Title
Quantification and Higher-Order Modal Logic

Author
Lauro, Greg

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Quantification and Higher-Order Modal Logic

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Logic & Philosophy of Science

by

Greg Lauro

Dissertation Committee:
Dean's Professor Kai F. Wehmeier, Chair
Associate Professor Sean Walsh
Associate Professor Gregory Scontras

2020

DEDICATION

In memory of G. Aldo Antonelli.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am fortunate that my graduate career united me with my now wife Kino, who has provided me with enduring love and support.

I owe immense thanks to my mentors throughout my university education: Aldo Antonelli, Sean Walsh and Kai Wehmeier. Their contributions made this journey possible, shaping my professional development.

I am also grateful for my other master's and doctoral committee members: Greg Scontras, Jeremy Heis and Jeffrey Helmrich. My research and broader education benefited greatly from our conversations together.

I am thankful for my opportunity to participate in the *Pedagogical Fellows* program at UCI. The training provided by Danny Mann and Matt Mahavongtrakul has been deeply influential in my teaching.

Many other faculty have been instrumental in my graduate education, and I remain indebted to each of them for the different ways in which they've supported me: Simon Huttegger, Arc Kocurek, Pen Maddy, Richard Mendelsohn, Lauren Ross, Brian Skyrms, and Ede Zimmermann.

Likewise, I would not be where I am today without the companionship of my fellow graduate student compatriots, past and present: Anna Bellomo, Justin Bruner, Elliott Chen, Jason Chen, Tom Colclough, Sam Eklund, Ben Feintzeig, Ethan Galebach, Kevin Kadowaki, Sam Kelso, Tobi Koch, Travis LaCroix, Fabio Lampert, Guilluame Massas, Chris Mitsch, Aydin Mohseni, Sarita Rosenstock, Gerard Rothfus, Hannah Rubin, Jeffrey Schatz, Tim Schmitz, Mike Schneider, Nick Smith, Will Stafford, Darby Vickers, and Tianyu Wu.

I would be remiss to not acknowledge the contributions of the administrative staff who've saved me from innumerable pitfalls: John Sommerhauser, Patty Jones, Jayne Lee, Dale Lee, Dan Paley, and Rachel Kraus-Lee.

Lastly, I would like to again thank my family: my wife, my parents and siblings, and our many pets. Their (sometimes literal) warmth has been a constant boon in my life.

Chapter Three is the product of joint work with Sean Walsh, tendered with permission.

VITA

Greg Lauro

EDUCATION

Ph.D. in Logic & Philosophy of Science **2020**
University of California, Irvine *Irvine, California*

M.A. in Philosophy **2017**
University of California, Irvine *Irvine, California*

B.A.S. in Mathematics, Linguistics, Philosophy (*summa.*) **2014**
University of California, Davis *Davis, California*

PUBLICATIONS

Structural Stability in Causal Decision Theory (with Simon Huttegger) **2020**
Erkenntnis

ABSTRACT OF THE DISSERTATION

Quantification and Higher-Order Modal Logic

By

Greg Lauro

Doctor of Philosophy in Logic & Philosophy of Science

University of California, Irvine, 2020

Dean's Professor Kai F. Wehmeier, Chair

This dissertation advances debates in modal metaphysics, philosophy of language and formal semantics by refining the logical systems traditionally employed in the analysis of natural language. Since linguistic phenomena underlie such debates, innovation in our approaches to semantics offers a means of resolution. My extension of these approaches through higher-order and alternative logics challenges established philosophical theses, while also providing a broad framework for the evaluation and comparison of philosophical and linguistic theories.

The first chapter extends an alternative semantics for quantification to the modal setting, establishing its cogency while clarifying the roles of identity in logic. The second chapter rebuts an argument by Timothy Williamson for necessitism in modal metaphysics by developing an admissible extension of his higher-order logic which, by his own methodology, undermines his claim. The third chapter (joint with Sean Walsh) develops a generalization of Montague's intensional type theory to allow for varying domains of objects across possible worlds, creating a framework suitable for the comparison of semantic theories. The fourth chapter considers the iterative application of reflection principles to formal theories of truth, revealing that what appears to be a purely mathematical choice between proof-theoretic reflection principles in fact commits one to competing stances on the properties of truth as a predicate.

Introduction: A Disneyland Semantics of Possible Worlds

> *Further, we see that there is no need of bringing in set theory to discuss [insolubilia], no need of a three-valued theory of some kind, no need of a "gappy" semantics condoning a truth-value hiatus, and no need of a Disneyland semantics of possible worlds, and the like.*
>
> — R.M. Martin, *Pragmatics, Truth, and Language*

This dissertation endorses two types of logical machinery, each the subject of much contention in the twentieth century. First, I accept the cogency of *higher-order predicate logic*, where the notions of predication and quantification are extended beyond the basic objects and their properties to permit the object language to talk of items like collections of properties or properties of properties. Second, I work closely with *intensions*, the meaning-components of expressions which determine their referents; in logic, intensions are sometimes conveyed indirectly through the use of modal operators, and they are sometimes taken as explicit objects of the system. These two concepts are often vitalized via *quantification*, a logical operation for generalization, and it is through this mechanism that their interaction comes to life.

Developments in logic have spurred advances and debates across disciplines: here, I focus on philosophy of language, metaphysics and linguistics. This dissertation follows this spirit, exploring how recent innovations in the areas of logic noted above have been employed (and sometimes misused) in philosophy. I build on such techniques and highlight how they can refine our understanding of concepts including identity, existence and truth.

Chapter 1 examines an alternative interpretation of variables due to Hintikka (1956) under which quantifiers exclude from their range the objectual values of any free variables and constants in their scope. I extend this interpretation to the setting of quantified modal logic with either constant or varying domains. This is of particular interest for modal metaphysics, where objectual quantification and identity are central to the debate. For these systems, I prove expressivity results which establish that one needn't give up the strength of the standard interpretation, even if one opts for Hintikka's alternative. Moreover, this analysis sheds light on the multifarious role of identity in logic.

More broadly, there are three morals from the project for philosophical logic. First, the semantics we assign to a given operator is not autogenous. This is a lesson we've seen in literature on the conditional in the past century, as well as in alternative semantics like dynamic semantics.[1] Second, relatedly, critical examinations of the orthodoxy can reveal errors and give further insight into the concepts rendered in a formal system.[2] Third, as is already apparent in classical logic, the interaction of quantifiers can be complex – in the chapter, the interplay between the modal operators and the objectual quantifiers complicate inter-translatability for the varying domain case.

Chapter 2 continues the theme of modal metaphysics, following Williamson (2010, 2013) into a second-order setting. Here, Williamson applies a novel criterion of expressivity to guide theory-choice, identifying a context in which necessitism is the more expressive theory,

[1]For more information, see Bennett (2003) on the former and Groenendijk and Stokhof (1991) as an exemplar of the latter.

[2]Another example of this can be found in the repudiation of Tarskian semantics by Wehmeier (2018).

giving us abductive evidence in favor of necessitism over contingentism. I argue that this result is not robust: after restoring parity between the power of object-quantification and world-quantification, Williamson's result no longer holds. Specifically, balance is obtained by adding new linguistic resources to the language, in the form of intensional operators with associated quantifiers as pioneered by Montague (1970).

For the actualism-possibilism debate, Williamson broke the standstill that the "choice between the possibilist and the actualist quantifier cannot be decided on formal grounds" (Fitting and Mendelsohn, 1998, p. 163). In doing so, he established the salience of higher-order logic as a means for adjudicating debates in metaphysics. Others writing in the same tradition, like Fritz and Goodman (2016), have continued in developing this philosophical space. However, these authors have approached the logical aspect from within the philosophical tradition: they expanded the usual first-order modal logic with a standard treatment of second-order logic. This approach essentially ignores innovations in formal semantics, notably the already higher-order modal logic that is Montague's intensional logic.

Chapter 3 formulates a generalization of intensional logic which is constructed to be compatible with the applications to metaphysics as desired by Stalnaker (2012) and the Williamson school.[3] The domain of objects associated to the base type e is permitted to vary across worlds, contra Montague. The notions of inner and outer domains are extended through the type hierarchy, and we examine natural constraints on the models which render them suitable for implementing theories of contingent propositions, situation semantics and perceptual reports.

While Montague's work was revolutionary in linguistics, giving rise to the field of formal semantics which united philosophers and linguists, his logical apparatus remained underappreciated outside of the study of language. Despite its elegance, the intensional logic was not immediately applicable to topics of interest in other areas of philosophy. Additionally,

[3] This chapter is the product of joint work with Sean Walsh.

the difficulty of amending the system gave logicians "the conviction that it is not a good idea to base a semantic theory on partial functions", as vocalized by Janssen (1986). Our project attempts to make the intensional apparatus more widely accessible, and hence we hope to provide an additional tool for advancing issues across philosophy.

Chapter 4 critiques a procedure advocated by Horsten and Leigh (2017) in which a modest formal theory of truth is expanded through the iterative application of proof-theoretic reflection principles. To explore this program, I articulate a new reflection principle and survey the closures of some popular candidate theories of truth (viz., typed Tarski biconditionals, Friedman-Sheard, and Kripke-Feferman) under these and more standard reflection principles. The results threaten the Horsten-Leigh project, for the pathology of certain base theories becomes explicit only after reflection, and the choice of reflection principle itself may lead a base theory to inconsistency. The most salient claim of the project for someone not invested in the Horsten and Leigh proposal is that reflection principles are semantically entangled, i.e., the semantic notion of truth is inextricable from the ostensibly syntactic notions involved in proof-theoretic reflection.

This project is a bit of a departure from the themes which unify the previous chapters. In spite of this, a few familiar notions present themselves in new ways. The two key concepts in the chapter are truth and provability, and the latter is famously analyzable as a modal operator.[4] There is a sense in which the logical systems studied are higher-order, as well, but the hierarchy orders the theories and their provability predicates rather than the variables and domains of quantification. Hence this is more akin to an infinitarily multi-modal logic than a modal higher-order logic. Nevertheless, the Barcan formulas are relevant in this context, being generally false and unprovable, while their converses are generally true and provably so. Incidentally, the 'pre-linguistic' Montague contributed to the topic with his Paradox of the

[4]See Boolos (1995) for an overview of provability logic.

Knower (Kaplan and Montague, 1960; Montague, 1963) and reflection theorems (Montague, 1961), though only the former features in the chapter.

Chapter 1

Quantified Modal W-Logic

1.1 Introduction

Hintikka (1956) identifies an alternative interpretation of variables under which quantifiers exclude from their range the objectual values of any free variables and constants in their scope. For example, in the formula $\forall x Rxy$, the evaluation of the universal $\forall x$ will exclude as candidates for x whatever value is assigned to y (by the background variable assignment). Hintikka calls this the *exclusive interpretation* of the variables.[1]

One motivation for this alternative comes from natural language, where naïve applications of the orthodox inclusive interpretation derive incorrect logical forms of expressions, or at least license undesired inferences. Consider the following sentences:

Kai owes everyone in the room five dollars. Therefore, he owes himself five dollars. (1.1)

[1]This view has a number of variants. We focus on Hintikka's preferred "weakly exclusive" interpretation. See also Rogers and Wehmeier (2012) for an analysis of these interpretations in connection with Tractarian logic.

The naïve formalizations are as follows:

$$\forall x Okx \therefore Okk \tag{1.2}$$

The inference in the formal rendition (1.2) is standardly valid, but the inference is likely unintended by the speaker. Similarly, the expression 'everybody loves somebody' (with wide scope for 'everybody') can be made true if each person is narcissistic, which again might run contrary to the intent of the speaker.

These undesired inferences can be blocked by more perspicuous use of language:

Kai owes everyone in the room *except himself* five dollars. (1.3)

$$\forall x(x \neq k \to Okx) \tag{1.4}$$

This approach implements the effect of the exclusive interpretation by the use of identity in the object language. Hence one might be tempted by the view that the exclusive interpretation arises from the analysis of imprecise speech. And indeed, there are uses where the inclusive interpretation is clearly intended, like in the Peano axioms:

Zero is not the successor of any number. (1.5)

$$\forall n \neg \mathsf{suc}(n, 0) \tag{1.6}$$

Here we intend for 0 to also fall under the range of quantifier $\forall n$.

Nevertheless, there are reasons to take the exclusive interpretation seriously. First, there seems to be no immediate reason why the inclusive interpretation needs to be taken as primary. Just as the expressions could be amended to block unintended inferences by the inclusive interpretation, we can do similarly to extend the range of the quantifiers to license the appropriate inferences. For example, we can translate (1.6) into the exclusive vernacular as

$$\forall n \neg \mathsf{suc}(n, 0) \wedge \neg \mathsf{suc}(0, 0) \tag{1.7}$$

to obtain the same truth conditions.[2]

Second, one's philosophical views might compel one to take the exclusive interpretation as primary. As displayed in (1.4), to get the exclusive reading using the inclusive quantifiers requires crucial use of an identity predicate. If one rejects identity as a binary relation (as in Wehmeier 2012), then the inclusive interpretation will not suffice for expressing routine elementary propositions. The natural solution for the identity-rejectionist is to adopt the exclusive interpretation as primary, for the inclusive reading can be obtained as necessary without additional expressive resources.

Given this background, Wehmeier (2004, 2008) advances the initial results of Hintikka (1956) to compare the expressive power of the two interpretations. In what follows, we will refer to classical logic as results from the inclusive interpretation with identity as $\text{FOL}^{=}$, and we will refer to the logic with the exclusive interpretation as W-logic (Wittgensteinian predicate logic).[3] Wehmeier proves the following expressivity results which vindicate the exclusive interpretation, glossed informally:

1. In signatures without constant symbols, $\text{FOL}^{=}$ is intertranslatable with W-logic.
2. In signatures with constant symbols, $\text{FOL}^{=}$ is intertranslatable with W-logic provided that distinct constants denote distinct objects in the domain.
3. In signatures with possibly co-denoting constant symbols, $\text{FOL}^{=}$ is intertranslatable with W-logic extended with a co-denotation predicate.

[2]Since there we be at most finitely-many constants and free variables in an expression, this trick of appending the excluded cases always suffices.

[3]This latter designation refers to the proposal of (Wittgenstein, 1971, 5.53ff.) on the elimination of the identity sign from logical notation, later systematized by Hintikka (1956) and Wehmeier (2012).

The co-denotation predicate mentioned in the third theorem allows assertions of the form $c \equiv d$ which expresses that the constants c and d designate the same object. Note that variables are not permitted to flank the co-denotation sign.

Now, the mechanism of quantification is obviously deployed in many logics besides FOL$^=$. For each of these in which the quantifiers are given the inclusive interpretation, it is natural to ask what analogues of the theorems above might hold for the exclusive interpretations. After all, the W-logicians might want to avail themselves of more than just classical logic.

In the context of quantified modal logic, Updike (2019) claims that the exclusive interpretation cannot, without incurring genuine philosophical costs, adequately express the sorts of metaphysical theses commonly articulated in this setting . In light of this challenge, I perform a systematic analysis of a number of common variations of quantified modal logic. For each, I prove the analogous expressivity theorems, noting what concessions (if any) are necessary on the part of the W-logician. I conclude by analyzing the multifarious role of identity in logic and highlighting a few future directions of study.

1.2 Preliminaries

We begin by specifying the routine details of quantified modal logic – of particular note are the clauses for W-satisfaction. We consider first the case of constant domain quantified modal logic.

Definition 1.1 (Language of QML). *The* language of quantified modal logic *(QML) consists of the following symbols:*

1. *The propositional connective* $\rightarrow$*, taking the rest as defined*
2. *The quantifier* $\forall$*, taking* $\exists$ *as defined*

3. *The modal operator $\Box$, taking $\Diamond$ as defined*

4. *Countably many variable symbols x_1, x_2, $x_3 \ldots$*

5. *For each $n > 0$, countably many n-ary predicate symbols P^n_1, P^n_2, $P^n_3 \ldots$*

6. *The propositional symbol $\bot$*

7. *Parentheses (and)*

As needed, we will use lower case roman letters to indicate metavariables over variable symbols, and we will use upper case roman letters to indicate metavariables over predicate symbols.

The language of quantified modal logic with identity *(QML$^=$) extends QML by the addition of the binary logical predicate =. When we wish to emphasize the W-logical counterpart of QML$^=$, we will write QMWL (*quantified modal W-logic*) in place of QML, even when there is no extension to the language.*

Definition 1.2 (Well-formed formula)**.** *An* atomic formula *is any expression of the form $Rx_1 \ldots x_n$ when R is an n-ary predicate symbol and x_i are variable symbols, or the propositional symbol $\bot$.*

The well-formed formulas *are defined inductively from the atomics in the usual manner.*

Definition 1.3 (QML Model, Constant Domain)**.** *A* constant domain QML model *$\mathcal{M}$ is given by a quadruple $\langle W, \mathcal{R}, \mathcal{D}, \mathcal{I} \rangle$ where W is a set of worlds, $\mathcal{R} \subset W \times W$ is an accessibility relation between worlds, $\mathcal{D} \neq \varnothing$ is a domain of objects, and $\mathcal{I}$ is an interpretation function assigning to each n-ary predicate P and each world w an extension $\mathcal{I}(P)(w) \subset \mathcal{D}^n$.*

When we wish to emphasize the underlying model $\mathcal{M}$, we will alternatively write W as $W(\mathcal{M})$, $\mathcal{D}$ as $|\mathcal{M}|$, and $\mathcal{I}(P)(w)$ as $P^{\mathcal{M}}(w)$.

Definition 1.4 (Satisfaction)**.** *We recursively define the notion of a* formula being satisfied at a world w in a model $\mathcal{M}$ by a variable assignment σ, *where a variable assignment is a function from the set of variable symbols to elements of the domain* $\mathcal{D}$.

