistically acceptable. In particular, (5.21) follows by instantiating x with $\underline{0}$ and syllogism. Next, (5.22) follows by instantiating x in F with $\underline{0}$; and (5.31) follows by relabeling x in F with another unused variable and then instantiating said unused variable with $\mathsf{S}(x)$. Furthermore, (5.23), (5.25), (5.26), (5.27), (5.29), (5.30), (5.32), (5.33), (5.34) all follow by syllogism. Moreover, (5.24) and (5.28) follow by contraposition. Finally, (5.35) follows by ex falso quodlibet. The same steps are made during the induction. Q.E.D.

The import of the ω-inconsistency from Theorem 5.3.3 is that, if Linnebo were to deny $\mathsf{4}_{\mathbf{N}}$ and accept $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$, no model of $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$ (if there are any) is an extension of the standard model of arithmetic with regards to the arithmetic vocabulary, and would include non-standard natural numbers. That is, the arithmetic vocabulary cannot be given its intended interpretation. This is precisely a problem because Linnebo is intending to provide a foundational alternative to set theory, which *can* give the arithmetic vocabulary its intended interpretation, in the sense that there are models of ZF that contain a copy of the standard model of arithmetic. Therefore, in accepting

$\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$ Linnebo has the burden of explaining why and how $\mathsf{4}_{\mathbf{N}}$ should be denied; of explaining why and how mathematicians are wrong about what the intended interpretation of arithmetic is; and of providing a mutual interpretability result between $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$ and set theory. Thus, this raises serious doubts as to whether $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$ can give rise to a formalisation of modal potentialism as standardly understood.

Of course, this assumes Linnebo can show that $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$ is consistent in the first place, which is a subtle and not so straightforward endeavour. In particular, $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$ combines arithmetic with plural logic, and is thus a theory expanding plural arithmetic. We know from Florio and Linnebo 2021, §7-8 that plural arithmetic is categorical given what they call the standard semantics for plural logic, which is a parallel to the full standard semantics for second order logic. That is, every model of plural arithmetic is isomorphic to the standard model $\mathbb{N}$ given the common standard semantics for plural logic. Thus, there are no models of $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$ given said semantics. However, there are also less common non-standard semantics for plural logic such as the plurality-based Henkin semantics developed and defended by Florio and Linnebo 2021, §8, which is a parallel to the Henkin semantics for second order logic. Given plurality-based Henkin semantics plural arithmetic is not categorical, and has non-standard models. Thus, there might be models of $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$ given such a semantics.[17]

To finish this section, recall that in this chapter we have added '**N**' on top of '$\Box$'. We did this because it is philosophically stronger to argue that Linnebo is committed to the legitimacy of introducing the type-free unary modal syntactic predicate **N**, couched in some theory of syntax; than to argue that this predicate should *also* replace the modal

[17]It would be an interesting avenue for further research to determine whether $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$ has models given plurality-based Henkin semantics. Väänänen and Wang 2015 show that second order arithmetic given Henkin semantics, though not categorical, *is* what they call internally categorical, which is a generalisation of categoricity; a theory E is internally categorical if all models of E in a fixed Henkin model are isomorphic according to said model. It is an interesting question whether one could also prove that plural arithmetic given plurality-based Henkin semantics is internally categorical. In such a case, any two copies of plural arithmetic contained in a given fixed model of $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$, if there is any such model, would be isomorphic, and in particular non-standard; which would further strengthen the worry that the arithmetic vocabulary cannot be given its intended interpretation.

sentential operator $\Box$. However, it is more theoretically economical to replace '$\Box$' with '$\mathbf{N}$', given that we are not interested in the behaviour of $\Box$. This naturally raises the question of what would happen if '$\mathbf{N}$' replaced '$\Box$'. First off, we immediately see that the two inconsistency results and the ω-inconsistency result have been formulated in such a way that they would still hold if we replaced '$\mathbf{N}$' with '$\Box$'; and thus the relevant parallel theories would still be ill-suited to give rise to a formalisation of modal potentialism. We also get a consistency result for the theory parallel to $\mathsf{MPFO}^{+} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}_{\mathbf{N}}$, by using another consistency result familiar from the literature on axiomatic theories of truth.

Definition 5.3.4. Let $\mathsf{HA} \cup \mathsf{K}_{\mathbf{N}} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{G}_{\mathbf{N}} \cup \mathsf{FO}'_{\mathbf{N}}$ denote the theory of the reduced language $\mathcal{L}_{\mathsf{PA}}$ that extends HA with the axiom schemata $\mathsf{K}_{\mathbf{N}}$, $\mathsf{D}_{\mathbf{N}}$, $\mathsf{G}_{\mathbf{N}}$, and

$$\forall x \mathbf{N}\overline{\ulcorner \phi(\overline{x}) \urcorner} \rightarrow \mathbf{N}\overline{\ulcorner \forall x \phi(x) \urcorner} \qquad (\mathsf{FO}'_{\mathbf{N}})$$

and is closed under the rule $\mathsf{Nec}_{\mathbf{N}}$.