We use $\vDash$ *for ordinary satisfaction in QML*$^=$*:*

1. $\mathcal{M} \vDash_w Rx_1 \dots x_n[\sigma]$ *iff* $\langle \sigma(x_1), \dots, \sigma(x_n) \rangle \in R^{\mathcal{M}}(w)$

2. $\mathcal{M} \nvDash_w \bot[\sigma]$, *always*

3. $\mathcal{M} \vDash_w A \to B[\sigma]$ *iff* $\mathcal{M} \nvDash_w A[\sigma]$ *or* $\mathcal{M} \vDash_w B[\sigma]$

4. $\mathcal{M} \vDash_w \forall x A[\sigma]$ *iff for all* $\tau \sim_x \sigma$*:* $\mathcal{M} \vDash_w A[\tau]$

5. $\mathcal{M} \vDash_w \Box A[\sigma]$ *iff for all* u *with* $w\mathcal{R}u$*:* $\mathcal{M} \vDash_u A[\sigma]$

We use $\Vdash$ *for W-satisfaction in QML:*

1. $\mathcal{M} \Vdash_w Rx_1 \dots x_n[\sigma]$ *iff* $\langle \sigma(x_1), \dots, \sigma(x_n) \rangle \in R^{\mathcal{M}}(w)$

2. $\mathcal{M} \nVdash_w \bot[\sigma]$, *always*

3. $\mathcal{M} \Vdash_w A \to B[\sigma]$ *iff* $\mathcal{M} \nVdash_w A[\sigma]$ *or* $\mathcal{M} \Vdash_w B[\sigma]$

4. $\mathcal{M} \Vdash_w \forall x A[\sigma]$ *iff for all* $\tau \sim_x \sigma$ *with* $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)]$*:* $\mathcal{M} \Vdash_w A[\tau]$

5. $\mathcal{M} \Vdash_w \Box A[\sigma]$ *iff for all* u *with* $w\mathcal{R}u$*:* $\mathcal{M} \Vdash_u A[\sigma]$

In the above definition, $\mathsf{FV}(B)$ *is the set of free variables in* B*. When* S *is a set, we write* $\tau[S]$ *to mean the pointwise image of* S *under* τ*. Currently, the only contrast with QML*$^=$ *is in the* $\forall$ *clause, where the constraint on repeated values captures the exclusive interpretation.*

1.3 Constant Domain Semantics for QMWL

We now proceed to prove the analogues of the intertranslatability theorems for the constant domain case. The strategy will be to provide truth-preserving translations between $\mathrm{QML}^{=}$ and QMWL.

Definition 1.5. *Let $\psi : \mathsf{Fmla}(QMWL) \to \mathsf{Fmla}(QML^{=})$ be identity on atomics, homomorphic for arrow and box, and*

$$\psi(\forall x A) = \forall x(\bigwedge_{y \in \mathsf{FV}(\forall x A)} x \neq y \to \psi(A))$$

Proposition 1.6. *For all $A \in \mathsf{Fmla}(QMWL)$, for all models $\mathcal{M}$, for all worlds $w \in W(\mathcal{M})$, for all $\mathcal{M}$-assignments σ, we have $\mathcal{M} \Vdash_w A[\sigma]$ iff $\mathcal{M} \vDash_w \psi(A)[\sigma]$.*

Proof. This proceeds by induction on formulas.

The base cases are immediate. We now consider the inductive steps in turn.

The arrow and box cases follow by rote via the induction hypothesis.

For the universal quantifier, we have the following: $\mathcal{M} \Vdash_w \forall x A[\sigma]$ iff for all $\tau \sim_x \sigma$ with $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)]$ we have $\mathcal{M} \Vdash_w A[\tau]$ (semantics for W-$\forall$), iff for all $\tau \sim_x \sigma$ with $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)]$ we have $\mathcal{M} \vDash_w \psi(A)[\tau]$ (induction hypothesis), iff for all $\tau \sim_x \sigma$ we have $\mathcal{M} \vDash_w \bigwedge_{y \in \mathsf{FV}(\forall x A)} x \neq y \to \psi(A)[\tau]$ (encoding the constraint on τ into the object language),[4] iff $\mathcal{M} \vDash_w \forall x(\bigwedge_{y \in \mathsf{FV}(\forall x A)} x \neq y \to \psi(A))[\sigma]$ (by the semantics for $\forall$). $\square$

We now move to demonstrate the reverse translation. Note that while we've assumed an expansive signature for purposes of illustration, in order to regiment the exclusive interpretation, all that is necessary is that one can form a tautology with one free variable and a

[4] Note that x is not free in $\forall x A$.

contradiction with two free variables, and so one could work in a much smaller language. Alternatively, the predicates defined below can be taken as primitive with the obvious semantic clauses.

Definition 1.7. *Define* $\top(x) := Rx \to Rx$ *for some arbitrary monadic predicate* R*, and define* $\bot(x,y) := \neg(Sxy \to Sxy)$ *for some arbitrary binary predicate* S*.*

Let $\varphi : \mathsf{Fmla}(QML^{=}) \to \mathsf{Fmla}(QMWL)$ *be identity on atomics except* $=$*, homomorphism for arrow and box, and*

$$\varphi(x = y) = \begin{cases} \top(x) & x, y \text{ identical variables} \\ \bot(x,y) & x, y \text{ distinct variables} \end{cases}$$

and

$$\varphi(\forall x A(x)) = \forall x \varphi(A(x)) \wedge \bigwedge_{y \in \mathsf{FV}(\forall x A(x))} \varphi(A(y))$$

In the second conjunct, we assume that the bound variables in $A(x)$ *have been renamed to avoid unintentional capture when the* y *is substituted for* x*.*[5]

In the above definition, the $\forall$ condition of the φ map generalizes the strategy described in the introduction for translating from (1.6) to (1.7). The condition for identity captures the belief that formal identity statements are either trivial or contradictory. The reader may also verify that this condition suffices to permit the definition of the finite numerical quantifiers ('there are at least n-many x's such that...'). This is explored in detail in Wehmeier (2008).

We now establish the adequacy of the translation.

Proposition 1.8. *For all* $A \in \mathsf{Fmla}(QML^{=})$*, for all models* $\mathcal{M}$*, for all worlds* $w \in W(\mathcal{M})$*, for all* $\mathcal{M}$*-assignments* σ *which are injective on* $\mathsf{FV}(A)$*, we have* $\mathcal{M} \vDash_w A[\sigma]$ *iff* $\mathcal{M} \Vdash_w \varphi(A)[\sigma]$*.*

[5]We use the familiar informal parenthetical notation of $A(x)$ to highlight the free variables of a formula for substitution.

Proof. This proceeds by induction on formulas.

The base cases for $\bot$ and non-logical predicates are immediate. For the identity case, $\mathcal{M} \vDash_w x = x[\sigma]$ always, and likewise for $\mathcal{M} \Vdash_w \top(x)[\sigma]$; similarly, $\mathcal{M} \vDash_w x = y[\sigma]$ never (by assumption of injectivity), and likewise for $\mathcal{M} \Vdash_w \bot(x,y)[\sigma]$.

The arrow and box cases proceed as before.

For the universal quantifier, we first proof the left-to-right direction: if $\mathcal{M} \vDash_w \forall x A[\sigma]$ then for all $\tau \sim_x \sigma$ we have $\mathcal{M} \vDash_w A(x)[\tau]$. Observe that while σ is injective on $\mathsf{FV}(\forall x A(x))$, τ might not be injective on $\mathsf{FV}(A(x))$, for it may assign a conflicting value to x. Consider any $\tau \sim_x \sigma$. There are two cases.

First, for any τ that is injective on $\mathsf{FV}(A(x))$, the induction hypothesis applies and we infer $\mathcal{M} \Vdash_w \varphi(A(x))[\tau]$. This holds for any $\tau \sim_x \sigma$ injective on $\mathsf{FV}(A(x))$, and hence it holds any $\tau \sim_x \sigma$ with $\tau(x) \notin \tau[\mathsf{FV}(\forall x A(x))]$ (by the partial injectivity of σ), and so we have $\mathcal{M} \Vdash_w \forall x \varphi(A(x))[\sigma]$.

Second, for any τ with $\tau(x) = \tau(y)$ for some $y \in \mathsf{FV}(\forall x A(x))$, since τ agrees on the value of x and y, it follows that $\mathcal{M} \vDash_w A(y)[\tau]$. Now τ was an x-variant of σ, and so $\mathcal{M} \vDash_w A(y)[\sigma]$ because x does not occur free in $A(y)$. By assumption σ is injective on $\mathsf{FV}(\forall x A(x)) = \mathsf{FV}(A(y))$, and so we can apply the induction hypothesis to obtain $\mathcal{M} \Vdash_w \varphi(A(y))[\sigma]$.

Collecting these cases together via the universal generalization over τ, we have that $\mathcal{M} \Vdash_w \forall x \varphi(A(x)) \wedge \bigwedge_{y \in \mathsf{FV}(\forall x A(x))} \varphi(A(y))[\sigma]$, which yields $\mathcal{M} \Vdash_w \varphi(\forall x A(x))[\sigma]$ as desired. The reverse direction proceeds similarly.

For the reverse direction, suppose we have that $\mathcal{M} \Vdash_w \varphi(\forall x A(x))[\sigma]$. Unpacking the translation and applying the semantics for conjunction, we have two conjuncts $\mathcal{M} \Vdash_w \forall x \varphi(A(x))[\sigma]$ and $\mathcal{M} \Vdash_w \bigwedge_{y \in \mathsf{FV}(\forall x A(x))} \varphi(A(y))[\sigma]$. We want to show that $\mathcal{M} \vDash_w \forall x A(x)[\sigma]$, i.e. for all $\tau \sim_x \sigma$ $\mathcal{M} \vDash_w A(x)[\tau]$. On $\mathsf{FV}(\forall x A(x))$ we know σ is injective by hypothesis. Hence if $\tau(x) \notin \tau[\mathsf{FV}(\forall x A(x))]$, we can apply the induction hypothesis to the first conjunct to obtain $\mathcal{M} \vDash_w A(x)[\tau]$. If instead $\tau(x) = \tau(y)$ for some $y \in \mathsf{FV}(\forall x A(x))$, we can infer $\mathcal{M} \Vdash_w \varphi(A(y))[\tau]$ from the second conjunct. Since x does not occur free in $A(y)$, τ is injective on $\mathsf{FV}(A(y))$, and so we can apply the induction hypothesis to obtain $\mathcal{M} \vDash_w A(y)[\tau]$. By the case hypothesis $\tau(x) = \tau(y)$, we then have $\mathcal{M} \vDash_w A(x)[\tau]$. Thus we've shown for any $\tau \sim_x \sigma$ $\mathcal{M} \vDash_w A(x)[\tau]$, hence $\mathcal{M} \vDash_w \forall x A(x)[\sigma]$. $\square$

Remark 1.9. While their proofs did not go into detail for the modal operators, the above propositions hold for arbitrary accessibility relations, because the modal operators do not differ between nor depend upon the two semantics. This will continue to be true when we add rigid constants below. However, permitting non-rigidity may complicate this.

1.3.1 Adding Constant Symbols

We now consider the addition of constant symbols to the language, in the usual manner.

Definition 1.10. *For what follows, we assume that the base language QML (and hence QML*$^=$ *and QMWL) has been extended with* $|I|$*-many constant symbols* c_i *for* $i \in I$ *a fixed index set. We then introduce the notion of a* term t*, which is either a variable or constant symbol. Accordingly, the atomic formulas are now built up from terms.*

We amend the semantics by considering the $\mathcal{M}$-interpretation relative to σ of a term $t^{\mathcal{M},\sigma}$, which is $\sigma(x)$ if t is a variable x, or $c^{\mathcal{M}}$ if t is a constant c (where $c^{\mathcal{M}} \in |\mathcal{M}|$). Note that in the current setting of constant domains with rigid constants, the interpretation of a term does not depend on the world of evaluation. The semantic clauses for $\vDash$ are standard, and we update the clauses for $\Vdash$ below.

Definition 1.11. *In what follows we let* $\mathsf{OC}(A)$ *be the constant symbols occurring in* A *and we let* $\mathsf{CV}^{\mathcal{M}}(A) = \{c^{\mathcal{M}} \mid c \in \mathsf{OC}(A)\}$ *be the objects designated by those constants:*

1. $\mathcal{M} \Vdash_w Rt_1 \ldots t_n[\sigma]$ *iff* $\langle t_1^{\mathcal{M},\sigma}, \ldots, t_n^{\mathcal{M},\sigma} \rangle \in R^{\mathcal{M}}(w)$

2. $\mathcal{M} \Vdash_w \forall x A[\sigma]$ *iff for all* $\tau \sim_x \sigma$ *with* $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)] \cup \mathsf{CV}^{\mathcal{M}}(A)$: $\mathcal{M} \Vdash_w A[\tau]$

In the quantifier clause, the idea is to assign a value to x distinct from all other objectual values mentioned in the subformula.

We now update our translations to accommodate constants and formulate the resulting intertranslatability propositions.

Definition 1.12. *We revise the map* $\psi : \mathsf{Fmla}(QMWL) \to \mathsf{Fmla}(QML^{=})$ *as follows, keeping all other clauses the same:*

$$\psi(\forall x A) = \forall x((\bigwedge_{y \in \mathsf{FV}(\forall x A)} x \neq y \wedge \bigwedge_{c \in \mathsf{OC}(\forall x A)} x \neq c) \to \psi(A))$$

Under this revision, we expect the following:

Proposition 1.13. *For all* $A \in \mathsf{Fmla}(QMWL)$*, for all models* $\mathcal{M}$*, for all worlds* $w \in W(\mathcal{M})$*, for all* $\mathcal{M}$*-assignments* σ*, we have* $\mathcal{M} \Vdash_w A[\sigma]$ *iff* $\mathcal{M} \vDash_w \psi(A)[\sigma]$*.*

Proof. We postpone this, for the proof is obvious given the proof of Proposition 1.18. □

Now we examine the reverse translation.

Definition 1.14. *Similarly, we update* $\varphi : \mathsf{Fmla}(QML^{=}) \to \mathsf{Fmla}(QMWL)$ *so that for identity we have*

$$\varphi(s = t) = \begin{cases} \top(s) & s, t \text{ identical terms} \\ \bot(s,t) & s, t \text{ distinct terms} \end{cases}$$

and for quantification we have

$$\varphi(\forall x A(x)) = \forall x \varphi(A(x)) \land \bigwedge_{y \in \mathsf{FV}(\forall x A(x))} \varphi(A(y)) \land \bigwedge_{c \in \mathsf{OC}(\forall x A(x))} \varphi(A(c))$$

The strategy in the above definition is the same as before, just with additional case breaks due to the addition of constants.

Under this revision, we expect the following:

Proposition 1.15. *For all $A \in \mathsf{Fmla}(QML^{=})$, for all models $\mathcal{M}$ for which $\cdot^{\mathcal{M}}$ is injective on constants, for all worlds $w \in W(\mathcal{M})$, for all $\mathcal{M}$-assignments σ which are injective on $\mathsf{FV}(A)$ with $\sigma(x) \notin \mathsf{CV}^{\mathcal{M}}(A)$ for each $x \in \mathsf{FV}(A)$, we have $\mathcal{M} \vDash_w A[\sigma]$ iff $\mathcal{M} \Vdash_w \varphi(A)[\sigma]$.*

We focus on a strengthening of the above proposition. Toward that end, we extend the language by addition of a binary predicate $\equiv$ for co-denotation.

Definition 1.16. *We extend the language QMWL with a fixedly-interpreted binary predicate $\equiv$ (hence QMWL is now distinct from QML). An atomic formula with $\equiv$ is well-formed only when flanked by constant symbols on each side, as in $c \equiv d$. The truth conditions are as follows: $\mathcal{M} \Vdash_w c \equiv d[\sigma]$ iff $c^{\mathcal{M}} = d^{\mathcal{M}}$.*

Having expanded the language, we must accordingly extend the translation.

Definition 1.17. *We extend ψ with $c \equiv d \mapsto c = d$.*

Then we have:

Proposition 1.18. *For all $A \in \mathsf{Fmla}(QMWL)$, for all models $\mathcal{M}$, for all $w \in W(\mathcal{M})$, for all $\mathcal{M}$-assignments σ, we have $\mathcal{M} \Vdash_w A[\sigma]$ iff $\mathcal{M} \vDash_w \psi(A)[\sigma]$.*

Proof. We proceed by induction on formulas. For the base case, we have three subcases. For A of the form $Rt_1 \ldots t_n$ or $\bot$, the result is immediate. The case for $\equiv$ follows from the definition of $\varphi(c \equiv d)$.

For the inductive step, the subcases for $\to$ and $\Box$ follow directly by induction hypothesis (as before).

For the universal quantifier, we have the following: $\mathcal{M} \Vdash_w \forall x A[\sigma]$ iff for all $\tau \sim_x \sigma$ with $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)] \cup \mathsf{CV}^{\mathcal{M}}(A)$ we have $\mathcal{M} \Vdash_w A[\tau]$ (semantics for W-$\forall$), iff for all $\tau \sim_x \sigma$ with $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)] \cup \mathsf{CV}^{\mathcal{M}}(A)$ we have $\mathcal{M} \vDash_w \psi(A)[\tau]$ (induction hypothesis), iff for all $\tau \sim_x \sigma$ we have $\mathcal{M} \vDash_w (\bigwedge_{y \in \mathsf{FV}(\forall x A)} x \neq y \wedge \bigwedge_{c \in \mathsf{OC}(\forall x A)} x \neq c) \to \psi(A)[\tau]$ (encoding the constraints on τ into the object language), iff $\mathcal{M} \vDash_w \forall x((\bigwedge_{y \in \mathsf{FV}(\forall x A)} x \neq y \wedge \bigwedge_{c \in \mathsf{OC}(\forall x A)} x \neq c) \to \psi(A))[\sigma]$ (semantics for $\forall$). □

Now we turn to the reverse translation.

Definition 1.19. *We redefine φ so that for identity:*

$$\varphi(s = t) = \begin{cases} \top(s) & s,t \text{ identical variables} \\ \bot(s,t) & s,t \text{ distinct variables} \\ s \equiv t & s,t \text{ constants} \\ \bot(s,t) & s,t \text{ one constant, one variable} \end{cases}$$

Under this revision, we have the following:

Proposition 1.20. *For all $A \in \mathsf{Fmla}(QML^{=})$, for all models $\mathcal{M}$, for all worlds $w \in W(\mathcal{M})$, for all $\mathcal{M}$-assignments σ which are injective on $\mathsf{FV}(A)$ with $\sigma(x) \notin \mathsf{CV}^{\mathcal{M}}(A)$ for each $x \in \mathsf{FV}(A)$, we have $\mathcal{M} \vDash_w A[\sigma]$ iff $\mathcal{M} \Vdash_w \varphi(A)[\sigma]$.*

Proof. We proceed by induction on formulas.

The base cases for $\bot$ and non-logical predicates are immediate. For the identity case, first, $\mathcal{M} \vDash_w x = x[\sigma]$ always, and likewise for $\mathcal{M} \Vdash_w \top(x)[\sigma]$; second, $\mathcal{M} \vDash_w x = y[\sigma]$ never (by assumption of injectivity), and likewise for $\mathcal{M} \Vdash_w \bot(x,y)[\sigma]$; third, $\mathcal{M} \vDash_w x = c[\sigma]$ never (by assumption), and likewise for $\mathcal{M} \Vdash_w \bot(x,c)[\sigma]$; lastly, $\mathcal{M} \vDash_w c = d[\sigma]$ iff $c^{\mathcal{M}} = d^{\mathcal{M}}$ iff $\mathcal{M} \Vdash c \equiv d[\sigma]$.

The arrow and box cases proceed as before.

For the universal quantifier, we first prove the left-to-right direction: if $\mathcal{M} \vDash_w \forall x A[\sigma]$ then for all $\tau \sim_x \sigma$ we have $\mathcal{M} \vDash_w A(x)[\tau]$. Consider any $\tau \sim_x \sigma$. There are three cases.

First, if τ is injective on $\mathsf{FV}(A(x))$ with $\sigma(z) \notin \mathsf{CV}^{\mathcal{M}}(A(z))$ for each $z \in \mathsf{FV}(A(x))$, then the induction hypothesis applies and we infer $\mathcal{M} \Vdash_w \varphi(A(x))[\tau]$. This holds for any $\tau \sim_x \sigma$ injective on $\mathsf{FV}(A(x))$ with $\sigma(z) \notin \mathsf{CV}^{\mathcal{M}}(A(x))$ for each $z \in \mathsf{FV}(A(x))$, and hence it holds for any $\tau \sim_x \sigma$ with $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)] \cup \mathsf{CV}^{\mathcal{M}}(A)$ (by the partial injectivity of σ), and so we have $\mathcal{M} \Vdash_w \forall x \varphi(A(x))[\sigma]$.

Second, if τ is such that $\tau(x) = \tau(y)$ for some $y \in \mathsf{FV}(\forall x A(x))$, then since τ agrees on the value of x and y, it follows that $\mathcal{M} \vDash_w A(y)[\tau]$. Now τ was an x-variant of σ, and so $\mathcal{M} \vDash_w A(y)[\sigma]$ because x does not occur free in $A(y)$. By assumption σ is injective on $\mathsf{FV}(\forall x A(x)) = \mathsf{FV}(A(y))$ with $\sigma(z) \notin \mathsf{CV}^{\mathcal{M}}(\forall x A(x)) = \mathsf{CV}^{\mathcal{M}}(A(y))$ for each $z \in \mathsf{FV}(F(y))$, and so we can apply the induction hypothesis to obtain $\mathcal{M} \Vdash_w \varphi(A(y))[\sigma]$.

Third, if τ is such that $\tau(x) = c^A$ for some $c \in \mathsf{OC}(\forall x A(x))$, then since τ agrees on the value of x and c, it follows that $\mathcal{M} \vDash_w A(c)[\tau]$. Now τ was an x-variant of σ, and so $\mathcal{M} \vDash_w A(c)[\sigma]$ because x does not occur free in $A(c)$. By similar considerations to the previous case, we can apply the induction hypothesis to obtain $\mathcal{M} \Vdash_w \varphi(A(c))[\sigma]$.

Collecting these cases together via the universal generalization over τ, we have that $\mathcal{M} \Vdash_w \forall x \varphi(A(x)) \wedge \bigwedge_{y \in \mathsf{FV}(\forall x A(x))} \varphi(A(y)) \wedge \bigwedge_{c \in \mathsf{OC}(\forall x A(x))} \varphi(A(c))[\sigma]$, which yields $\mathcal{M} \Vdash_w \varphi(\forall x A(x))[\sigma]$

as desired. The reverse direction proceeds analogously to the proof of Proposition 1.8, and so we omit it. $\square$

1.4 Variable Domain Semantics for QMWL

We revise the definition of a QML model to allow for variable domains:

Definition 1.21 (QML Model, Variable Domain). *A* variable domain QML model $\mathcal{M}$ *is given by a quintuple* $\langle W, \mathcal{R}, \mathcal{D}, (D(w))_{w\in W}, \mathcal{I}\rangle$ *where* W *is a set of worlds,* $\mathcal{R} \subseteq W \times W$ *is an accessibility relation between worlds,* $\mathcal{D} \neq \varnothing$ *is an outer domain of objects, each* $D(w) \subseteq \mathcal{D}$ *in the family is the non-empty domain at* w*, and* $\mathcal{I}$ *is an interpretation function assigning to each* n*-ary predicate* P *an extension* $\mathcal{I}(P)(w) \subseteq \mathcal{D}^n$ *at world* w *and to each constant symbol* c *an object* $\mathcal{I}(c) \in \mathcal{D}$*. Again we work with rigid constants, and so we do not take the interpretation of constants to be world-relative.*

For now, we do not require that $\bigcup_{w\in W} D(w) = \mathcal{D}$*, i.e. we permit objects in the outer domain that don't belong to the domain of any possible world.*

Similarly we update the satisfaction clauses:

Definition 1.22. *For both semantics, we do not enforce the being constraint (Williamson, 2013, §4.1), which states that an atomic predication holding of some terms requires that the values of those terms belong to the inner domain of the world of evaluation.*[6] *That is, we do not require that* $\mathcal{I}(P)(w) \subseteq D(w)^n$*.*

The only revision is then in the quantifier clauses:

- $\mathcal{M} \vDash_w \forall x A[\sigma]$ *iff for all* $\tau \sim_x \sigma$ *with* $\tau(x) \in D(w)$*:* $\mathcal{M} \vDash_w A[\tau]$

[6] This constraint finds prior study as "predicate actualism" in Fine (1981).

- $\mathcal{M} \Vdash_w \forall x A[\sigma]$ *iff for all* $\tau \sim_x \sigma$ *with* $\tau(x) \in D(w)$ *and* $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)] \cup \mathsf{CV}^{\mathcal{M}}(A)$*:* $\mathcal{M} \Vdash_w A[\tau]$

Hence we use actualist quantification, where the quantifiers range over only the objects in the domain of the world of evaluation. Note that variable assignments are in general still permitted to assign values outside of the inner domain of a world.

We continue in the full languages of QML⁼ and QMWL with co-denotation, and so proceed directly to strongest intertranslatability propositions.

Proposition 1.23. *Let ψ be as described in Definition 1.17. For all $A \in \mathsf{Fmla}(QMWL)$, for all models $\mathcal{M}$, for all worlds $w \in W(\mathcal{M})$, for all $\mathcal{M}$-assignments σ, we have $\mathcal{M} \Vdash_w A[\sigma]$ iff $\mathcal{M} \vDash_w \psi(A)[\sigma]$.*

Proof. This proceeds analogously to Proposition 1.18. The extra constraint on the quantifier clause is shared between the two semantics, and hence does not interfere with the inductive step. □

So the variable domain semantics do not pose any difficulties for translating from QMWL to QML⁼. However, the reverse translation requires adjustment. To see this, we examine a simple example.

Example 1.24. Consider the formula $\exists x\ x = c$. In the orthodox semantics, $\mathcal{M} \vDash_w \exists x\ x = c[\sigma]$ iff there is $\tau \sim_x \sigma$ with $\tau(x) \in D(w)$ such that $\mathcal{M} \vDash x = c[\tau]$, i.e. $c^{\mathcal{M}} \in D(w)$.

But consider the translation $\varphi(\exists x\ x = c) = \exists x \bot(x, c) \vee c \equiv c$. In the Wittgensteinian semantics, the first disjunct is always false and the second disjunct is always true, not just when $c^{\mathcal{M}} \in D(w)$.

The truth conditions for the original formula amount to 'c exists at w'. Thus one way forward is to amend the translation and express this by employing an existence predicate.

Definition 1.25. *Let the* language of QMWL with existence *($QMWL^{\mathcal{E}}$) be the extension of QMWL by a unary predicate $\mathcal{E}$. Semantically, we have $\mathcal{M} \Vdash_w \mathcal{E}t[\sigma]$ iff $t^{\mathcal{M},\sigma} \in D(w)$, i.e. the object denoted by the term t exists at w.*

We revise the quantifier clause for the φ map from Definition 1.19:

$$\varphi(\forall x A(x)) = \forall x \varphi(A(x)) \wedge \bigwedge_{y \in \mathsf{FV}(\forall x A(x))} (\mathcal{E}y \to \varphi(A(y))) \wedge \bigwedge_{c \in \mathsf{OC}(\forall x A(x))} (\mathcal{E}c \to \varphi(A(c)))$$

We want the following:

Proposition 1.26. *For all $A \in \mathsf{Fmla}(QML^{=})$, for all models $\mathcal{M}$, for all worlds $w \in W(\mathcal{M})$, for all $\mathcal{M}$-assignments σ which are injective on $\mathsf{FV}(A)$ with $\sigma(x) \notin \mathsf{CV}^{\mathcal{M}}(A)$ for each $x \in \mathsf{FV}(A)$, we have $\mathcal{M} \vDash_w A[\sigma]$ iff $\mathcal{M} \Vdash_w \varphi(A)[\sigma]$.*

Proof. We proceed by induction on formulas

The base case and inductive steps for arrow and box proceed as in Proposition 1.20.

For the case of the universal quantifier, we have $\mathcal{M} \vDash_w \forall x A(x)[\sigma]$ iff for all $\tau \sim_x \sigma$ with $\tau(x) \in D(w)$ we have $\mathcal{M} \vDash_w A(x)[\tau]$. Observe that since $\tau(x) \in D(w)$, then if $\tau(x) = c^{\mathcal{M}} \in \mathsf{CV}^{\mathcal{M}}(A)$ then $\mathcal{M} \vDash_w \mathcal{E}c[\sigma]$ (equivalently, under τ). Similarly, if $\tau(x) = \tau(y)$ with y free in $\forall x A(x)$, we have $\mathcal{M} \vDash_w \mathcal{E}y[\sigma]$ (equivalently, under τ, since σ is an x-variant). This case is then analogous to the case in Proposition 1.20. ☐

We've established that QML$^{=}$ and QMWL$^{\mathcal{E}}$ are intertranslatable, and hence have the same expressive power. Now, we'd like to show that some extension to base QWML like the one we've pursued is necessary to establish intertranslatability. To accomplish this, we'll show that the addition of an existence predicate strictly increases the expressive power of QWML. The strategy will be to provide two models which are indistinguishable in base QWML but can be distinguished when an existence predicate is added.

Example 1.27. We work in QMWL without the existence predicate. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be models with outer domain $\omega \cup \{u\}$ and two worlds w_1 and w_2, where ω is the set of natural numbers and $u \notin \omega$ is arbitrary, e.g., the successor ordinal of ω. Let the domains be such that $D_{\mathcal{M}_1}(w_1) = \omega$ and $D_{\mathcal{M}_2}(w_1) = \omega \cup \{u\}$, and $D_{\mathcal{M}_1}(w_2) = D_{\mathcal{M}_2}(w_2) = \omega \cup \{u\}$. Let the accessibility relation in each model be universal.

For each constant c in the signature, let $c^{\mathcal{M}_1} = c^{\mathcal{M}_2} = 0$. For each predicate R in the signature, interpret R as the empty relation at each world in each model.

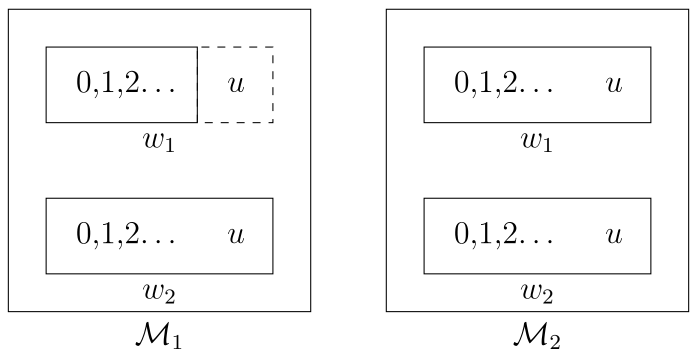

Figure 1.1: The models $\mathcal{M}_1$ and $\mathcal{M}_2$ are indistinguishable in QMWL.

Let $\vec{0}$ be the variable assignment such that $\vec{0}(x) = 0$ for all x. Then we have the following:

Lemma 1.28. *For all QMWL-formulas A, all variable assignments σ, and all choices of indexes i, j, k, l for worlds and models, we have that $\mathcal{M}_i \Vdash_{w_j} A[\sigma]$ iff $\mathcal{M}_k \Vdash_{w_l} A[\vec{0}]$.*

Proof. The base cases follow simply by the interpretations specified in the models. For the inductive step, the cases for arrow and box follow immediately by the induction hypothesis. The case provided below for the universal quantifier is tedious but unilluminating.

For the case of the universal quantifier, suppose that the formula is of the form $\forall x A$. Then $\mathcal{M}_i \Vdash_{w_j} \forall x A[\sigma]$ iff for all $\tau \sim_x \sigma$ with $\tau(x) \in D(w_j)$ and $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)] \cup \mathsf{CV}^{\mathcal{M}}(A)$ we have $\mathcal{M}_i \Vdash_{w_j} A[\tau]$. Applying the induction hypothesis, this holds iff for all $\tau \sim_x \sigma$ with $\tau(x) \in D(w_j)$ and $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)] \cup \mathsf{CV}^{\mathcal{M}}(A)$ we have $\mathcal{M}_k \Vdash_{w_l} A[\vec{0}]$. The quantification over variable assignments is vacuous, so equivalently, this holds iff for all $\tau \sim_x \vec{0}$ with $\tau(x) \in D(w_l)$ and $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)] \cup \mathsf{CV}^{\mathcal{M}}(A)$ we have $\mathcal{M}_k \Vdash_{w_l} A[\vec{0}]$. Applying the induction hypothesis in the reverse direction, this holds iff for all $\tau \sim_x \vec{0}$ with $\tau(x) \in D(w_l)$ and $\tau(x) \notin \tau[\mathsf{FV}(\forall x A)] \cup \mathsf{CV}^{\mathcal{M}}(A)$ we have $\mathcal{M}_k \Vdash_{w_l} A[\tau]$. And this holds iff $\mathcal{M}_k \Vdash_{w_l} \forall x A[\vec{0}]$. □

Proposition 1.29. *For all QMWL-formulas φ and all variable assignments σ, we have that $\mathcal{M}_1 \Vdash_{w_j} \varphi[\sigma]$ iff $\mathcal{M}_2 \Vdash_{w_j} \varphi[\sigma]$ (with $j \in \{1, 2\}$).*

Proof. This follows immediately via two appeals to Lemma 1.28. □

Corollary 1.30. *The extended language QMWL$^{\mathcal{E}}$ is strictly more expressive than QMWL. Likewise, the language for QML$^{=}$ is strictly more expressive than QMWL.*

Proof. The first claim follows from Proposition 1.29 and the observation that $\mathcal{M}_1 \nVdash_{w_2} \forall x \Box \mathcal{E} x$ and $\mathcal{M}_2 \Vdash_{w_2} \forall x \Box \mathcal{E} x$. The second claim was established by Propositions 1.23 and 1.26. □

1.5 Philosophical Costs of the Existence Predicate

In the context of variable domain modal logic, we have found that the W-logician is in need of an existence predicate to adequately express certain propositions which can be articulated in QML$^{=}$.[7] Classically, the existence predicate has been employed to mediate between actualist and possibilist discourse in the framework of quantified modal logic. It is commonly held that the "choice between the possibilist and the actualist quantifier cannot be decided on formal grounds" because the existence predicate permits their intertranslation (Fitting and Mendelsohn, 1998, p. 163). In contrast to the present case, it was the translation from possibilism (constant domains) to actualism (variable domains) which required a primitive existence predicate. The reverse translation did not require this, for the existence predicate

[7]Even if the language is extended in some other way, the proof of Corollary 1.30 reveals that the chosen machinery must capture the essence of an existence predicate.

was definable via identity. If the stipulation of an existence predicate by the W-actualist does represent some philosophical cost, it is in a sense already incurred by the standard actualist by virtue of its definability.

This is the clearest argument against the existence predicate posing a substantial cost for the W-actualist. By parceling out the existential function of the identity sign, the W-logician actually gains some flexibility over the standard logician. If the W-actualist admits an existence predicate, they fare no worse than the standard actualist. But if the W-actualist views the existence predicate as presenting a genuine cost, they can retreat to the more austere language without it. This obviously gives up intertranslatability, but from the current perspective, this would present an advantage since the standard actualist would be committed to the undesirable baggage.

If one still wishes to press the W-actualist on this matter, the remaining available line is try to find a significant difference between a defined existence predicate and one which is primitive or stipulated. To explore this idea, we review the debate about the concept of an existence predicate.[8]

While the nature of existence is a broad philosophical topic with discussions reaching back into antiquity, the question of whether existence is a predicate owes much to the ontological argument for the existence of God. Kant (1998, A598/B626) notably rebuts the ontological argument on the basis that "being is evidently not a real predicate... it is merely the positing of a thing... in the logical use it is merely the copula of judgment", something which does not expand a concept. Given the comparatively restrictive form of logical form provided in the Metaphysical Deduction, we shouldn't read too much into the use of "predicate" over, say, 'property' – what is clear is that Kant argues that existence is not the sort of thing that behaves in a judgment like the concept 'dog' or 'red'.

[8] In this section, we forego an extended discussion of Updike's specific criticisms. While the issues raised are salient, they detract from the primary discussion, and so we address them separately in Appendix A.