Theorem 5.3.4. $\mathsf{HA} \cup \mathsf{K}_{\mathbf{N}} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{G}_{\mathbf{N}} \cup \mathsf{FO}'_{\mathbf{N}}$ *is consistent.*

Proof. We will prove that $\mathsf{HA} \cup \mathsf{K}_{\mathbf{N}} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{G}_{\mathbf{N}} \cup \mathsf{FO}'_{\mathbf{N}}$ couched in Classical Logic, i.e. $\mathsf{PA} \cup \mathsf{K}_{\mathbf{N}} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{G}_{\mathbf{N}} \cup \mathsf{FO}'_{\mathbf{N}}$, is consistent. Then, since Classical Logic is strictly stronger than Intuitionistic Logic, this will imply that $\mathsf{HA} \cup \mathsf{K}_{\mathbf{N}} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{G}_{\mathbf{N}} \cup \mathsf{FO}'_{\mathbf{N}}$ is consistent. But first, we need a lemma:

Lemma 5.3.1. *Let* E *be a Classical theory of the reduced language* $\mathcal{L}_{\mathsf{PA}}$ *that contains* PA*, the schemata* $\mathsf{D}_{\mathbf{N}}$ *and*

$$\mathbf{N}\overline{\ulcorner \phi \vee \psi \urcorner} \rightarrow \mathbf{N}\overline{\ulcorner \phi \urcorner} \vee \mathbf{N}\overline{\ulcorner \psi \urcorner} \qquad (\vee_{\mathsf{inf}})$$

and is closed under the rule $\mathsf{Nec}_{\mathbf{N}}$*. Then* E *proves the schema* $\mathsf{G}_{\mathbf{N}}$*.*

Proof.

$$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner \phi \urcorner} \rightarrow \neg \mathbf{N}\overline{\ulcorner \neg \phi \urcorner} \qquad \mathsf{D}_{\mathbf{N}} \qquad (5.44)$$

$$\mathsf{E} \vdash \neg \mathbf{N}\overline{\ulcorner \phi \urcorner} \vee \neg \mathbf{N}\overline{\ulcorner \neg \phi \urcorner} \qquad (5.44) \qquad (5.45)$$

$$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner \neg \mathbf{N}\overline{\ulcorner \phi \urcorner} \vee \neg \mathbf{N}\overline{\ulcorner \neg \phi \urcorner} \urcorner} \qquad \mathsf{Nec}_{\mathbf{N}} \qquad (5.46)$$

$$\mathsf{E} \vdash \mathbf{N}\overline{\ulcorner\neg\mathbf{N}\overline{\ulcorner\phi\urcorner}\urcorner} \vee \mathbf{N}\overline{\ulcorner\neg\mathbf{N}\overline{\ulcorner\neg\phi\urcorner}\urcorner} \qquad \vee_{\mathsf{inf}}, (5.46) \qquad (5.47)$$

$$\mathsf{E} \vdash \neg\mathbf{N}\overline{\ulcorner\neg\mathbf{N}\overline{\ulcorner\phi\urcorner}\urcorner} \rightarrow \mathbf{N}\overline{\ulcorner\neg\mathbf{N}\overline{\ulcorner\neg\phi\urcorner}\urcorner} \qquad (5.47) \qquad (5.48)$$

Q.E.D.

Then, by Friedman and Sheard 1987, p. 6 the Classical theory $\mathsf{PA} \cup \mathsf{K}_{\mathbf{N}} \cup \mathsf{D}_{\mathbf{N}} \cup \vee_{\mathsf{inf}} \cup \mathsf{FO}'_{\mathbf{N}}$ of the reduced language $\mathcal{L}_{\mathsf{PA}}$ that contains PA, $\mathsf{K}_{\mathbf{N}}$, $\mathsf{D}_{\mathbf{N}}$, $\vee_{\mathsf{inf}}$, and $\mathsf{FO}'_{\mathbf{N}}$; and is closed under $\mathsf{Nec}_{\mathbf{N}}$ is consistent.[18] Hence, by Lemma 5.3.1 the Classical theory $\mathsf{PA}\cup\mathsf{K}_{\mathbf{N}}\cup\mathsf{D}_{\mathbf{N}}\cup\mathsf{G}_{\mathbf{N}}\cup\mathsf{FO}'_{\mathbf{N}}$ is consistent. Q.E.D.

All in all, however, a type-free Intuitionistic account of **N** cannot give rise to a formalisation of modal potentialism, irrespective of whether it replaces, or is added on top of, '□'. Theorem 5.3.3 suggests that $\mathsf{D}_{\mathbf{N}}$, which is a necessary condition of modal potentialism, is the culprit of the inconsistencies, assuming **N** satisfies $\mathsf{K}_{\mathbf{N}}$ and $\mathsf{Nec}_{\mathbf{N}}$.