Frege, having provided the foundations for modern logic, sharpened this idea. In the *Grundlagen*, he draws an analogy between existence and number, treating existence as a second-level concept. First, Frege (1884, §46) connects the number zero with the denial of existence, writing that to assert "Venus has 0 moons" is to hold that "there simply does not exist any moon or agglomeration of moons for anything to be asserted of". We do not connect a first-level concept with these absent moons; rather, "a property is assigned to the concept 'moon of Venus', namely that of including nothing under it" – hence we assign the second-level concept of zero to the first-level concept 'moon of Venus'. In clarifying the typology of concepts, Frege introduces a similar example before giving a direct commentary on existence:

> In this respect existence is analogous to number. Affirmation of existence is in fact nothing but denial of the number nought. Because existence is a property of concepts the ontological argument for the existence of God breaks down. [...] what is true is, that this can never be so direct a matter as it is to assign some component of a concept as a property to an object falling under it. (Frege, 1884, §53)

So Frege would not endorse an existence predicate which applies to objects. Instead, existential claims about objects are mediated through first-level concepts by assigning or denying the second-level concept of existence (or zero) to those first-level concepts.

This view was influential on later philosophers of logic. Carnap (1932) argues that a predicative reading of *to be* "feigns a predicate where there is none" (pp. 73-74), going on to claim that the quantificational apparatus subsumes this role. This is similarly expressed in Quine's famous dictum "to be is to be the value of a bound variable" (p. 34, 1948). Quine leveraged this idea further in formulating his criterion of ontological commitment. Quine's derision toward 'subsistence' earlier in the paper, together with his general modal skepticism, suggest that he would likewise oppose the notion of existence as a predicate applying to particulars.

However, given that we're already deeply invested in the framework of possible worlds, we should not let this prejudice us too much against an existence predicate. Frege and Quine, for their own reasons, did not dedicate serious consideration to the logical analysis of modal contexts. In light of the modal argument against descriptivist theory of names, let us see what support we can give for an existence predicate.[9]

Recall that we find ourselves in variable domain setting. We use the actualist quantifiers, which locally seem to fulfill Quine's criterion of ontological commitment: at a given world, the quantifiers range over exactly those objects which exist at that world. But the semantics obviously reflect a broader commitment in the outer domain which potentially exceeds the domain of any given world. And when we allow modal operators in the scope of the actualist quantifiers, we can easily escape the domain of the world.[10] For example, the formula $\exists x \Diamond Px$ could be true without the object existing at the witnessing world. In modal formulas, we outstrip the expressivity of a single world. When we consider $\exists x \Diamond \forall y x \neq y$, we import the witness x to a world where it fails to exist, loosely speaking. Similarly, if we help ourselves to an actuality operator for effect, the formula $\Diamond \exists x \mathsf{A} \forall y x \neq y$ has us import that non-existent witness into the actual world. If the Quinean criterion is to hold any water in this setting, our commitment should be to the outer domain rather than merely the inner domain (as a naïve application would suggest).[11] However, quantification in non-modal formulas still effectively ranges over the inner domain. Taken together, these discriminatory capabilities suggest that we in some sense already have access to the information encoded by an existence predicate.

[9] I allude to Kripke (1972).

[10] Quine would of course object to this arrangement. However, his objections were targeted toward a reading of the box as analyticity, which is not our current focus. Nevertheless, it's worth noting that these objections were generally on the mark. See Burgess (1997) for a discussion of this topic.

[11] In Definition 1.21 we were lax in allowing that $\bigcup_{w \in W} D(w) \subsetneq \mathcal{D}$. The criterion could be taken as requiring equality (i.e. forbidding objects which exist at no world), which is the more common formulation of the semantics.

The above reasoning accepts the Quinean view that existential quantification is existence-entailing. This assumption has been challenged in the actualism-possibilism debate. More recently, Parsons (1980, pp. 32-36) argues that objectual quantification is neutral on the basis of inference. Zalta (2012, p. 50-52) provides an argument from natural language (among others) for accepting an existence predicate with the existential quantifier not being existence-entailing. These authors sketch the space for one to accept the existence predicate while maintaining some skepticism over existence as a property (though they generally hold no such skepticism). Others like Hintikka (1969a, p. 33) ultimately view an existence predicate as harmless if "redundant". Others still, like Berto (2013, Ch. 4), likewise view the sort of logical existence predicate under discussion as harmless, though only because they argue existence is actually a stronger, non-logical and genuine property. Hence if the W-actualist errs by admitting an existence predicate, the damage seems to be just a bloated vocabulary.

1.6 Concluding Remarks and Open Questions

In motivating his ontology, Frege (1892, pp. 194-195) notes the pernicious ambiguity of the copular verb 'ist', with its different uses having categorically distinct renditions in formal logic. Russell (1903, p. 64 ft.) labels the readings as "assert[ing] Being", "identity" and "predication".[12] These authors rightfully recognized the danger of failing to attend to these differences by treating them uniformly. While they disentangle the verb 'is' into its different forms, with its function of identity being assigned the identity sign, we find that the identity sign itself is similarly ambiguous. Classical logicians overload the identity predicate with three distinct roles: logical, for manipulating the range of quantification; empirical, for expressing co-reference of names; and ontological, for expressing existence as a property. By moving to W-logic, we carve out the divergent functions of identity into three distinct and unitary devices: quantification, co-denotation and predicative existence.

[12]Wittgenstein (1971, 3.323) recounts this distinction in the *Tractatus* as well.

To finish, we would like to highlight a few areas for future study.

Thus far, we have only examined systems with rigid constants. This is standard for quantified modal logic, but the alternative is worth considering to see if it poses any difficulty for the W-logician. One would expect that an update to the semantics for co-denotation could handle this, though the influence of quantifiers may be nontrivial, especially if they are amended to range over individual concepts.

The discussion of the being constraint and necessitism-contingentism debate reveals the salience of higher-order considerations. Consequently, it would be of interest to explore higher-order (modal) W-logics. In connection with the above, one could also examine intensional logics which already permit non-rigid designation like the Montagovian system IL axiomatized by Gallin (2016) or the variable domain intensional type theory we present in Chapter 3.

Consideration of intensional logics with explicit world parameters motivates another philosophical question: does adopting the exclusive interpretation for objects commit one to the exclusive interpretation for modal quantifiers? In the modal logics considered so far, the choice between inclusive and exclusive quantification over worlds doesn't seem impactful. One reason for this can be seen in the translation of first-order modal logic into classical first-order logic with sorts – the restricted placement of world variables limits the scope effects of modal quantifiers. But in a richer modal system there is greater opportunity for interaction.

Chapter 2

Against Williamson's Argument for Necessitism: Neutralizing the Modalized Geach-Kaplan

2.1 Introduction

Williamson (2010, 2013) provides a defense of necessitism over contingentism, or in the older terminology, a defense of possibilism over actualism. While the contingentist believes that we quantify only over objects that exist at the actual world, the necessitist believes that we quantify over all possible objects. Williamson argues for necessitism on abductive grounds pertaining to expressibility: the necessitist is able to articulate certain distinctions for which there is no acceptable contingentist translation or paraphrase.

The critical point in Williamson's argument is a move from first-order modal logic to second-order modal logic, which thereby increases the power of object-quantification but not of world-quantification: that is, in standard second-order modal logic we can quantify over collections of objects, but there is no second-order analogue of the modal operators which gives us the effect of quantification over collections of worlds. I claim that Williamson's argument for necessitism is an artifact of this imbalance, and that by restoring parity in the strength of the quantifiers, there is no longer any expressive difference between necessitism and contingentism. Specifically, balance is obtained by adding new linguistic resources to the language, in the form of intensional operators with associated quantifiers. I then defend the legitimacy of this extension of the language in the debate. Authors such as Fritz (2013) and Goodman (2016a) have contributed to the development of the positions of necessitism and contingentism. Goodman (2016b) provides one of the first critical evaluations of Williamson's argument, sketching a number of potential objections. In particular, in section §2.5 of the article, Goodman proposes a strategy of adding higher-order quantification. He ultimately takes a pessimistic view of its prospects because he finds that its naïve implementation trivializes the matter by pushing necessitism into the background theory. This chapter shows that the approach of enriching quantification can succeed when we attend carefully to the sorts of quantifiers we add. Hence, the goal is to directly engage and respond to Williamson's central argument for necessitism by means of a thorough and non-trivial formal result.

The outline of the chapter is as follows. In section §2.2, I introduce Williamson's methodology in abstract terms and analyze its application to some simple examples. Section §2.3 introduces the background needed to contextualize and state Williamson's result. Afterward, in section §2.4, I show that the addition of new linguistic resources in the form of intensional operators overturns Williamson's theorem from the prior section, undermining his argument for necessitism. The following sections §2.5 and §2.6 secure the warrant for these additional resources. I conclude in section §2.7 by summarizing the argument against Williamson.

2.2 Williamson on the Structure of the Debate

In order to understand how Williamson obtains his result and evaluate how robust the argument is, we begin by formalizing his methodology. In doing so, we can apply the technique to other debates as a means of testing the method. Should we judge it as sound, we then obtain a new tool for resolving theoretical disputes.

Let us fix a formal language, which for the sake of simplicity we take to only consist of relation symbols. Further, we assume that we have a unary relation 'concrete' in the language, so that some objects are concrete and others are not. We want to consider two theories, T_0 and T_1, which we might think about as either given via axioms in this language or via a class of models in this language, and which are respectively advocated by two agents, whom we call Player 0 and Player 1. Then we say that a *neutral* sentence is one where all quantifiers are bound to the concrete objects (and if there are higher-order quantifiers then all of these are bound to concepts whose instantiations are concrete). Further, let us say that a *Williamson-map* for Player 1 is a function taking any sentence φ in the language to a neutral sentence φ^1 in the language such that Player 1's theory T_1 entails that φ is equivalent to the neutral φ^1. We define a Williamson map for Player 0 similarly by changing subscripts and superscripts from 1 to 0. Finally, we say that *Player 1 wins the Williamson debate* if there is a Williamson-map for Player 0 but not for Player 1; and we say that *Player 0 wins the Williamson debate* if there is a Williamson-map for Player 1 but not for Player 0. Note that it is the player who does *not* admit a Williamson-map that wins the debate.

The idea behind the Williamson methodology is that the two opposing theories should share a common language, with no linguistic disputes.[1] The debate should instead concern deep

[1]Williamson (2013, p. 308) emphasizes that the debate should not be "merely verbal". However, Williamson expresses that he does not intend his methodology to apply to all disputes: he agrees that it would be unreasonable to place such a "demand on *any* theorist to find neutral equivalents for *any* of their opponents' utterances" (Williamson, 2016, p. 648, emphasis mine). Williamson seems to think that the methodology applies primarily to debates in which the interlocutors share a common language where there is a "neutral zone" of uncontested facts. Moreover, the debate should regard the distinctions licensed by one's

(possibly modal) properties of the objects of the domain. However, their present surface characteristics should not be in dispute – the two agree on C and any (non-modal) relations holding among the Cs. By introducing the Williamson-maps, the players have a method for 'distilling down' the theory-laden statements of the opposing player into neutral talk (which the opposing player considers equivalent against their background theory). The claim is then that if only one side permits this reduction, then the opposing side which does not permit such a reduction is at an expressive advantage, which is taken as evidence in support of that opposing theory.

To get a better sense for this methodology let's examine a few examples. For our first example, we work in the language of set theory. The two theories we consider share a common base theory, consisting of the usual axioms of Zermelo-Frankel set theory, except that we omit the axiom of Infinity and instead add the axiom $\forall x(Cx \leftrightarrow$ 'x is a set of finite rank'). A set has finite rank when it is built-up from the empty set in finitely many operations. So our additional axiom states that the concrete objects of set theory are just those that are finitely constructed. Now we consider extending this base theory in incompatible ways. We let T_0 extend the base theory by adding the axiom of Infinity. Hence Player 0 believes that there are also infinite objects in set theory, such as the collection of all finite ordinals (the set of natural numbers, in other words). We let T_1 extend the base theory by adding the negation of the axiom of Infinity. So Player 1 denies that there are any sets beyond the finite. Consider the map on formulas where we restrict predication and quantification to the concrete predicate, e.g. the quantification $\forall x \psi(x)$ is replaced by $\forall x(Cx \rightarrow \psi(x))$. This will be a Williamson-map for Player 1, since their set theoretic universe consisted of just the objects of finite rank, so finite quantification is the same as unbounded quantification by their lights. However, there is no Williamson-map for Player 0. For, if there were, we

language and one's theory (p. 646), and the fatal error is to dismiss distinctions to which one is otherwise "committed to regarding as genuine" (Williamson, 2013, p. 364). For instance, he claims that Goodman's example of the favorite set (of incompossibles) "goes beyond the dialectical set-up of the book" since the concepts under discussion go "beyond the neutral zone" (p. 648). The examples presented here fit the criteria he does identify and are well within the general spirit of the dialectic.

could define a truth predicate for T_0, contradicting Tarski's theorem on the undefinability of truth.[2] Because there is a Williamson-map for Player 1 but not for Player 0, we have that Player 0 wins the Williamson-debate. This is abductive evidence in favor of T_0, set theory with the axiom of Infinity. Thus this example provides a case in which Williamson's method succeeds. That is, the prevailing theory is our preferred one, as most mathematicians and many philosophers accept the axiom of Infinity.

Now we consider an example whose results appear to run counter to the orthodox viewpoint. Here, we work in the language of first-order arithmetic. Again, we will consider to theories which extend a base theory in incompatible ways. We start by defining a formula $\sigma(x) \equiv$ 'x is the shortest proof of $0 = 1$ in PA^1'. We take the base theory to consist of the first-order Peano axioms (the system PA^1) plus the axiom $\forall x(Cx \leftrightarrow \forall y(\sigma(y) \to x < y))$. Intuitively, this axiom stipulates that the concrete numbers are those that come before the number which codes the proof of a contradiction in PA^1, if there is such a number. In other words, we can think of these numbers as the consistent initial segment of PA^1. Now we let T_0 be the extension by $Con(PA^1)$, the consistency statement for PA^1. And we let T_1 be the extension by $\neg Con(PA^1)$. So Player 0 believes that Peano arithmetic is consistent, and Player 1 believes that Peano arithmetic is inconsistent. In terms of our formula $\sigma(x)$ above, Player 0 thinks that all numbers are concrete since there is no proof of a contradiction (because PA^1 is consistent and hence there is no contradiction to be proven); Player 1 thinks that only some of the numbers are concrete and eventually one encounters a number which codes the proof of a contradiction.

As before, our task is now to investigate which of these theories admit Williamson-maps. The map which restricts predication and quantification to the concrete predicate suffices as a Williamson-map for Player 0. This follows by the same reasoning in our set theory example: Player 0 thinks everything is concrete, so the concrete-restricted quantifiers appear the same

[2]Specifically, if there were such a map $\varphi \mapsto (\varphi)^0$, we can define $T(\ulcorner\varphi\urcorner)$ by $V_\omega \vDash \ulcorner\varphi^0\urcorner$. Then we would have that T_0 proves each instance of the following string of biconditionals: $T(\ulcorner\varphi\urcorner) \Leftrightarrow (V_\omega \vDash \ulcorner(\varphi)^0\urcorner) \Leftrightarrow (\varphi)^0 \Leftrightarrow \varphi$.

as the unrestricted ones. For Player 1, it is provable that there is no such Williamson-map. As before, doing so would enable us to define a truth predicate for Peano arithmetic in contradiction of Tarski's theorem.[3] Since there is a Williamson map for Player 0 but not for Player 1, we have that Player 1 wins the Williamson debate. This suggests that we should adopt $\neg Con(PA)$. However, it is widely held that if one accepts PA^1, one is committed to accepting its consistency (see Halbach (2011, §22.1) and Dean (2015)).

At first glance, this casts some doubt on the value of the methodology. But it is important to note that this case differs from the previous one in that here the two theories disagree on the extension of C, i.e. which objects are concrete. It's possible that by refining the technique by restricting how the theories extend the base theory, we will obtain the 'expected' or preferred outcome in the previous example. A guiding principle moving forward is that the methodology is applicable only to disputes in which there is no disagreement regarding which objects are concrete.

We briefly discuss one more example, detailed in Appendix §B.1. We again work in the language of set theory and consider competing views in contemporary philosophy of mathematics on the structure of the set-theoretic universe. The debate regards the axiom of constructibility, which acts essentially as a limit on how rich or complex the set theoretic hierarchy is. In the literature, the dispute is unsettled, but the popular view is to deny the axiom of constructibility.[4] We make three observations. First, in accord with our moral from the previous example, both positions agree on what the concrete objects are, namely, the constructible sets. Second, when working through the technical details, we find that this result does not involve Tarski's undefinability theorem. Third, the outcome of applying

[3]Any candidate Williamson-map $\varphi \mapsto (\varphi)^1$ for Player 1 by definition restricts to the concrete, yielding the following chain of equivalences in T_1: $\varphi \Leftrightarrow (\varphi)^1 \Leftrightarrow \exists\pi\sigma(\pi) \wedge ([0,\ldots,\pi] \vDash \ulcorner\varphi\urcorner)$. Then define the truth predicate $T(\ulcorner\varphi\urcorner)$ by the right-hand side of the equivalence. Since PA^1 can express truth predicates for finite initial segments (Hájek and Pudlák, 1998, Theorem 2.26), we again contradict Tarski's theorem on the undefinability of truth.

[4]For a discussion of the case against constructibility, see Maddy (1997) – Ch. 6 of Part III in particular highlights some of the virtues of the existence of $0^\sharp$. And Koellner (2010) explains some of the virtues of the dissenting view.

the method to this case favors the prevailing theory (the denial of constructibility). Again, this is a live debate in philosophy of mathematics. Hence this shows that the technique has applicability to substantive disputes in philosophy and therefore does more than merely validate uncontested views, as in the toy examples we considered previously.

We close this section by noting that the technique is not always conclusive. It is entirely possible for neither player to win the debate, giving us no evidence to adopt one theory over the other. This happens when both players have Williamson-maps, or when neither player does. For instance, consider a case in which Cx is defined by $x = x$. Here, the condition for a object to be concrete is just a logical truth and hence the content of a player's theory will have no influence on which objects they view as concrete, since every object will be. In this case, it's trivial for both players to have a Williamson-map, since the restriction to concrete objects doesn't change the quantifiers at all. Loosely speaking, in the circumstance where the extension of the concrete predicate is too large, both players will be able to define a Williamson-map. On the other hand, modifying our first example from set theory, consider replacing the axiom that specified the extension of the concrete objects with $\forall x(Cx \leftrightarrow x = \varnothing)$, which states that the empty set is the only concrete object. For either player, having a Williamson-map would mean being able to encode a truth-predicate, in contradiction of Tarski's theorem. So in the case where the extension of the concrete predicate is too narrow, neither player will be able to define a Williamson-map.