5.4 Concluding remarks

The aim of this chapter has been to show that Linnebo's account of modal potentialism in (2018) is a very reasonable attempt at satisfying the platonist thesis of the independence of mathematical objects, based on the dynamic picture of Fregean abstraction; it is not a mistake that semantic interpretations figure so heavily in Linnebo's account of dynamic abstraction. Moreover, his solution to the bad company problem requires that semantic interpretations can be indefinitely extended, i.e. that the D schema is satisfied. This chapter has shown that Linnebo is committed to the legitimacy of introducing a primitive modal syntactic predicate **N** of formulae in his object language, next to the modal sentential operator □. This is due to the syntactic nature of his intended interpretation of interpretational necessity; it is most appropriately formalised using the mention predicate

[18] The consistent set of sentences (where truth converges) that Friedman and Sheard 1987, p. 6 construct for subset D, more commonly known as the system FS, provides the consistency result. Here, T-Cons is $\mathsf{D}_{\mathbf{N}}$, ∀-inf is $\mathsf{FO}'_{\mathbf{N}}$, T-Comp implies $\vee_{\mathsf{inf}}$, T-Intro is $\mathsf{Nec}_{\mathbf{N}}$, and their base theory contains A-10, which is $\mathsf{K}_{\mathbf{N}}$. They inductively define a sequence of Classical models with initial model $M_0 := \langle \mathbb{N}, \emptyset \rangle$, and inductive step $M_{n+1} := \langle \mathbb{N}, \{\overline{\ulcorner\phi\urcorner} \mid M_n \vDash \phi\}\rangle$. Then by compactness $\mathrm{Th}_\infty := \{\phi \mid \exists k \forall n (n > k \rightarrow M_n \vDash \phi)\}$ is a Classically consistent set of sentences that contains FS, and in particular contains $\mathsf{PA} \cup \mathsf{K}_{\mathbf{N}} \cup \mathsf{D}_{\mathbf{N}} \cup \mathsf{FO}'_{\mathbf{N}} \cup \vee_{\mathsf{inf}}$. One can see FS as an axiomatisation of the semantics of the Revision Theory of Truth (discussed in a modal context in chapter 3) for all finite levels.

approach. However, the inconsistencies in this chapter show that **N** and □ must behave radically differently. Hence, Linnebo's intended interpretation of interpretational necessity and his proof theoretic approach, i.e. the behaviour of □ do not match up. More generally, Linnebo's intended interpretation of interpretational necessity is an example that provides yet another warning of the risks of paradox, in particular what goes wrong if one does not use the most appropriate formalisation when formalising philosophical notions.[19]

How then should **N**, i.e. Linnebo's interpretational modality behave? Some of the choices Linnebo has available for **N** are parallel to the choices available for a syntactic truth predicate; he can take inspiration from approaches to *typed* theories of truth, e.g. Tarski's hierarchy of truth; or from approaches to *type-free* theories of truth, such as Kripke 1975.[20] Probably the most promising choice Linnebo has available is to take inspiration from Stern 2015, §4; where a *modal* syntactic predicate is added alongside a syntactic *truth* predicate, and the behaviour of the modal syntactic predicate is characterised strictly by its interaction with the syntactic truth predicate. See also Chapter 2, footnote 15. This meaningfully preserves in a sense provided by Stern 2015 the modal principles Linnebo is interested in.[21] Whatever choices Linnebo makes, it will require that he further develops or changes his account of dynamic abstraction. Whether dynamic abstraction can then (1) give rise to absolute generality via the modalised quantifiers; (2) provide an adequate

[19]There is another general worry applying to Linnebo's account of modal potentialism in (2018) which also argues that the operator approach is not the most appropriate formalisation of the platonist modal potentialist's primitive modality. Recently, much work, such as that of Bacon 2018, has been done towards developing modal logicism, where modalities are reduced to higher-order logic. In particular, type-theoretic accounts of metaphysical necessity are developed which argue that metaphysical necessity is the broadest necessity operator, i.e. is absolute. Linnebo, and platonist modal potentialists in general, understand interpretational necessity as different from (any subspecies of) metaphysical necessity. That is because they take pure mathematical objects to be metaphysically necessary; yet they take certain mathematical objects, such as ordinals, to be precisely not interpretationally necessary. However, if interpretational necessity were to be a modal sentential or higher-order operator, we would have a contradiction with the position that metaphysical necessity is the broadest necessity operator; whatever is metaphysically necessary would be interpretationally necessary. This is another point that Linnebo must address. I would like to thank James Studd for pointing out this worry to me.

[20]For an example of a formalisation of the former see Feferman 1991. For a comprehensive analysis of the latter see either Friedman and Sheard 1987 or Leigh and Rathjen 2012, given a Classical or Intuitionistic background logic respectively. Leigh and Rathjen 2012 also investigate a classical truth predicate over an Intuitionistic base theory, which seems to best capture Semi-Intuitionistic Logic.