2.3 Williamson and the Modalized Geach-Kaplan

Williamson wants to apply this methodology to the debate between necessitists and contingentists. Recall that necessitists hold that all (the same) objects compose the domain of each possible world, with each object being possibly concrete (possibly existing) at each world. Contingentists hold that ontology is contingent, i.e. that while objects may come in and out of being across worlds, there are no merely possible objects. As we saw in §2.2, a perspicuous logical axiomatization is required to apply his methodology, so Williamson begins by formalizing the two positions.

He axiomatizes necessitism with the following (Williamson, 2010, p. 678):

$$\Box\forall x\,\Box\,\exists y(x = y) \wedge \forall x\,\Diamond\,Cx \wedge \bigwedge_R \Box\forall\bar{x}(R\bar{x} \rightarrow \bigwedge_j Cx_j) \tag{Aux[Nec]}$$

The first conjunct is the basic statement of necessitism (NNE): 'necessarily everything is necessarily something' (Williamson, 2013, p. 38). Informally, this captures the idea that entities are persistent and acts as a pseudo-Barcan formula.[5] As a result, models of Aux[Nec] will validate the Barcan formulas. In other words, the quantifiers will have the same range at each world. The second rules out any objects which never obtain. In the last conjunct, R is restricted to non-logical vocabulary – it states that only concrete objects can be related, or similarly that talk of non-concrete (nonexistent) objects must be bound by modal operators. This is what Williamson (2013, pp. 148-149) calls the *being constraint*.

Contingentism is more succinctly captured (Williamson, 2010, p. 688):

$$\Box\forall x Cx \tag{Aux[Con]}$$

This asserts that there are no merely possible objects. In the model theory, for new objects to come into existence, the domain at each world must change (increase in this case). Since the Barcan formula forces non-increasing domains, the contingentist must reject it. In models of Aux[Con], quantification at each world will be independent and possibly differing in range. Figure 2.1 provides an example of how the two models might appear.

[5]See Fitting and Mendelsohn (1998, pp.181-182).

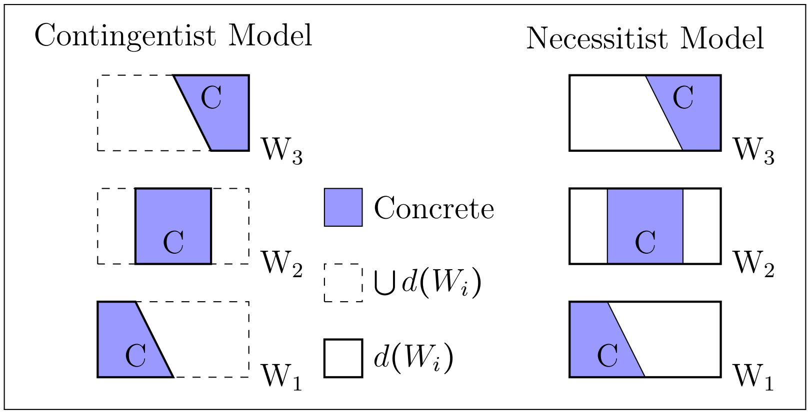

Figure 2.1: Visualizing the debate

As demanded by his methodology, Williamson begins by defining the $(\cdot)^{\mathrm{Con}}$ map, which restricts formulas down to concrete objects:

$$
\begin{aligned}
(Fx_1 \ldots x_n)^{\mathrm{Con}} &= Fx_1 \ldots x_n \wedge Cx_1 \wedge \ldots \wedge Cx_n \\
(\neg A)^{\mathrm{Con}} &= \neg(A)^{\mathrm{Con}} \\
(A \wedge B)^{\mathrm{Con}} &= (A)^{\mathrm{Con}} \wedge (B)^{\mathrm{Con}} \\
(\Diamond A)^{\mathrm{Con}} &= \Diamond(A)^{\mathrm{Con}} \\
(\exists x A)^{\mathrm{Con}} &= \exists x(Cx \wedge (A)^{\mathrm{Con}}) \\
(Xx_1 \ldots x_n)^{\mathrm{Con}} &= Xx_1 \ldots x_n \wedge X \leq C \\
(\exists X A)^{\mathrm{Con}} &= \exists X((A)^{\mathrm{Con}} \wedge X \leq C)
\end{aligned}
$$

In this, $X \leq C$ abbreviates $\forall x_1 \ldots \forall x_n(Xx_1 \ldots x_n \to (Cx_1 \wedge \ldots \wedge Cx_n))$. As one can see, the $(\cdot)^{\mathrm{Con}}$ map is a second-order extension of the classic map used to translate between possibilist and actualist quantification (Fitting and Mendelsohn, 1998, p. 106).

Again, the reason we want to restrict down to concrete objects is because the debate is supposed to center around the modal properties of objects. Which objects are concrete, while potentially a question of philosophical import, is not at stake here: the necessitist and the contingentist do not in principle dispute the extension of C at any world. As Williamson (2013, p. 367) writes: 'The contingentist denies what the necessitist affirms *in a common language*' (emphasis mine).

Since the interlocutors have a baseline of agreement, the goal of the Williamson-maps is to convert arbitrary talk to that which concerns just the uncontentious neutral portion of the domain. We say a formula φ is *neutral* if there is some formula ψ such that $\vDash \varphi \leftrightarrow (\psi)^{\mathrm{Con}}$, i.e. $(\psi)^{\mathrm{Con}}$ is true in exactly the same models as φ. Note that $(\psi)^{\mathrm{Con}}$ is neutral trivially, for the definition reduces to a tautology $\theta \leftrightarrow \theta$. Like the classic possibilism-actualism translation, $(\cdot)^{\mathrm{Con}}$ should allow us to move between different discourses by finding neutral equivalents. When we're dealing with only first-order formulas, Williamson (2010, pp. 730-734) proves that for every formula φ, we have that $\mathrm{Aux}[\mathrm{Con}] \vDash \varphi \leftrightarrow (\varphi)^{\mathrm{Con}}$ and there is neutral ψ such that $\mathrm{Aux}[\mathrm{Nec}] \vDash \varphi \leftrightarrow \psi$.[6] This is a restatement in the our setting of the classic result that possibilist-validities formulated using possibilist quantification are actualist-validities when we relativize the quantifiers appropriately (Fitting and Mendelsohn, 1998, Proposition 4.8.2, p. 106).

So we know that one can always find a neutral middle ground for anything their opponent might assert, at least for first-order expressions. In other words, for first-order quantified modal logic, neither the necessitist nor the contingentist wins the Williamson debate, for each player has a Williamson-map.

This nice correspondence breaks down when we move up to second-order formulas. The sentence Williamson uses to show this fact is a modal version of the Geach-Kaplan sentence, which found wide exposure in Boolos (1998, p. 57): 'Some critics admire only each other'.[7]

[6]The exact statement here is a slight weakening of Williamson's result. He goes on to define a $(\cdot)^{\mathrm{Nec}}$ map so that $\mathrm{Aux}[\mathrm{Nec}] \vDash \varphi \leftrightarrow (\varphi)^{\mathrm{Nec}}$. So both sides have a constructive procedure for translating formulas. See Theorems 1.8 and 1.21 in Williamson (2010).

[7]See Oliver and Smiley (2013) for a detailed history of this example.

This natural language sentence can be formalized as:

$$\exists X \forall x \forall y (\exists z X z \wedge ((X x \wedge A x y) \rightarrow (X y \wedge x \neq y))) \tag{GK}$$

A sentence is *first-orderizable* if it is equivalent across all models to a first-order sentence. Geach and Kaplan's famous result is:

Theorem 2.1. *There is a sentence of second-order logic which is non-first-orderizable. Indeed, the Geach-Kaplan sentence (GK) is non-first-orderizable.*

For the traditional proof, which elegantly uses models of Peano arithmetic, see Appendix §B.2.1.

The sentence Williamson uses is a modal version of the Geach-Kaplan sentence:

$$\exists X (\exists x X x \wedge \exists x \neg X x \wedge \forall x \forall y (\underline{\Diamond R x y \rightarrow (X x \rightarrow X y)})) \tag{MGK}$$

Indeed, compare (MGK) with a minor rearrangement of (GK):

$$\exists X \forall x \forall y (\exists z X z \wedge \underline{A x y \rightarrow ((X x \rightarrow X y)} \wedge (X x \rightarrow x \neq y))) \tag{GK}$$

Williamson then proves:

Theorem 2.2. *There is a sentence φ of second-order quantified modal logic such that there is no neutral ψ with Aux[Nec] $\vDash \varphi \leftrightarrow \psi$. Indeed, (MGK) is such a sentence (Theorem 2.15 in Williamson (2010, p. 743)).*

The proof Williamson (2010) provides passes through a number of small lemmas, and as a result the relation between (GK) and (MGK) is not illuminated. However, see Appendix §B.2.1 for a direct proof which uses the same methods as the traditional proof of Theorem 2.1.

What this means is that the necessitist does not have a Williamson map. But it is true that for every formula φ of second-order quantified modal logic, we have that $\mathrm{Aux[Con]} \vDash \varphi \leftrightarrow (\varphi)^{\mathrm{Con}}$ (Williamson, 2010, Theorem 2.5, p. 738) – in other words, the contingentist has a Williamson-map. Thus we have that the necessitist wins the Williamson debate over second-order quantified modal logic, giving us abductive support in favor of necessitism.

2.4 Neutralizing the Modalized Geach-Kaplan with Additional Resources

We have seen that given Williamson's formalization of the debate, the necessitist wins the debate in the case of second-order quantified modal logic. Taking this methodology to give genuine evidence in support of one theory over another, we see that necessitism becomes the favored metaphysical theory. Results using this technique are sensitive to the formalizations provided. Our goal in this section is to investigate how changing the underlying theories might affect Williamson's result.

We expand the language by adding functions from worlds to subsets of the domain of the frame. We reserve the variables Δ, Γ and Θ for these functions. Our overall aim is to add new resources so that, in the expanded system, Theorem 2.2 does not hold. In particular, that means the language should be such that each sentence has a neutral equivalent from the perspective of the necessitist. Hence we focus on the setting of necessitism. Recall that necessitist models are constant domain. For each such function $\Delta : W \to \mathcal{P}(D)$ we require that $\Delta(w) \subset int(C)(w)$, where $int(C)(w)$ provides the interpretation of C at w, or in other words, the concrete objects at world w. These functions can be interpreted as providing the intensions of properties. We likewise add quantifiers for these entities.[8]

[8] These resources are amenable to the view of *non-domain-inclusion actualism* described in Bennett (2005, p. 302). Section §2.5 presents their general utility in modal metaphysics.

We now investigate whether (MGK) becomes neutralizable in this linguistic framework. We begin by defining a sentence:

$$\begin{aligned}\exists\Delta\Big(&\Diamond\, \exists x(Cx \wedge x \in \check{\Delta}) \\ &\wedge \Diamond\, \exists x(Cx \wedge \Box x \notin \check{\Delta}) \\ &\wedge \uparrow \Box\forall x \in C \downarrow\uparrow \Box\forall y \in C \downarrow (\Diamond Rxy \to (\Diamond x \in \check{\Delta} \to \Diamond y \in \check{\Delta}))\Big)\end{aligned} \tag{INT}$$

This sentence makes use of the Vlach operators $\uparrow$ and $\downarrow$ and the extension operator $\check{}$ from intensional logic.[9] Informally, the $\uparrow$ operator stores the current world of evaluation, and the $\downarrow$ operator recalls that world for the evaluation – this lets one 'shift' the world of evaluation, allowing, e.g., for the object of quantification at one world to be considered at a different world. The extension operator $\check{}$ gives the denotation of an object at the world of evaluation, e.g. $\check{\Delta}$ gives the extension of the intension Δ at a particular world.

In our new setting, it is provable that (MGK) is equivalent to (INT) across models of Aux[Nec]. For ease of readability, the precise statement and proof of this result is presented as Proposition B.1 in Appendix §B.2. So our new expressive resources have allowed us to formulate an equivalent to (MGK). If (INT) were to count as neutral, then Williamson's proof of Theorem 2.2 is now blocked. By our old definition of neutrality, (INT) is not neutral.[10] However, the previous definition involved a map on formulas – we've since expanded our language and hence added new formulas, so we should investigate if we can extend the map in a way which preserves our intuitive notion of neutrality. Doing so could potentially increase the set of neutral formulas and therefore possibly neutralize (MGK) in the new sense of neutrality.

[9]The Vlach operators are employed by Williamson (2010) and find preliminary discussion in Fine (2005).
[10]For if it were, then (MGK) would likewise be neutral in the old sense, contradicting the theorem.

We would like to motivate a revised definition of neutrality which incorporates the new linguistic resources. The technical details are cumbersome, but the central idea is this: we introduce a $(\cdot)^{\mathrm{Nec}}$ map which replaces rigid second-order quantification with quantification over intensions and is otherwise similar to the $(\cdot)^{\mathrm{Con}}$ map.[11] Our neutral equivalent (INT) from before is nothing but $(MGK)^{\mathrm{Nec}}$. With this in mind, we can generalize our reasoning about (INT) and (MGK) and prove that for any sentence A, $\mathrm{Aux}[\mathrm{Nec}] \vDash (A)^{\mathrm{Nec}} \leftrightarrow A$. This shows Williamson's Theorem 2.2 does not generalize to the setting featuring the intensional operators. The precise formulation and statement of the results can be found in Appendix §B.2 (Theorem B.2, Lemma B.3 and Corollary B.4). It remains to be argued that the $(\cdot)^{\mathrm{Nec}}$ map is deserving of the appellation 'neutral', given how the term has been used previously in the debate. We pursue this matter in §2.6.

2.5 The Permissibility of the New Resources

In the previous section, we showed that the expansion of the language by new resources for intensional operators restores expressive parity so that Williamson's result no longer holds. In addition to the issue of neutrality, which we will discuss in the next section, the force of argument against the robustness of Williamson's result further hinges on the permissibility of the new resources. The intensional operators were developed by linguists to capture certain nuances of natural language not expressible in classical first-order logic. Hence the legitimacy of these operators is grounded in natural language.

As an illustration of the use of these resources, we will review some classic examples in Montague grammar (Montague, 1970, 1973), following Dowty et al. (1981). We will build

[11] In the proposal sketched by Goodman (2016b, p. 628), he takes the map to pass over the higher-order quantifiers without effect: $(\forall X^{\langle t_1,\ldots,t_n\rangle} A)^{\mathrm{Con}} = \forall X (A)^{\mathrm{Con}}$. The approach I propose treats the matter of neutrality for the higher-order resources as non-trivial.

up to an intensional rendering of the Geach-Kaplan sentence (GK), which makes natural use of the intensional operators employed in the proofs of the previous section.

The intension operator $^\frown$ gives the intension of an expression (e.g. a proposition), yielding a function from worlds to the denotation of that expression (e.g. the extension of that proposition at that world). This operator is roughly an inverse to the extension operator $_\vee$ introduced earlier.[12]

Intensional operators can be used to express the *de re*/*de dicto* distinction, and our utilization of the operators can be viewed as a particular instance of the distinction. Recall that the distinction is borne out of attributions like 'John believes that Miss America is bald'.[13] In the *de dicto* interpretation, John believes the full proposition 'Miss America is bald'; the particular individual 'Miss America' is not his focus. This is reflected in the formalism by $\mathrm{Bel}(j, {}^\frown B(m))$. Here, j ('John') stands in the believing relation $\mathrm{Bel}(x, p)$ to the proposition ${}^\frown B(m)$ 'Miss America is bald'.

In the *de re* interpretation, John believes of the particular individual known as Miss America that she is bald. Using lambda abstraction, we can formulate the expression as $\lambda x.\mathrm{Bel}(j, {}^\frown B(x))(m)$. Read more perspicuously, this expression says that Miss America is the individual for which John believes the proposition that they are bald. Rendered in this manner, we can quantify out the individual m: $\exists z \lambda x \mathrm{Bel}(j, {}^\frown B(x))(z)$.

Now, in the above examples, the distinction regarded the individual satisfying the proposition. A similar distinction can be seen regarding the property the individual fell under. Consider the sentence 'John believes that Martha is a freemason'. The most natural interpretation is the *de dicto* reading $\mathrm{Bel}(j, {}^\frown F(m))$. Or it could be that John believes *of the freemasons* that Martha is a member. In other words, it could be that John has no vivid understanding of who the freemasons are, but nevertheless believes that Martha is

[12]Their precise relation is described in Theorem 2 (p. 128) and Theorem 4 (p. 130) of Gamut (1991).

[13]This example is due to Dowty et al. (1981, pp. 164-169).

among them. In that case, we have the *de re* reading $\lambda Y.\mathrm{Bel}(j, {}^\frown Y(m))(F)$. In the second-order setting, we can quantify out the second-order object F picking out the freemasons: $\exists Z \lambda Y.\mathrm{Bel}(j, {}^\frown Y(m))(Z)$.

The distinction can also be applied to the extensions of properties. Echoing the Twin Earth thought experiment Putnam (1973), consider the sentence 'John believes that the sample is water'. While Putnam and Kripke famously hold that natural kind terms designate rigidly, some philosophers, such as Chalmers (2006) and Davies and Humberstone (1980), have suggested that there is some facet of the meaning of the term 'water' which possesses certain contingent features. If one wanted to explicate the sentence that way, using standard rigid second-order quantification is inappropriate to capture the meaning of the sentence: the extension of the term 'water' could vary across worlds, i.e. its molecular structure could be H_2O in one world but XYZ in another. Using the devices of the previous section, 'water' would be best treated as a function from worlds to sets of objects. In that case, we would render the *de dicto* reading as $\mathrm{Bel}(j, {}^\frown(\check{\Delta}(m)))$. Then we can distinguish the *de re* formulation as in the previous example and obtain $\lambda\Gamma.\mathrm{Bel}(j, {}^\frown(\check{\Gamma}(m)))(\Delta)$. Once we accept that rendering, we can quantify out, as before: $\exists\Theta\lambda\Gamma.\mathrm{Bel}(j, {}^\frown(\check{\Gamma}(m)))(\Theta)$.