[21]Gorbow 2024 has been working on a response along these lines.

solution to the bad company problem; and (3) underpin an account of modal potentialism is unclear, and not a given.

Linnebo's account of modal potentialism in (2018) is not the only approach whose intended interpretation of the primitive modality is most appropriately formalised using the mention predicate approach. For example, Studd's approach in (2019) employs modalities concerned with shifts in interpretations of a fixed language, understood as shifts in 'semantic content' (Studd 2019, p. 148). In particular, an interpretation is a 'function mapping ... well-formed expressions to ... semantic values' (Studd 2019, p. 105). Thus, it seems Studd is susceptible to the inconsistencies in this chapter. Moreover, some philosophers such as Yu 2017 use approaches to modal potentialism to underpin a potentialist theory of propositions. This chapter makes it clear that Linnebo's interpretational necessity is ill-suited for a potentialist theory of propositions. The primitive modality involved in a potentialist theory of propositions is clearly concerned with how reality is; whereas Linnebo's intended interpretation of interpretational necessity 'is concerned with how the language is interpreted, not with how reality is' (Linnebo 2018, p. 61-62).

Further work is needed to show whether other approaches to modal potentialism involve mention modalities, thus likely susceptible to the challenges set by this chapter, and what the ramifications are. At the very least, every account of modal potentialism must have a response to these challenges. If the responses cannot stand the test of scrutiny, a further challenge is posed: if the primitive modality in modal potentialism cannot be understood interpretationally nor circumstantially, how *should* it be understood?

6
Conclusion

In contemporary philosophical discourse modality plays a central theoretical role. The three main approaches to formalising modalities and modal statements are the operator approach, the mention predicate approach, and the use predicate approach. They respectively formalise a modality as a sentential operator; as a first-order predicate that predicates on names of sentences / formulae, or names of propositions; and as a higher-order predicate that predicates on interpreted sentences or formulae, or propositions / facts / events / states of affairs themselves. These three approaches are not equally supported, and the mention predicate approach has fallen by the wayside because it is prone to paradox, the most notable being Montague's Paradox. However, this conclusion is too quick. In particular, there are modalities in the literature that are most appropriately formalised using the mention predicate approach. Therefore, in this thesis I have critically assessed and contributed to topics relating to the mention predicate approach.

In chapter 2 we provided the necessary logical background to be able to critically assess and contribute to topics relating to the mention predicate approach. We focussed on formalising modalities as first-order unary predicates that predicate on names of sentences, and we considered Peano Arithmetic as the theory of syntax we worked in. We then proved the central Diagonal Lemma and the Uniform Diagonal Lemma. Finally, we used

these Lemmas to prove Tarski's Undefinability of Truth, the Liar Paradox, and ultimately Montague's Paradox.

In chapter 3 we considered what happens if we block diagonalisation, with example the framework of Stern 2014c; Stern 2015. We then saw that blocking diagonalisation does not succeed in providing a uniform strategy for avoiding the paradoxes of the mention predicate approach. I further argued that diagonalisation is a necessary aspect of any formalisation of the mention predicate approach. We thus concluded that blocking diagonalisation does not provide a means of addressing the paradoxes of the mention predicate approach. However, we tentatively observed that we can nonetheless use Stern's framework when formalising a representational modality that mentions a class of representations that are structurally too poor to prove diagonalisation on them. Such a tentative claim naturally raises a fruitful question for further research: can we make this constructive takeaway of Stern's framework more than just tentative? Are there examples of representational modalities? If so, what kinds of classes of representations do they mention members of? What kinds of structure do they have? What behaviour do these representational modalities have? Are there limitative results as to the scope of their behaviour?

In chapter 4 we provided a characterisation theorem for the paradoxes of the mention predicate approach. We first considered more paradoxes besides Montague's Paradox that the mention predicate approach is prone to. These then made clear that we need to categorise when paradox arises, and when it does not. To explore this question of categorisation, we introduced a classical possible-world predicate-semantics for the mention predicate approach. We then proved that certain classes of multimodal frames admit no pw-models. We also showed that when expanding the classical possible-world P-semantics to a classical two-dimensional possible-world P-semantics, we cannot have the expected truth-conditions for well known indexicals of two-dimensional semantics conceived of as syntactic predicates. Finally, we proved that a multimodal frame $\mathfrak{F}$ admits a pw-model on every interpretation iff $\mathfrak{F}$ is converse well-founded. This result functions

as a categorisation theorem indicating when paradox arises in this generalised classical possible-world P-semantics.