Similar motivations can be found if we analyze the Geach-Kaplan sentence (GK). Suppose we embed it in a belief context: 'Some critics are such that John believes that they admire only each other'. In one context of utterance, it could be that John (say, as a sociologist) comes to this conclusion from study of the behavior of critics and has no particular group in mind, in line with a *de dicto* reading. Or more faithfully to the presented syntax, we have a *de re* interpretation where John could have simply observed a particular group of critics who exhibit this property, formalized as $\exists X \mathrm{Bel}(j, {}^\frown(\forall x \forall y(\exists z Xz \wedge ((Xx \wedge Axy) \to (Xy \wedge x \neq y)))))$. That group of individuals, call them Δ, could shift membership across worlds while still being a group of co-admiring critics. For instance, it could be that John's interests change across worlds: in one scenario, he's interested in theater and so forms his observation about a group of theater critics; in another, he might be interested in literature and so forms his observation about a group of literary critics. The group could also shift membership independent of John's dispositions. It could be that at various worlds, the opportunity cost for becoming a critic is lower, and hence that necessary group of co-admiring critics is larger in size. So we might render the sentence using the intensional objects to obtain $\lambda\Gamma.\mathrm{Bel}(j, {}^\frown(\forall x \forall y(\exists z \check{\Gamma}(z) \wedge ((\check{\Gamma}(x) \wedge Axy) \rightarrow (\check{\Gamma}(y) \wedge x \neq y))))))(\Delta)$. To be more faithful to the original sentence we can remove the name Δ by quantifying out as before. This gives:

$$\exists\Theta\lambda\Gamma.\mathrm{Bel}(j, {}^\frown(\forall x \forall y(\exists z \check{\Gamma}(z) \wedge ((\check{\Gamma}(x) \wedge Axy) \rightarrow (\check{\Gamma}(y) \wedge x \neq y))))))(\Theta) \tag{2.1}$$

This last rendering provides the correct semantics for the intensional Geach-Kaplan sentence. The formalization in (2.1) makes crucial use of the intensional resources introduced in §2.4.

The above natural language examples motivate the utility of the intensional operators. In particular, a correct formalization of the intensional Geach-Kaplan requires them. However, one concern the contingentist might have is that these resources outstrip the linguistic resources available to them, allowing one to draw distinctions beyond what the contingentist believes possible. But to evaluate this we would have to go beyond the austere version of contingentism expressed in Aux[Con] and rather study the robust developments of the contingentist perspective, such as Stalnaker and Plantinga, and so we leave this to future work.[14]

[14]Recent work by Fritz and Goodman (2017) studies formal developments of contingentism using related but distinct logical devices. They examine the debate as extended by generalized quantification and infinitary formulas. Evidence suggests that generalized quantification occurs in natural language (see Barwise and Cooper (1981)), but the intensional operators proposed here clearly fulfill the goal of increasing the power of world-quantification while still having linguistic legitimacy (and the case is worse for infinitary formulas). However, their philosophical commentary (in particular section §3) still sheds some light on the issue at hand.

2.6 The Neutrality of the New Resources

As mentioned previously, in addition to the permissibility of the resources, we also need to argue that their usage is neutral. One way to see this is to appeal to a model theoretic definition of neutrality given by Fritz (2013, p. 654):

> "To make this formally precise, define two Kripke models $M, M' \in \mathsf{P}$ [the models satisfying the *being constraint*] to *[concrete]-coincide* if they have the same set of worlds W, the same actual world, and their interpretation functions i and i' agree on the [concrete] things... Now we can define a sentence in $L_{\Box\mathbb{Q}}$ to be *neutral* if for any two [concrete]-coinciding Kripke models $M, M' \in \mathsf{P}$, $M \vDash \varphi$ if and only if $M' \vDash \varphi$."[15]

Now, Fritz introduces this semantic criterion for neutrality because he is expanding the language to extend Williamson's results to generalized quantifiers. Williamson's syntactic criteria (the $(\cdot)^{\mathrm{Con}}$ and $(\cdot)^{\mathrm{Nec}}$ maps) are sensitive to changes to the language – any addition or alteration requires one to explain the application of the map to expressions containing the new resources, as we did in §2.4. The advantage of a model-theoretic criterion is that it can be applied uniformly to different languages.

In the context of the debate, the concept of Fritz-neutrality also helps to clarify what exactly is meant by neutrality: the neutral sentences are those which have fixed truth-value when one restricts attention to just the concrete objects (i.e. to the potential exclusion of the non-concrete). Dually, if a sentence is non-neutral and hence carries metaphysical weight, then it must be that fixing the concrete part of the model is insufficient to determine the sentence's truth-value. This is a type of supervenience, where a sentence's status as neutral supervenes on facts about concrete objects.

[15] To make terminology consistent, I have replaced the instances of Fritz's preferred term "chunky" with "concrete".

From this perspective, it seems natural to expect that intensional resources are Fritz-neutral because they are reifications of meaning, and as such might be naturally seen to supervene on the concrete. Further, we can prove that the new linguistic devices are Fritz-neutral for sentences (Proposition B.5 in Appendix §B.2). In particular, any sentence $(A)^{\mathrm{Nec}}$ which the necessitist views as equivalent to A will be Fritz-neutral. This completes the task of establishing that every sentence has a neutral equivalent in the eyes of the necessitist in the richer linguistic setting.

2.7 Conclusion

Williamson was correct to identify a disparity between the expressive power of necessitism and contingentism in standard second-order quantified modal logic. However, this expressive imbalance is rooted in fundamental imbalance in object-quantification over world-quantification in the formal framework. By equalizing the power of quantification through the addition of resources from intensional logic, I show the expressive mismatch to be tenuous. Williamson (2013) reflects on the methodology, noting that "the robustness of such arguments should be tested by extending them to a variety of logical settings" (p. 364). We have hence undermined Williamson's argument for necessitism by the exact means by which he believes it should be evaluated.

Chapter 3

A Variable Domain Intensional Type Theory

3.1 Introduction

The traditional intensional theory of types due to Montague uses a constant first-order domain of individuals D_e and a constant, fixed set of indexes I (worlds, times, etc.) and builds a natural type-structure on top of this, where the indexes are hidden from quantification in the object-language.[1] While suitable for many purposes, such constancy has two disadvantages. First, the use of a constant first-order domain validates the Barcan formula, which is ill-motivated on many understandings of the modality. Second, the use of a single set of indexes makes the domain of propositions – as well as the domain of the other intensional types – invariant across indexes, whereas one might rather have had the intuition that propositions too can be contingent.

[1]Standard presentations of this include Gamut (1991, Chapters 4-6) and Dowty et al. (1981). See Gamut (1991, p. 146) and Dowty et al. (1981, Preface) for how these systems relate to Montague's original papers Montague (1974).

The natural and well-understood manner of falsifying Barcan in the first-order setting is to allow the domain of the individuals to vary with the indexes, and hence to use a collection of inner domains $\{D_{i,e} : i \in I\}$ in addition to an outer domain D_e of individuals. In this, i is an index from I and e is the type reserved for first-order individuals. The natural way of falsifying Barcan in the first-order setting is to let the inner-domains $D_{i,e}$ be small subsets of the outer-domain D_e, whose size and members vary with the index i. The inclusion $D_{i,e} \subseteq D_e$ is important for the semantics since it allows variables to be assigned to the outer domain D_e, while at index i the quantifiers only range over the values assigned to the inner domain $D_{i,e}$ itself. If there were not this inclusion $D_{i,e} \subseteq D_e$, there would be no clear choice of where to assign the variables.

However, there is a "mismatch" problem which prevents the obvious exportation of this solution from the type of individuals to higher-types. For, recall that in the type-theoretic setting, one has types not only for first-order entities of type e, but also for truth-values t, as well as functional types $\langle a, b\rangle$ associated to maps from entities of type a to entities of type b. Now, associated to the type $\langle e, e\rangle$ of functions from individuals to individuals, the natural idea would be to define $D_{i,\langle e,e\rangle} = \{f : D_{i,e} \to D_{i,e}\}$. The problem then is: how does one conceive of the higher-type outer-domain $D_{\langle e,e\rangle}$? If one sets $D_{\langle e,e\rangle} = \{F : D_e \to D_e\}$, then one does not have $D_{i,\langle e,e\rangle} \subseteq D_{\langle e,e\rangle}$. This is because a function from the smaller set $D_{i,e}$ to itself is not a function from the larger set D_e to itself. That is, there is a "mismatch" between the input sets of these inner-domain and outer-domain functions.

A natural solution quickly comes to mind, beginning with the observation that every function f from $D_{i,e}$ to itself *extends* to a function $\widetilde{f}$ from D_e to itself. Hence, this suggests that instead of having $D_{i,a} \subseteq D_a$ for each type a, we should merely insist on their being "transfer" maps $T_{i,a} : D_{i,a} \to D_a$, which perhaps are injective and which perhaps agree when there is overlap between the indexes. If we do this for the ground-level type e, then at the higher-level type $\langle a, b\rangle$ we simply lift from f to $\tilde{f}$ so that the diagram in Figure 3.1 commutes.

$$\begin{array}{ccc} D_{i,a} & \xrightarrow{T_{i,a}} & D_a \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle \widetilde{f}} \\ D_{i,b} & \xrightarrow{T_{i,b}} & D_b \end{array}$$

Figure 3.1: Lifting the functional types

To say that the diagram commutes is simply to say that $\widetilde{f}(T_{i,a}(x)) = T_{i,b}(f(x))$ for all x from $D_{i,a}$. And this is just to say that to each predication $f(x) = y$ in the inner-domain there is a corresponding predication $\widetilde{f}(T_{i,a}(x)) = T_{i,b}(y)$ in the outer-domain. However, one might want to lift f to $\widetilde{f}$ in a different way than one lifts g to $\widetilde{g}$, so long as the diagram commutes; or, similarly, one might want sometimes to lift f to $\widetilde{f}$, and other times lift f to a distinct way of defining $\widetilde{f}$, so long as the diagram commutes. Hence, this suggests making the map f to $\widetilde{f}$ part of the definition of a model, which one may supervaluate upon.

Our idea is thus to extend from inner-domains to outer-domains at the higher-type levels by making choices about how to map from the former to the latter. The choices one makes in selecting representatives for equivalence classes can be conceived of as such a process of extension. For instance, suppose that $D_{i,e}$ is the set of equivalence classes ${}^{D_e}\!/_{E_i} = \{[x]_{E_i} : x \in D_e\}$ on D_e by an equivalence relation E_i. To use Frege's famous projective geometry example, suppose that D_e contains straight lines in the plane, and suppose that dE_id' holds iff d, d' are parallel, so that $D_{i,e}$ can be conceived as the set of directions.[2] Perhaps other indexes j might involve equivalence relations E_j which are coarser (e.g. which discretize to only finitely many directions) or which are finer (e.g. which build-in orientations). In such settings, the injective transfer maps $T_{i,e} : D_{e,i} \to D_e$ occur as right inverses to the natural projections $\pi_{i,e} : D_e \to D_{e,i}$ given by $\pi_{i,e}(x) = [x]_{E_i}$. For instance, it is natural to pick $T_{i,e}$ to be the line which goes through the origin, so that the image of $D_{i,e}$ under $T_{i,e}$ resembles the face of a compass. Further, there is often a natural choice for how to lift from f to $\widetilde{f}$.

[2]See Frege (1884, §64 ff.).

For instance, if f of type $\langle e, e\rangle$ rotates a direction 45° clockwise, then it would be natural to have $\widetilde{f}$ mimic this behavior on straight lines.

This idea about how to abstractly model the contingency of "what there is" mirrors Stalnaker's idea for how to abstractly model the contingency of "what can be said." Stalnaker (2012, Appendix A) associates to each index i an equivalence relation E_i on the set of indexes I, and then he takes the propositions at i to be subsets of ${}^{I}\!/_{E_i}$, where this is the set of equivalence classes. Stalnaker explains: "One point represents the actual world, and the cells of the partition that is induced by its equivalence relation represent the maximal consistent propositions. The other points represent realizations of various counterfactual possibilities, and the partition cells that are induced by their equivalence relations represent the maximal consistent propositions that would exist if they were realized." (p. 62). The intuition behind this is that $w_1 E_i w_2$ iff w_1, w_2 are the same from the perspective of i, and so at this index nothing can be said which distinguishes w_1 from w_2. The challenge for how to import this Stalnakerian idea into the type-theoretic setting is to decide how the equivalence $w_1 E_i w_2$ interacts with the individuals in w_1 and w_2. One approach would be to require that the image of $D_{w_1,e}$ in the outer-domain under the injection $T_{w_1,e}$ be the same as the image of $D_{w_2,e}$ in the outer-domain under the injection $T_{w_2,e}$. Then, nothing can be said which distinguishes the two indexes since they are identical from the perspective of the outer-domain.

In what follows, we systematically develop the semantics adumbrated above. The end result is a system which is conservative over previous semantics, in several respects. First, the traditional intensional theory of types of Montague is a special case of our semantics, when there is no variance between the indexes. Hence the object-language of our semantics is the same as that of Montague, with two minor exceptions. First, as is now common, we build in a basic type for degrees d to handle adjectival expressions. Second, we add the Vlach operators, which is a now-familiar way to emulate the effect of certain restricted forms of index quantification. Practically, this requires that we evaluate not at a single index, but at a finite sequence of indexes.

It is worth highlighting one over-riding idea which we pursue throughout the chapter: namely, the transfer maps $T_{i,a} : D_{i,a} \to D_a$ from the inner-domains to the outer-domains are the natural higher-order analogues of the existence predicates from ordinary first-order variable domain modal logic. Hence, familiar questions arise about the extent to which the existence predicates – together with the Vlach operators – have the effect of full quantification over the outer domain or over the indexes. In addition to the philosophical interest in whether actualist and possibilist quantifiers are mere notational variants of one another, these questions speak to whether higher-order modal logic is really a distinctive logic or whether it is rather a baroque rendition of simple type theory with a world-type.

This chapter is organized as follows. In §3.2 we describe the general way of extending from f to $\widetilde{f}$, which amounts to a method of defining the transfer maps $T_{i,a} : D_{i,a} \to D_a$ for all higher-types a. In §3.3 we set out the formal specification of the object-language and define the semantics. In §3.5 we examine the extent to which Stalnaker's proposal for modeling contingent propositions can be formalized our setting. In sections §§3.6-3.7, we examine two related constructions of models: one associated to direct limits and one associated to inverse limits. In §3.6, we look at Kratzerian situation semantics in terms of a direct-limit variation of semantics. In §3.7, we look at van der Does and van Lambalgen's logic of perception based on inverse limits, and show how no great departure from the Montagovian tradition is required to handle this, once the tradition is modified to permit for variable domains.

3.2 Frames

3.2.1 Types

The *types* are defined inductively as follows:

- e, t, d are types
- if a, b are types, then $\langle a, b\rangle$ is a type
- if b is a type, then $\langle s, b\rangle$ is a type.

We further say that a is an *extended type* if a is a type or a is s. Hence, while s is not a type, it is an extended type. Finally, we say that e, t, d, s are the *basic extended types*. (Note that e.g. $\langle s, s\rangle$ and $\langle e, s\rangle$ are not even extended types. The extended types correspond to the subscripts a on our domains $D_{u,a}$ which we develop in subsequent sections.)

As in the intensional theory of types, e is reserved for first-order objects, t is reserved for truth-values, and s is reserved worlds or times – we use the neutral term "indexes" as a stand-in for one of these. Similarly, d is reserved for e.g. scales for heights or weights, and we use the neutral term "degrees" as a stand-in for one of these. It is obvious how to extend the system presented here to one in which there were types for both heights and weights, or one in which there were extended types for both worlds and times. But, for the sake of simplicity, we restrict attention in what follows to the four basic extended types e, t, d, s.

Sometimes in what follows, we need the notion of the *degree* of an extended type, which is defined recursively by

$$\deg(a) = 1 \text{ if } a \text{ is a basic extended type}, \quad \deg(\langle a, b\rangle) = \max(\deg(a), \deg(b)) + 1$$

Note that to be well-defined, this requires that if X is the set of extended types, then $X^2 \cap X$ is empty. An easy way to enforce this is to assume that e, t, d, s have the same set-theoretic rank and to assume the ordered pairing operation increases rank.

Finally, sometimes in what follows we use a natural strict partial order $<$ and partial order $\leq$ on extended types. The strict partial order $<$ is the transitive closure of: $a < \langle a, b \rangle$ and $b < \langle a, b \rangle$ and $s < \langle s, b \rangle$ and $b < \langle s, b \rangle$. Finally, the partial order $\leq$ is just defined by $c \leq d$ iff $c < d$ or $c = d$. Note that e, t, s, d are incomparable under these orders. We define $\mathrm{pred}_{\leq}(c)$ to be the set of basic extended types a such that $a \leq c$. A quick induction shows that if $\deg(c) = n$ then this set has size $\leq 2^n - 1$. The idea behind the order is that the components of the frames and models of types $< c$ are precisely what one needs to construct the components of type c.

3.2.2 Definition of frames, notation

The basic starting point of our semantics is the following definition:

A *base frame* $\langle I, B, D, S \rangle$ is given by a family of maps and non-empty sets $S_{i,a} : B_{i,a} \to D_a$ for each i in I and basic extended type a such that:

- $S_{i,e} : B_{i,e} \to D_e$ is a function.
- $S_{i,t} : B_{i,t} \to D_t$ is the identity function on the set $B_{i,t} = D_t = \{0, 1\}$
- $S_{i,d} : B_{i,d} \to D_d$ is a linear-order homomorphism between linear orders $B_{i,d}$ and D_d, whose ordering we write as $\leq_{i,d}$ and $\leq_d$, respectively.
- $S_{i,s} : B_{i,s} \to D_s$ is function and $D_s = I$
- D_e, D_t, D_d, D_s are four distinct sets.

To aid in describing base frames $\langle I, B, D, S\rangle$, we keep in mind the mnemonics which are indicative of their intended interpretation. Elements i , j, k, u, v, w of I are indexes, which in paradigmatic cases are worlds or times or indexes of models. The $B_{i,e}$-sets are the inner-domains associated to index i, and the D_e set is the outer-domain. The mnemonic is that 'D' reminds us of 'domain' and 'B' reminds us of 'basic domain' and is an alphabetic predecessor of 'D' (the letter 'C' is obviously reserved for constants, which we will introduce later). Elements of $B_{i,e}, D_e$ are typically written with variables x, y, z. Part of what is distinctive about our approach is that we extend the inner-outer distinction to all basic extended types, and we develop "transfer" maps to send the inner domains to the outer domains. We chose 'S' for the name of the transfer maps on the base frames because it is the alphabetic predecessor of 'T', and as will become clear in a few paragraphs, we use S to define T.

Before turning to the definition of frame, we make four brief remarks on the definition of base frame:

- We do not build any accessibility relation into the base frames *per se*, and only add this in later in §3.3.3 when covering the semantics. That said, we chose the word "frame" because it plays much the same role in our semantics here as worlds plus accessibility relations play in propositional modal logic: together with a valuation, it is what determines the notion of a model.

- In applications were the degrees are not relevant, we simply set $S_{i,d} = S_{i,t}$ and $B_{i,d} = D_d = \{0, 1\}$ with the natural ordering. That is, we just treat the domains associated to type d as a redundant copy of the domains associated to type t, modulo renaming for disjointness.

- To avoid certain set-theoretic pathologies, we assume that there is an ordinal rank β such that for all indexes i and all basic extended types a and all elements u of $B_{i,x} \cup D_x$

we have $rank(u) = \beta$. This precludes $B_{i,x} \in B_{j,y}$ and similar things like a function $f : B_{i,x} \to B_{i,y}$ being an element of $B_{j,z}$. Practically, this is just like treating the elements of the B-sets and the D-sets as urelemente. The requirement that D_e, D_t, D_d, D_s are four distinct sets plays a similar role.

- It might be natural to replace $\{0,1\}$ by other complete Boolean algebras (or complete distributive lattices) and to make $S_{i,t} : B_{i,t} \to D_t$ be homomorphisms of these structures, but we do not pursue that here.

Here is then the fundamental notion of frame:

A *frame* $\langle I, B, D, S, T\rangle$ is given by a base frame $\langle I, B, D, S\rangle$ and a family of maps and non-empty sets $T_{i,a} : B_{i,a} \to D_a$ for each index i and extended type a such that:

- $T_{i,a} = S_{i,a}$ if a is a basic extended type.

- If a, b are types, then every element of $B_{i,\langle a,b\rangle}$ is a function $f : B_{i,a} \to B_{i,b}$ such that for all x_0, x_1 from $B_{i,a}$:

$$T_{i,a}(x_0) = T_{i,a}(x_1) \text{ implies } T_{i,b}(f(x_0)) = T_{i,b}(f(x_1)) \qquad (\textit{frame:1})$$

  Further, every element of $D_{\langle a,b\rangle}$ is a function from D_a to D_b, and the map $T_{i,\langle a,b\rangle} : B_{i,\langle a,b\rangle} \to D_{\langle a,b\rangle}$ makes the following diagram commute:

$$\begin{tikzcd} B_{i,a} \arrow[r, "T_{i,a}"] \arrow[d, "f"'] & D_a \arrow[d, "T_{i,\langle a,b\rangle}(f)"] \\ B_{i,b} \arrow[r, "T_{i,b}"] & D_b \end{tikzcd} \qquad (\textit{frame:2})$$

- If b is a type then every element of $B_{i,\langle s,b\rangle}$ is a function $f : B_{i,s} \to D_b$ such that for all x_0, x_1 from $B_{i,s}$:

$$T_{i,s}(x_0) = T_{i,s}(x_1) \text{ implies } f(x_0) = f(x_1) \qquad (\textit{frame:1s})$$

  Further every element of $D_{\langle s,b\rangle}$ is a function from D_s to D_b, and the map $T_{i,\langle s,b\rangle} : B_{i,\langle s,b\rangle} \to D_{\langle s,b\rangle}$ makes the following diagram commute:

$$\begin{tikzcd} B_{i,s} \arrow[r, "T_{i,s}"] \arrow[dr, "f"'] & D_s \arrow[d, "T_{i,\langle s,b\rangle}(f)"] \\ & D_b \end{tikzcd} \qquad (\textit{frame:2s})$$

The condition (*frame:2*) is the formal version of the diagram from §3.1. The reason why (*frame:2s*) is a triangle and not a square like (*frame:2*) can be illustrated by considering the case of $b = e$, and suppose that D_e is intended to model persons. In this case, the intension associated to a role like "the treasurer" would have type $\langle s, e\rangle$, and since we want to allow the role to be unoccupied at a world of evaluation, and hence we would want to map directly to the outer domain rather than to an intermediary inner domain.

The following are some properties which frames may or may not have:

- *constant*: $T_{i,a}$ are surjective and are the identity on their domain.

- *full*: if $\langle a, b\rangle$ is a type, then $B_{i,\langle a,b\rangle}$ is the entire space of functions from $B_{i,a}$ to $B_{i,b}$ and likewise for $D_{\langle a,b\rangle}$.

- *minimal*: if $\langle a, b\rangle$ is a type and f, g in $B_{i,\langle a,b\rangle}$ are such that $T_{i,\langle a,b\rangle}(f)$ and $T_{i,\langle a,b\rangle}(g)$ agree on $T_{i,a}(B_{i,a})$, then they are identical; and likewise if b is a type and f, g in $B_{i,\langle s,b\rangle}$ are such that $T_{i,\langle s,b\rangle}(f)$ and $T_{i,\langle s,b\rangle}(g)$ agree on $T_{i,s}(B_{i,s})$, then they are identical.

- *strongly minimal*: if $\langle a,b\rangle$ is a type and f in $B_{i,\langle a,b\rangle}$ and g in $B_{j,\langle a,b\rangle}$ are such that $T_{i,b}(f(B_{i,a})) = T_{j,b}(g(B_{j,a}))$ and such that $T_{i,\langle a,b\rangle}(f), T_{j,\langle a,b\rangle}(g)$ agree on $T_{i,a}(B_{i,a}) \cup T_{j,a}(B_{j,a})$, then $T_{i,\langle a,b\rangle}(f) = T_{j,\langle a,b\rangle}(g)$; and likewise if b is a type and f in $B_{i,\langle s,b\rangle}$ and g in $B_{j,\langle s,b\rangle}$ are such that $f(B_{i,s}) = g(B_{j,s})$ and $T_{i,\langle s,b\rangle}(f), T_{j,\langle s,b\rangle}(g)$ agree on $T_{i,s}(B_{i,s}) \cup T_{j,s}(B_{j,s})$, then $T_{i,\langle s,b\rangle}(f) = T_{j,\langle s,b\rangle}(g)$.

- *inclusive*: $T_{i,a}$ are the identity on their domain.

- *injective*: $T_{i,a}$ are injective.

- *cohesive*: $T_{i,a} \restriction (B_{i,a} \cap B_{j,b}) = T_{j,b} \restriction (B_{i,a} \cap B_{j,b})$; further, if $B_{i,a} = B_{j,b}$ then $D_a = D_b$.

- *bounded*: if f is in $B_{i,\langle a,b\rangle}$ then $(T_{i,\langle a,b\rangle}(f))(D_a) \subseteq T_{i,b}(f(B_{i,a}))$; likewise f is in $B_{i,\langle s,b\rangle}$ then $(T_{i,\langle s,b\rangle}(f))(D_s) \subseteq f(B_{i,s})$.

Obviously one can relativize these properties to certain indexes or types. And obviously base frames can have these same properties, but with the maps T replaced by the maps S.

As the names suggest, strongly minimal implies minimal:

Proof. Suppose that strongly minimal is satisfied. To verify minimality, suppose $\langle a,b\rangle$ is a type and f,g in $B_{i,\langle a,b\rangle}$ are such that $T_{i,\langle a,b\rangle}(f)$ and $T_{i,\langle a,b\rangle}(g)$ agree on $T_{i,a}(B_{i,a})$. This agreement and two applications of (*frame:2*) give that for all x from $B_{i,a}$ one has: $T_{i,b}(f(x)) = (T_{i,\langle a,b\rangle}f)(T_{i,a}(x)) = (T_{i,\langle a,b\rangle}g)(T_{i,a}(x)) = T_{i,b}(g(x))$. From this follows that $T_{i,b}(f(B_{i,a})) = T_{j,b}(g(B_{j,a}))$. By strong minimality we then have that $T_{i,\langle a,b\rangle}(f)$ and $T_{i,\langle a,b\rangle}(g)$ are identical. The same argument works for the type $\langle s,b\rangle$. □

3.2.3 Extending from base frames to frames

The following proposition tells us that we can always extend base frames to full frames, and tells us what properties are preserved in this extension:

Proposition 3.1.

- *Every base frame can be extended to a full strongly minimal bounded frame.*
- *If the base frame is injective (resp. and/or cohesive, constant), then it can be extended to an injective (resp. and/or cohesive, constant) full strongly minimal bounded frame frame.*
- *But if the base frame is inclusive and but not constant, then it cannot be extended to an inclusive frame.*

Proof. For the first item in the proposition, we build the frame in ω-many stages, taking care of extended types of degree n at stage n of the construction. As we go along, we construct a well-order $\leq_n$ on $D_n = \bigcup_{\deg(b)\leq n} D_b$ such that $m > n$ implies that $(D_n, \leq_n)$ is a proper initial segment of $(D_m, \leq_m)$.

At stage $n = 1$, we simply set $T_{i,a} = S_{i,a}$ for all extended types a of degree n, and we choose an arbitrary well-order $\leq_1$ on D_1. (Note that $\leq_1$ is distinct from the linear orders $\leq_d$ associated to type d). Suppose that stage n has been completed.

For stage $n+1$, first consider $\langle a, b\rangle$ where a, b are types of degree $\leq n$. Then we define $B_{i,\langle a,b\rangle}$ to be the set of all functions $f : B_{i,a} \to B_{i,b}$ such that for all x_0, x_1 from $B_{i,a}$, if $T_{i,a}(x_0) = T_{i,a}(x_1)$ then $T_{i,b}(f(x_0)) = T_{i,b}(f(x_1))$. (Note that $B_{i,\langle a,b\rangle}$ is non-empty: given any point y of $B_{i,b}$ the constant function which sends everything to y is an element of $B_{i,\langle a,b\rangle}$). Further, we define $D_{\langle a,b\rangle}$ to be the set of all functions from D_a to D_b. Given an element f of $B_{i,\langle a,b\rangle}$, define $T_{i,\langle a,b\rangle}(f)$ as follows:

$$T_{i,\langle a,b\rangle}(f)(y) = \begin{cases} T_{i,b}(f(x)) & \text{if } y = T_{i,a}(x) \text{ for some } x \text{ in } B_{i,a}, \\ \min_{\leq_n}(T_{i,b}(f(B_{i,a})) & \text{otherwise.} \end{cases} \tag{3.1}$$

On its first case break, this is well-defined by the constraints placed on f as a member of $B_{i,\langle a,b\rangle}$. And this definition in the first case break suffices to satisfy the commutative condition. On its second case break, note that $f(B_{i,a})$ is a non-empty subset of $B_{i,b}$, and hence $T_{i,b}(f(B_{i,a}))$ is a non-empty subset of D_b, which in turn is subset of D_n, and thus $\min_{\leq_n}(T_{i,b}(f(B_{i,a}))$ picks out an element of $T_{i,b}(f(B_{i,a}))$ and hence D_b. Also, obviously this construction is strongly minimal and bounded.

To continue the stage $n+1$ construction, consider $\langle s,b\rangle$ where b is a type of degree $\leq n$. Then we define $B_{i,\langle s,b\rangle}$ to be the set of all functions $f: B_{i,s} \to D_b$ such that for all x_0, x_1 from $B_{i,s}$, if $T_{i,s}(x_0) = T_{i,s}(x_1)$ then $f(x_0) = f(x_1)$. (Again, $B_{i,\langle s,b\rangle}$ is non-empty: given any point y of D_b the constant function which sends everything to y is an element of $B_{i,\langle s,b\rangle}$). Further, we define $D_{\langle s,b\rangle}$ to be the set of all functions from D_s to D_b. Given an element f of $B_{i,\langle s,b\rangle}$, define $T_{i,\langle s,b\rangle}(f)$ as follows:

$$T_{i,\langle s,b\rangle}(f)(y) = \begin{cases} f(x) & \text{if } y = T_{i,s}(x) \text{ for some } x \text{ in } B_{i,s}, \\ \min_{\leq_n}(f(B_{i,s})) & \text{otherwise.} \end{cases} \tag{3.2}$$

On its first case break, this is well-defined by the constraints placed on f as a member of $B_{i,\langle s,b\rangle}$. And this definition in the first case break suffices to satisfy the commutative condition. On its second case break, note that $f(B_{i,s})$ is a non-empty subset of D_b, which in turn is subset of D_n, and thus $\min_{\leq_n}(f(B_{i,s}))$ picks out an element of D_b. Likewise, this construction is evidentally strongly minimal and bounded.

To finish the stage $n+1$ construction, we merely set $\leq_{n+1}$ to be an arbitrary well-order of D_{n+1} which extends the well-order $\leq_n$ on D_n.

For the second item in the proposition, we simply check that the construction in the previous paragraph preserves injectivity and/or cohesiveness (as well as the much stronger constraint of constancy), as we go along through the induction.

For injectivity, first suppose that $T_{i,\langle a,b\rangle}(f) = T_{i,\langle a,b\rangle}(g)$. To show that $f = g$, suppose that x is from $B_{i,a}$; we must show that $f(x) = g(x)$. Let $y = T_{i,a}(x)$. Then one has the following, where the middle identity follows from the supposition of this paragraph and the two other identities follow from the construction in (3.1):

$$T_{i,b}(f(x)) = T_{i,\langle a,b\rangle}(f)(y) = T_{i,\langle a,b\rangle}(g)(y) = T_{i,b}(g(x))$$

Then by induction applied to the type b of degree $\leq n$, we have that $f(x) = g(x)$, which is what we wanted to show. Similarly, to finish the proof of injectivity, suppose that $T_{i,\langle s,b\rangle}(f) = T_{i,\langle s,b\rangle}(g)$. To show that $f = g$, suppose that x is from $B_{i,s}$; we must show that $f(x) = g(x)$. Let $y = T_{i,s}(x)$. Then one has the following, where the middle identity again follows the supposition of this part of the injectivity proof and the other two identities from the the construction in (3.2):

$$f(x) = T_{i,\langle s,b\rangle}(f)(y) = T_{i,\langle s,b\rangle}(g)(y) = g(x)$$

For cohesiveness, suppose first that f is an element of both $B_{i,\langle a,b\rangle}$ and $B_{j,\langle c,d\rangle}$ where $\langle a,b\rangle$, $\langle c,d\rangle$ are types with a,b,c,d having degree $\leq n$. We must show that both $T_{i,\langle a,b\rangle}(f)$ and $T_{j,\langle c,d\rangle}(f)$ as defined in (3.1) are equal. Since f is an element of both $B_{i,\langle a,b\rangle}$ and $B_{j,\langle c,d\rangle}$, we know that its domain is $B_{i,a} = B_{j,c}$ and hence by induction hypothesis of cohesiveness for degrees $\leq n$ we have that $D_a = D_c$ and $T_{i,a} = T_{j,c}$. Hence $T_{i,\langle a,b\rangle}(f)$ and $T_{j,\langle c,d\rangle}(f)$ have the same domain, namely $D_a = D_c$. Supposing that y is in $D_a = D_c$, it suffices to show $T_{i,\langle a,b\rangle}(f)(y) = T_{j,\langle c,d\rangle}(f)(y)$. First suppose that $y = T_{i,a}(x) = T_{j,c}(x)$ for some x in $B_{i,a} = B_{j,c}$. Then $f(x)$ is in both $B_{i,b}$ and $B_{j,d}$. By induction hypothesis of cohesiveness for degrees $\leq n$, we have $T_{i,b}(f(x)) = T_{j,d}(f(x))$ and thus $T_{i,\langle a,b\rangle}(f)(y) = T_{j,\langle c,d\rangle}(f)(y)$ by (3.1). Note that by generalizing across all y in this first case, we have that $T_{i,b}(f(B_{i,a})) = T_{j,d}(f(B_{j,c}) \subseteq D_n$ and hence that $\min_{\leq_n}(T_{i,b}(f(B_{i,a})) = \min_{\leq_n}(T_{j,d}(f(B_{j,c}))$. By the second case break in (3.1), we then have that $T_{i,\langle a,b\rangle}(f)(y) = T_{j,\langle c,d\rangle}(f)(y)$ when y is not in the image of $T_{i,a} = T_{j,c}$. Hence, both when this does and when this does not happen to y, we have that $T_{i,\langle a,b\rangle}(f)(y) = T_{j,\langle c,d\rangle}(f)(y)$, and thus $T_{i,\langle a,b\rangle}(f) = T_{j,\langle c,d\rangle}(f)$.