Chapter 4 suggests many fruitful avenues for further research. A first big picture avenue for further research would be to generalise the framework and results of this chapter to multiple first-order binary modal satisfaction predicates, so as to also account for de re modal statements. In particular, how do the paradoxes of satisfaction considered in section 3.2.1 affect which pw-models admit pw-models on no / every interpretation? A second avenue would be to find more complete necessary and sufficient conditions for questions 1 and 3, which recall ask which multimodal frames admit no / a pw-model respectively. In particular, can we generalise the results of Halbach, Leitgeb, et al. 2003 to include contingent vocabulary, and to the multimodal setting? Can we further develop these results? A third avenue would be to explore whether there are any constructive results or further limitations, i.e. (philosophical) benefits, to a classical two-dimensional possible-world P-semantics. What kinds of interactions between reference and evaluation worlds do not pose an issue? A fourth avenue for further research would be to further explore the model theory of the classical possible-world P-semantics. In particular do we have modal frame correspondences, where modal predicate formulae define classes of frames?[1] The last avenue for further research we will consider, would be to investigate in what ways the classical possible-world P-semantics, and in particular the classification result of Theorem 4.4.1, can contribute to providing a unified strategy for avoiding paradox in the mention predicate approach. What must give?

In chapter 5 we provided an example of a modality in the philosophical literature that is most appropriately formalised using the mention predicate approach. In particular, we critically assessed the account of modal potentialism in the philosophy of mathematics developed by Linnebo 2018. We first considered Linnebo's framework. We then showed that due to the syntactic nature of Linnebo's intended interpretation of the primitive modality he is committed to the legitimacy of introducing a primitive modal syntactic predicate to his object language. Subsequently, we showed that principles for this modal

[1] See Blackburn et al. 2001, §3 for more on modal frame correspondences in the operator approach.

syntactic predicate that are fundamental to modal potentialism are inconsistent; no modal syntactic predicate can underpin an account of modal potentialism. Thus, Linnebo provides an account of modal potentialism that is incoherent. More generally, we see that this example provides yet another warning of the risks of paradox, in particular what goes wrong if one does not use the most appropriate formalisation when formalising philosophical notions. Many fruitful avenues for further research are discussed throughout chapter 5, mostly in section 5.4. The main avenue for further research is that Linnebo must respond to the challenges set by this chapter, and further develop or change his account of modal potentialism in (2018). These challenges also raise the question of which other accounts of modal potentialism are susceptible to them, and more broadly of how we should understand the modal potentialist's primitive modality.

Works Cited

Abusch, Dorit (2020). "Possible-Worlds Semantics for Pictures". In: *The Wiley Blackwell Companion to Semantics*. Ed. by Daniel Gutzmann et al. John Wiley & Sons, Ltd.

Bacon, Andrew (2018). "The Broadest Necessity". In: *Journal of Philosophical Logic* 47.5, pp. 733–785.

— (2023). *A Philosophical Introduction to Higher-Order Logics*. Routledge.

Barwise, Jon (2017). *Admissible Sets and Structures*. Perspectives in Logic. Cambridge University Press.

Berry, Sharon (2024). *A Logical Foundation for Potentialist Set Theory*. Cambridge University Press.

Blackburn, Patrick, Maarten de Rijke, and Yde Venema (2001). *Modal Logic*. Cambridge University Press.

Boolos, George (1993). *The Logic of Provability*. Cambridge University Press.

Boolos, George, John P. Burgess, and Richard C. Jeffrey (2007). *Computability and Logic*. 5th ed. Cambridge University Press.

Cappelen, Herman and Ernest Lepore (2007). *Language Turned on Itself: The Semantics and Pragmatics of Metalinguistic Discourse*. Oxford University Press.

Carnap, Rudolf (1934). *Logische Syntax der Sprache*. Springer.

— (1946). "Modalities and Quantification". In: *The Journal of Symbolic Logic* 11.2, pp. 33–64.

Cellucci, Carlo (2022). *The Theory of Gödel*. Springer.

Chellas, Brian (1980). *Modal Logic: An Introduction*. Cambridge University Press.

Cook, Roy T. and Øystein Linnebo (2018). "Cardinality and Acceptable Abstraction". In: *Notre Dame Journal of Formal Logic* 59.1, pp. 61–74.

Copeland, Brian John (2002). "The Genesis of Possible Worlds Semantics". In: *Journal of Philosophical Logic* 31.2, pp. 99–137.

Davidson, Donald (1979). "Quotation". In: *Theory and Decision* 11.1, pp. 27–40.

Davies, Martin and Lloyd Humberstone (1980). "Two Notions of Necessity". In: *Philosophical Studies* 38.1, pp. 1–30.

Dummett, Michael (1978). "The philosophical significance of Gødel's theorem". In: *Truth and Other Enigmas*. Harvard University Press, pp. 186–214.

Escher, Maurits Cornelis (1956). *Print Gallery*. Lithograph.

Evans, Gareth (1979). "Reference and Contingency". In: *The Monist* 62.2, pp. 161–189.

Feferman, Solomon (1960). "Arithmetization of Metamathematics in a General Setting". In: *Fundamenta Mathematicae* 49.1, pp. 35–92.

— (1991). "Reflecting on Incompleteness". In: *The Journal of Symbolic Logic* 56.1, pp. 1–49.

Fine, Kit (2005). "Our Knowledge of Mathematical Objects". In: *Oxford Studies in Epistemology*. Ed. by Tamar Szabo Gendler and John Hawthorne. Vol. 1. Oxford University Press, pp. 89–109.

— (2006). "Relatively Unrestricted Quantification". In: *Absolute Generality*. Ed. by Agustín Rayo and Gabriel Uzquiano. Oxford University Press, pp. 20–44.

Fischer, Martin and Johannes Stern (2015). "Paradoxes of Interaction?" In: *Journal of Philosophical Logic* 44.3, pp. 287–308.
Florio, Salvatore and Øystein Linnebo (2021). *The Many and the One: A Philosophical Study of Plural Logic*. Oxford University Press.
Friedman, Harvey and Michael Sheard (1987). "An Axiomatic Approach to Self-Referential Truth". In: *Annals of Pure and Applied Logic* 33.1, pp. 1–21.
Geach, Peter Thomas (1976). "Critical Notice". In: *Mind* 85.339, pp. 436–449.
Gödel, Kurt (1931). "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I". In: *Monatsheft für Mathematik und Physik* 38, pp. 173–198.
Goldblatt, Robert (2003). "Mathematical Modal Logic: A View of its Evolution". In: *Journal of Applied Logic* 1.5, pp. 309–392.
Gorbow, Paul (2024). *Potential as a Predicate*. Presented at IFIKK, Oslo, 2024-11-20.
Grabmayr, Balthasar, Volker Halbach, and Lingyuan Ye (2023). "Varieties of Self-Reference in Metamathematics". In: *Journal of Philosophical Logic* 52, pp. 1005–1052.
Gupta, Anil (1978). "Modal Logic and Truth". In: *Journal of Philosophical Logic* 7, pp. 441–472.
— (1980). *The Logic of Common Nouns*. Yale University Press.
— (1982). "Truth and Paradox". In: *Journal of Philosophical Logic* 11, pp. 1–60.
Hájek, Petr and Pavel Pudlák (1993). *Metamathematics of First-Order Arithmetic*. Springer Berlin.
Halbach, Volker (2001). "Semantics and deflationism". Unpublished habilitation thesis, University of Konstanz.
— (2006). "How Not to State T-Sentences". In: *Analysis* 66.4, pp. 276–280.
— (2008). "On a Side Effect of Solving Fitch's Paradox by Typing Knowledge". In: *Analysis* 68.2, pp. 114–120.
— (2014). *Axiomatic Theories of Truth*. 2nd ed. Cambridge University Press.
— (2021). "The Fourth Grade of Modal Involvement". In: *Modes of Truth: The Unified Approach to Truth, Modality, and Paradox*. Ed. by Carlo Nicolai and Johannes Stern. Routledge, pp. 209–230.
Halbach, Volker and Graham Leigh (2024). *The Road to Paradox: A Guide to Syntax, Truth, and Modality*. Cambridge University Press.
Halbach, Volker, Hannes Leitgeb, and Philip Welch (2003). "Possible-Worlds Semantics for Modal Notions Conceived as Predicates". In: *Journal of Philosophical Logic* 32.2, pp. 179–223. URL: http://www.jstor.org/stable/30226940 (visited on 04/26/2022).
Halbach, Volker and Albert Visser (2014a). "Self-Reference in Arithmetic I". In: *The Review of Symbolic Logic* 7.4, pp. 671–691.
— (2014b). "Self-Reference in Arithmetic II". In: *The Review of Symbolic Logic* 7.4, pp. 692–712.
Halbach, Volker and Shuoying Zhang (2017). "Yablo without Gödel". In: *Analysis* 77.1, pp. 53–59.
Hellman, Geoffrey (1993). *Mathematics without Numbers: Towards a Modal-Structural Interpretation*. Oxford University Press.
Hintikka, Jaakko (1957a). "Modality as Referential Multiplicity". In: *Ajatus* 20, pp. 49–64.
— (1957b). "Quantifiers in Deontic Logic". In: *Societas Scientiarum Fennica, Commentationes Humanarum Litterarum* 23.4, p.23.
Horsten, Leon and Hannes Leitgeb (2001). "No Future". In: *Journal of Philosophical Logic* 30.3, pp. 259–265.
Hughes, George Edward and Maxwell John Cresswell (1996). *A New Introduction to Modal Logic*. Routledge.

Kanger, Stig (1957). *Provability in Logic*. Acta Universitatis Stockholmiensis. Stockholm Studies in Philosophy; 1. Stockholm: Almqvist & Wiksell.

Kaplan, David and Richard Montague (1960). "A Paradox Regained". In: *Notre Dame Journal of Formal Logic* 1.3, pp. 79–90.

Kaye, Richard (1991). *Models of Peano Arithmetic*. Oxford Logic Guides. Clarendon Press, Oxford.

Kneale, Martha and William Kneale (1962). *The Development of Logic*. Clarendon Press, Oxford.

Kratzer, Angelika (1977). "What 'Must' and 'Can' Must and Can Mean". In: *Linguistics and Philosophy* 1.3, pp. 337–355.

Kripke, Saul (1959). "A Completeness Theorem in Modal Logic". In: *Journal of Symbolic Logic* 24.1, pp. 1–14.

— (1963a). "Semantical Analysis of Modal Logic I. Normal Modal Propositional Calculi". In: *Mathematical Logic Quarterly* 9.5-6, pp. 67–96.

— (1963b). "Semantical Considerations on Modal Logic". In: *Acta Philosophica Fennica* 16, pp. 83–94.

— (1965). "Semantical Analysis of Modal Logic II. Non-normal Modal Propositional Calculi". In: *The Theory of Models, Proceedings of the 1963 International Symposium at Berkeley*. Ed. by Alfred Tarski J. W. Addison Leon Henkin. Studies in Logic and the Foundations of Mathematics. North-Holland Publishing Company, Amsterdam, pp. 206–220.

— (1975). "Outline of a Theory of Truth". In: *The Journal of Philosophy* 72.19, pp. 690–716.

— (1980). *Naming and Necessity*. Blackwell, Oxford.

Laan, Boaz D (2025). "How Should We Understand the Modal Potentialist's Modality?" In: *Philosophia Mathematica* 33.2, pp. 242–252. DOI: 10.1093/philmat/nkaf007.

Leigh, Graham and Michael Rathjen (2012). "The Friedman-Sheard Programme in Intuitionistic Logic". In: *The Journal of Symbolic Logic* 77.3, pp. 777–806.

Lewis, Clarence Irving (1912). "Implication and the Algebra of Logic". In: *Mind* 21, pp. 522–531.

— (1914). "The Calculus of Strict Implication". In: *Mind* 23, pp. 240–247.

Lewis, Clarence Irving and Cooper Harold Langford (1932). *Symbolic Logic*. New York; London: Century.

Lewis, David (1986). *On The Plurality of Worlds*. Blackwell, Oxford.

Lewy, Casimir (1947). "Truth and Significance". In: *Analysis* 8.2, pp. 24–27.

Linnebo, Øystein (2009). "Introduction". In: *Synthese* 170.3, pp. 321–329.

— (2018). *Thin Objects*. Oxford University Press.

Linnebo, Øystein and Stewart Shapiro (2019). "Actual and Potential Infinity". In: *Noûs* 53.1, pp. 160–191.

Löb, Martin Hugo (1955). "Solution of a Problem of Leon Henkin". In: *Journal of Symbolic Logic* 20.2, pp. 115–118.

MacColl, Hugh (1906). *Symbolic Logic and its Applications*. England: Longmans, Green, and Co.

Malinas, Gary (1991). "A Semantics for Pictures". In: *Canadian Journal of Philosophy* 21.3, pp. 275–298.

McGee, Vann (1985). "How Truthlike Can a Predicate Be? A Negative Result". In: *Journal of Philosophical Logic* 14.4, pp. 399–410.

Meredith, Carew Arthur and Arthur Norman Prior (1996). "Interpretations of different modal logics in the 'Property Calculus'". In: *Logic and Reality: Essays on the Legacy of Arthur Prior*. Ed. by Brian John Copeland. Clarendon Press, Oxford.

Monk, James Donald (1976). *Mathematical Logic*. Springer New York.

Montague, Richard (1963). "Syntactical Treatments of Modality, With Corollaries on Reflexion Principles and Finite Axiomatizability". In: *Acta Philosophica Fennica* 16, pp. 153–167.

Mulligan, Kevin (2010). "The Truth Predicate vs the Truth Connective. On Taking Connectives Seriously." In: *Dialectica* 64.4, pp. 565–584.
Paseau, Alexander (2008). "Fitch's Argument and Typing Knowledge". In: *Notre Dame Journal of Formal Logic* 49.2, pp. 153–176.
— (2009). "How to type: reply to Halbach". In: *Analysis* 69.2, pp. 280–286.
Peacocke, Christopher (1978). "Necessity and Truth Theories". In: *Journal of Philosophical Logic* 7.473-500.
Peirce, Charles Sanders et al. (1994). *The Collected Papers of Charles Sanders Peirce*. InteLex Corporation.
Picollo, Lavinia (2018). "Reference in Arithmetic". In: *The Review of Symbolic Logic* 11.3, pp. 573–603.
— (2020). "Alethic Reference". In: *Journal of Philosophical Logic* 49.3, pp. 417–438.
Plantinga, Alvin (1978). *The Nature of Necessity*. Oxford University Press.
Prior, Arthur Norman (1971). *Objects of Thought*. Clarendon Press, Oxford.
Putnam, Hilary (1972). "The Meaning of 'Meaning'". In: *Minnesota Studies in the Philosophy of Science* 7, pp. 131–193.
Quine, Willard Van Orman (1943). "Notes on Existence and Necessity". In: *The Journal of Philosophy* 40.5, pp. 113–127.
— (1946). "Concatenation as a Basis for Arithmetic". In: *The Journal of Symbolic Logic* 11.4, pp. 105–114.
— (1947). "The Problem of Interpreting Modal Logic". In: *The Journal of Symbolic Logic* 12.2, pp. 43–48.
— (1953). "Three Grades of Modal Involvement". In: *Proceedings of the XIth International Congress of Philosophy*. Vol. 14. North-Holland Publishing Company, Amsterdam.
— (1956). "Quantifiers and Propositional Attitudes". In: *Journal of Philosophy* 53.5, pp. 177–187.
— (1977). "Intensions Revisited". In: *Midwest Studies in Philosophy* 2.1, pp. 5–11.
— (1980). "From a logical point of view: 9 logico-philosophical essays". In: 2nd ed. Harvard University Press. Chap. Reference and Modality, pp. 139–159.
Rahman, Shahid and Juan Redmond (2008). "Hugh MacColl and the birth of logical pluralism". In: *Handbook of the History of Logic*. Ed. by Dov Gabbay and John Woods. Vol. 4. Elsevier, pp. 533–604.
Recanati, François (2000). *Oratio Obliqua, Oratio Recta: An Essay on Metarepresentation*. MIT Press.
— (2001). "Open Quotation". In: *Mind* 110.439, pp. 637–687.
Rogers, Hartley (1967). *Theory of Recursive Functions and Effective Computability*. McGraw-Hill.
Russell, Bertrand (2002). "Letter to Frege". In: *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Ed. by Jean van Heijenoort. Harvard University Press.
Scambler, Chris (2021). "Can all things be counted?" In: *Journal of Philosophical Logic* 50.5, pp. 1079–1106.
Schweizer, Paul (1992). "A Syntactic Approach to Modality". In: *Journal of Philosophical Logic* 21.1, pp. 1–31.
Sider, Theodore (2010). *Logic for Philosophy*. Oxford University Press.
Smullyan, Raymond (1992). *Gödel's Incompleteness Theorems*. Oxford University Press.
Stalnaker, Robert (1978). "Assertion". In: *Syntax and Semantics, Volume 9: Pragmatics*. Ed. by P. Cole and J. Morgan. Academic Press, New York, pp. 315–332.
— (2023). *Propositions: Ontology and Logic*. Oxford University Press.

Stern, Johannes (2014a). "Modality and Axiomatic Theories of Truth I: Friedman – Sheard". In: *The Review of Symbolic Logic* 7.2, pp. 273–298.
— (2014b). "Modality and Axiomatic Theories of Truth II: Kripke – Feferman". In: *The Review of Symbolic Logic* 7.2, pp. 299–318.
— (2014c). "Montague's Theorem and Modal Logic". In: *Erkenntnis* 79.3, pp. 551–570.
— (2015). *Toward Predicate Approaches to Modality*. Vol. 44. Trends in Logic. Springer.
Studd, James (2013). "The Iterative Conception of Set: A (Bi-) Modal Axiomatisation". In: *Journal of Philosophical Logic* 42.5, pp. 697–725.
— (2016). "Abstraction Reconceived". In: *The British Journal for the Philosophy of Science* 67.2, pp. 579–615.
— (2019). *Everything, more or less: A defence of generality relativism*. Oxford University Press.
Tarski, Alfred (1983). "The Concept of Truth in Formalized Languages". In: *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Ed. by John Corcoran. 2nd ed. Originally published in Polish in 1933. Indianapolis: Hackett, pp. 152–278.
The Unicorn Tapestries (c. 1495-1505). Tapestry. The Metropolitan Museum of Art, The Cloisters, New York, United States of America.
Troelstra, A.S. and D. van Dalen (1988). *Constructivism in Mathematics: Volume 1*. Vol. 121. Studies in Logic and the Foundations of Mathematics. Elsevier.
Väänänen, Jouko and Tong Wang (2015). "Internal Categoricity in Arithmetic and Set Theory". In: *Notre Dame Journal of Formal Logic* 56.1, pp. 121–134.
Visser, Albert (1989). "Semantics And The Liar Paradox". In: *Handbook of Philosophical Logic*. Ed. by Dov Gabbay and Franz Günthner. Vol. 4. Dordrecht: Reidel, pp. 617–706.
Von Leibniz, Freiherr Gottfried Wilhelm, Austin Farrer, and E.M. Huggard (1951). *Theodicy : essays on the goodness of God, the freedom of man, and the origin of evil*. London: Routledge and Kegan Paul.
Williamson, Timothy (2013). *Modal Logic as Metaphysics*. Oxford University Press.
Wittgenstein, Ludwig et al. (2014). *Tractatus Logico-Philosophicus*. London: Routledge.
Yu, Andy Demfree (2017). "A Modal Account of Propositions". In: *Dialectica* 71.4, pp. 463–488.