Second, suppose that f is an element of both $B_{i,\langle s,b\rangle}$ and $B_{j,\langle s,d\rangle}$ where b, d have degree $\leq n$. We must show that both $T_{i,\langle s,b\rangle}(f)$ and $T_{j,\langle s,d\rangle}(f)$ as defined in (3.2) are equal. Since f is an element of both $B_{i,\langle s,b\rangle}$ and $B_{j,\langle s,d\rangle}$, we know that its domain is $B_{i,s} = B_{j,s}$ and hence by induction hypothesis of cohesiveness for degrees $\leq n$ that $T_{i,s} = T_{j,s}$. Further, $T_{i,\langle s,b\rangle}(f)$ and $T_{j,\langle s,d\rangle}(f)$ have the same domain, namely $D_s = I$. Letting y be in $D_s = I$, it suffices to show $T_{i,\langle s,b\rangle}(f)(y) = T_{j,\langle s,d\rangle}(f)(y)$. First suppose that $y = T_{i,s}(x) = T_{j,s}(x)$ for some x in $B_{i,s} = B_{j,s}$. Then per (3.2) we have $T_{i,\langle s,b\rangle}(f)(y) = f(x) = T_{j,\langle s,d\rangle}(f)(y)$. Note that $B_{i,s} = B_{j,s}$ implies $f(B_{i,s}) = f(B_{j,s}) \subseteq D_b \cap D_d \subseteq D_n$ and hence that $\min_{\leq n}(f(B_{i,s}) = \min_{\leq n}(f(B_{j,s}))$. By the second case break in (3.2), we then have that $T_{i,\langle s,b\rangle}(f)(y) = T_{j,\langle s,d\rangle}(f)(y)$ when y is not in the image of $T_{i,s} = T_{j,s}$. Hence, both when this does and when this does not happen to y, we have that $T_{i,\langle s,b\rangle}(f)(y) = T_{j,\langle s,d\rangle}(f)(y)$, and thus $T_{i,\langle s,b\rangle}(f) = T_{j,\langle s,d\rangle}(f)$.

Third, suppose we have an element f in $B_{i,\langle a,b\rangle}$ and $B_{j,c}$ where c has degree one. But this cannot happen due to the convention about set-theoretic ranks.

Fourth and finally, suppose f is in $B_{i,\langle a,b\rangle}$ and $B_{j,\langle s,d\rangle}$ where $\langle a,b\rangle$ is a type and d is a type and where a, b, d have degree $\leq n$. But then f would have domain $B_{i,a}$ and $B_{j,s}$ and by cohesiveness for degrees $\leq n$ we would have that $D_a = D_s$. By our remark about set-theoretic ranks, this implies that a also has degree one, which contradicts our assumption on frames that D_e, D_t, D_d, D_s are distinct sets.

Finally, for cohesiveness, suppose that $B_{i,\langle a,b\rangle} = B_{j,\langle c,d\rangle}$, where a, b, c, d have degree $\leq n$ (and where a, c can possibly be s). An element of this set has domain $B_{i,a} = B_{j,c}$. Hence by induction hypothesis of cohesiveness for degree $\leq n$, we have that $D_a = D_c$. By the same argument as the previous paragraph, either both $a \neq s$ and $c \neq s$, or both $a = s$ and $c = s$. In the first case, using the constant functions mentioned above, we can obtain likewise that the co-domains $B_{i,b}, B_{j,d}$ are equal. Hence by induction hypothesis again, we have that $D_b = D_d$. Given how $D_{\langle a,b\rangle}$ and $D_{\langle c,d\rangle}$ were constructed (i.e. the set of all functions ...), we thus obtain their identity as well. In the second case, using again the constant functions, we can obtain likewise that the co-domains D_b, D_d are equal. And again, given how $D_{\langle s,b\rangle}$ and $D_{\langle s,d\rangle}$ were constructed, we thus obtain their identity as well.

The argument for constancy proceeds much like the proof of injectivity. In particular, constancy implies that only the first case break in the definitions in (3.1)-(3.2) are used, from which one can see that constancy on types of degree $\leq n$ carries over to types of degree $n+1$.

For the last item, suppose that the base frame is inclusive but not constant. Then one has at least one pair i, a such that $B_{i,a}$ is a proper subset of $D_{i,a}$. Suppose that some frame extending it is inclusive. But consider the type $\langle a, a\rangle$. The definition requires that $B_{i,\langle a,a\rangle}$ is non-empty, and so choose some element f of it. Since $f : B_{i,a} \to B_{i,a}$ and since $B_{i,a}$ is a proper subset of $D_{i,a}$, obviously it is not the case that $f : D_{i,a} \to D_{i,a}$.

□

3.2.4 Constructions on frames

Here we enumerate various constructions on frames which we employ in what follows.

One natural construction is a sum construction. Suppose that $\langle I, B, D, S\rangle$ and $\langle I', B', D, S'\rangle$ are two base frames with the same third component and with disjoint first components. That is, they have the same outer domain and disjoint index sets. Then $\langle I \cup I', B \cup B', D, S \cup S'\rangle$ is a base frame defined in the obvious piecewise manner. Further, if $\langle I, B, D, S, T\rangle$ is a frame extending the base frame $\langle I, B, D, S\rangle$, and if similarly $\langle I', B', D, S', T'\rangle$ is a frame extending the base frame $\langle I, B', D, S'\rangle$, and if for the outer domains relative to each type are the same, then one has that $\langle I \cup I', B \cup B', D, S \cup S', T \cup T' \rangle$ is a frame extending the base frame $\langle I \cup I', B \cup B', D, S \cup S' \rangle$. Hence, for instance, if the two frames individually both have properties like constant, full, minimal, inclusive, or injective, then the union too has this property. This is just because these properties concern properties of the indexes taken one by one. (The same is not true of cohesive, and the union of two cohesive frames need not be cohesive).[3]

3.3 Semantics

3.3.1 Well-formed expressions

A *signature* L is given by a set of ordered pairs consisting of a constant symbol along with its type.

Supposing that L is a signature, we now define the well-formed expressions of the signature, along with their type. To aid in this, we write $\alpha : a$ to indicate that a well-formed expression α has type a. However, it should be kept in mind that $\alpha : a$ is in the metalangauge, and says that the object-language expression α is associated to type a.

Now we formally define the *well-formed expressions* of the signature. As a base case of the definition, we suppose that variables of types a and the constants of type a are both well-formed expression of type a. As the inductive step of the definition, suppose that we already have the following well-formed expressions, wherein v is a variable:

$$\alpha : a \quad \beta : b \quad \gamma : \langle a, b\rangle \quad \xi : \langle s, a\rangle \quad v : a \quad \psi : t \quad \varphi : t \quad \delta : d \quad \epsilon : d$$

Then we define the further well-formed expressions:

[3] This construction is similar to the disjoint union construction in Blackburn et al. (2001, pp. 52 ff.).

$$
\begin{array}{llll}
(\gamma(\alpha)) : b & (\lambda v.\ \beta) : \langle a, b\rangle & (\alpha = \beta) : t & (\delta \leq \epsilon) : t \\
(\varphi \wedge \psi) : t & (\varphi \vee \psi) : t & (\neg\varphi) : t & (\varphi \rightarrow \psi) : t \\
(\forall\ v\ \varphi) : t & (\exists\ v\ \varphi) : t & (\Box\varphi) : t & (\Diamond\varphi) : t \\
(^\frown\alpha) : \langle s, a\rangle & (\smile\xi) : a & (\uparrow\alpha) : a & (\downarrow\alpha) : a \\
(\Pi\ v\ \alpha) : t & (\Sigma\ v\ \alpha) : t & (\mathtt{L}\varphi) : t & (\mathtt{M}\varphi) : t
\end{array}
$$

The first three rows should be familiar from type theory and modal logic. The fourth row pertains to the Montagovian *intension* operator $^\frown$ and *extension* operator $\smile$, as well as the two *Vlach operators* $\uparrow$ and $\downarrow$. In the fifth row, we have the possibilist analogues $\Pi, \Sigma, \mathtt{L}, \mathtt{M}$ of $\forall, \exists, \Box, \Diamond$. Hence, we say that *actualist well-formed expressions* are the subset of the well-formed expressions which do not contain $\Pi, \Sigma, \mathtt{L}, \mathtt{M}$.

As usual, we drop outermost parentheses from well-formed expressions, and we assume that the unary symbols such as negation and the modalities bind tightly.

3.3.2 Preliminaries

As preliminaries to the definition of the semantics relative to a frame, we first take note of considerations pertaining to the variable assignments, the accessibility relations, and the Vlach operators.

If a frame $\langle I, B, D, S, T\rangle$ is given, then a *variable assignment* g relative to the frame is given by a family of maps g_a which maps the variables of type a to the elements of D_a. If a is a type and v is a variable of type a, then variable assignments g, h are *v-variants* if they agree on all variables distinct from v.

The semantics for the modal operators make use of an accessibility relation R on the index set I. We simply add the accessibility relation R to the base frame (resp. frame) and write $\langle I, R, B, D, S\rangle$ (resp. $\langle I, R, B, D, S, T\rangle$). Recall that we say that R is the *universal* accessibility relation on I simply if $R = I \times I$. These are widely used, of course, due to the correspondence with $\mathsf{S5}$. If no accessibility relation is present, it is presumed that the universal accessibility relation is being deployed. Finally, we write $R[i] = \{j \in I : iRj\}$ for the set of indexes accessible from index i.

The semantics for the Vlach operators make use of a finite sequence of indexes – the so-called "storage indexes" – along with an index of evaluation. We write finite sequences of indices with lower-case Greek letters σ, τ, ρ, as well as subscripted and superscripted versions thereof. We include, as a degenerate case, the length zero sequence $\varnothing$ as a finite sequence. We concatenate in the usual way: if σ is of length ℓ and τ is of length m, then $\sigma\tau$ is the length $\ell + m$ sequence which consists of σ followed by τ. Often, we append length one sequences to other sequences. If σ again has length ℓ, then σw has length $\ell + 1$, and its last entry is w. Likewise, $w\sigma$ is the length $\ell + 1$ sequence whose first entry is w and which is followed by σ. Note that one has to be careful *not* to read σi as the result of inputting i into σ, but rather as the concatenation of σ with a length one sequence whose sole entry is i. Finally, if i, j are two distinct indexes, then we define $\sigma \equiv_{i,j} \tau$ iff they have the same length and one can be obtained from the other by replacing some instances of i by j, or vice-versa.

3.3.3 Definition of denotation

These preliminaries in place, we now turn to the definition of model and denotation.

If L is a signature, then an *L-model* $\mathcal{M}$ is given by a frame $M = \langle I, R, B, D, S, T\rangle$ and an interpretation function $\cdot^{\mathcal{M}}$, which identifies a distinguished element $c^{\mathcal{M}}$ of $D_{\langle s,a\rangle}$ for each constant symbol c of type a. Note that we are using the familiar convention that the models are written with cursive letters $\mathcal{M}, \mathcal{N}, \ldots$, while their underlying frames are written with the corresponding Roman letters $M, N, \ldots$

If $\mathcal{M}$ is a model, and i is an index and σ is a finite sequence of indexes and g is a variable assignment relative to the frame M, then we define the *denotation* $[\![\alpha]\!]_{\mathcal{M},i,\sigma,g}$ of each well-formed expression α of type a by induction on α. As a base case, if v is a variable of type a and c is a constant of type a, then we define:

- $[\![v]\!]_{\mathcal{M},i,\sigma,g} = g(v)$ and $[\![c]\!]_{\mathcal{M},i,\sigma,g} = c^{\mathcal{M}}(i)$

For the inductive step of the definition, suppose that we have already defined the semantics for the following well-formed expressions with the displayed types, wherein v is a variable:

$$\alpha : a \quad \beta : b \quad \gamma : \langle a, b\rangle \quad \xi : \langle s, a\rangle \quad v : a \quad \psi : t \quad \varphi : t \quad \delta : d \quad \epsilon : d$$

Then we further define:

- $[\![\gamma(\alpha)]\!]_{\mathcal{M},i,\sigma,g} = [\![\gamma]\!]_{\mathcal{M},i,\sigma,g}([\![\alpha]\!]_{\mathcal{M},i,\sigma,g})$ and $[\![\lambda v.\beta]\!]_{\mathcal{M},i,\sigma,g}(u) = [\![\beta]\!]_{\mathcal{M},i,\sigma,h}$ where h is the v-variant of g with $h(v) = u$, and $[\![\alpha = \beta]\!]_{\mathcal{M},i,\sigma,g} = 1$ iff $[\![\alpha]\!]_{\mathcal{M},i,\sigma,g} = [\![\beta]\!]_{\mathcal{M},i,\sigma,g}$, and $[\![\delta \leq \epsilon]\!]_{\mathcal{M},i,\sigma,g} = 1$ iff $[\![\delta]\!]_{\mathcal{M},i,\sigma,g} \preceq_d [\![\epsilon]\!]_{\mathcal{M},i,\sigma,g}$, wherein $\preceq_d$ is the linear order on D_d.

- $[\![\varphi \wedge \psi]\!]_{\mathcal{M},i,\sigma,g} = \inf([\![\varphi]\!]_{\mathcal{M},i,\sigma,g}, [\![\psi]\!]_{\mathcal{M},i,\sigma,g})$, and $[\![\varphi \vee \psi]\!]_{\mathcal{M},i,\sigma,g} = \sup([\![\varphi]\!]_{\mathcal{M},i,\sigma,g}, [\![\psi]\!]_{\mathcal{M},i,\sigma,g})$, and $[\![\neg\varphi]\!]_{\mathcal{M},i,\sigma,g} = 1 - [\![\varphi]\!]_{\mathcal{M},i,\sigma,g}$, and $[\![\varphi \rightarrow \psi]\!]_{\mathcal{M},i,\sigma,g} = [\![\neg\varphi \vee \psi]\!]_{\mathcal{M},i,\sigma,g}$.

- $[\![\forall\ v\ \varphi]\!]_{\mathcal{M},i,\sigma,g} = \inf\{[\![\varphi]\!]_{\mathcal{M},i,\sigma,h} : h\ v\text{-variant of } g \text{ with } h(v) \in T_{i,a}(B_{i,a})\}$, and $[\![\exists\ v\ \varphi]\!]_{\mathcal{M},i,\sigma,g} = \sup\{[\![\varphi]\!]_{\mathcal{M},i,\sigma,h} : h\ v\text{-variant of } g \text{ with } h(v) \in T_{i,a}(B_{i,a})\}$, and $[\![\Box\varphi]\!]_{\mathcal{M},i,\sigma,g} = \inf\{[\![\varphi]\!]_{\mathcal{M},j,\sigma,g} : j \in T_{i,s}(B_{i,s}) \cap R[i]\}$, and $[\![\Diamond\varphi]\!]_{\mathcal{M},i,\sigma,g} = \sup\{[\![\varphi]\!]_{\mathcal{M},j,\sigma,g} : j \in T_{i,s}(B_{i,s}) \cap R[i]\}$.

- $[\![{}^\frown\alpha]\!]_{\mathcal{M},i,\sigma,g}(k) = [\![\alpha]\!]_{\mathcal{M},k,\sigma,g}$, and $[\![\smile\xi]\!]_{\mathcal{M},i,\sigma,g} = [\![\xi]\!]_{\mathcal{M},i,\sigma,g}(i)$, and $[\![\uparrow\alpha]\!]_{\mathcal{M},i,\sigma,g} = [\![\alpha]\!]_{\mathcal{M},i,\sigma i,g}$, and $[\![\downarrow\alpha]\!]_{\mathcal{M},i,\sigma k,g} = [\![\alpha]\!]_{\mathcal{M},k,\sigma,g}$, and $[\![\downarrow\alpha]\!]_{\mathcal{M},i,\emptyset,g} = [\![\alpha]\!]_{\mathcal{M},i,\emptyset,g}$

- $[\![\Pi\ v\ \varphi]\!]_{\mathcal{M},i,\sigma,g} = \inf\{[\![\varphi]\!]_{\mathcal{M},i,\sigma,h} : h\ v\text{-variant of } g\}$, and $[\![\Sigma\ v\ \varphi]\!]_{\mathcal{M},i,\sigma,g} = \sup\{[\![\varphi]\!]_{\mathcal{M},i,\sigma,h} : h\ v\text{-variant of } g\}$, and $[\![\mathtt{L}\varphi]\!]_{\mathcal{M},i,\sigma,g} = \inf\{[\![\varphi]\!]_{\mathcal{M},j,\sigma,g} : j \in R[i]\}$, and $[\![\mathtt{M}\varphi]\!]_{\mathcal{M},i,\sigma,g} = \sup\{[\![\varphi]\!]_{\mathcal{M},j,\sigma,g} : j \in R[i]\}$.

In these, the infs, sups, and complement operations are from the Boolean algebra structure on $D_t = \{0, 1\}$.

Note that the outer domains of the models are really just constant domain models, and that the definition of denotation proceeds just like normal, except that we use the transfer maps T to bound the actualist quantifiers $\forall$ and $\exists$, as well as the modal operators $\Box$ and $\Diamond$. Hence, as mentioned previously, the transfer maps T serve as the higher-order analogue of the existence-predicates of first-order modal logic.

One variation on this semantics involves replacing the modal logic by a multimodal logic. This results in a minor variation to the above in which we expand the language by new pairs of boxes and diamonds, and then add to the models a new accessibility relation for every such pair. If we just need a couple extra pairs, we might use something like $\blacksquare$ and $\blacklozenge$, or $\boxdot$ and $\diamonddot$. In other settings, such as temporal logic, there are received symbols for the various modalities, and we will just revert to those. When needed, we introduce the possibilist analogues of these, as we go along.

To simplify certain discussions, we assume that our models come equipped with constant symbols which rigidly pick out the elements of D_e and D_t and D_d, and we do not introduce separate notation for these constant symbols. For instance, if α is of type t, then $\alpha = 1$ is a well-formed expression with a constant symbol on the right-hand side of the identity, and we have that $[\![\alpha = 1]\!]_{\mathcal{M},i,\sigma,g} = 1$ iff $[\![\alpha]\!]_{\mathcal{M},i,\sigma,g} = 1$. We do the same with D_e and D_d, and typically this causes no confusion since in applications the sets D_e and D_d come equipped with natural choices of names for its elements. For instance, if D_e consists of three people Anne, Bill, Claire, then we just take their names as constant symbols which rigidly denote these people. Likewise, if D_d is the rational numbers with its standard ordering, we just write well-formed expressions like $\alpha \leq 2.5$, where 2.5 is a constant symbol which always denotes the rational number halfway between two and three.

Finally, we close with some natural definitions made available by the semantics:

- If a is type and y is a variable of type a, then the *existence predicate* $\mathsf{E}_a(y)$ *of type* a is the well-formed expression $\exists\, x\; x = y$, where x is a variable symbol of type a different than y.

- If $\mathcal{M}$ is a model, then a well-formed expression φ of type a *exists at indexes* i, σ *and variable assignment* g *in* $\mathcal{M}$ if $[\![\mathsf{E}_a(\varphi)]\!]_{\mathcal{M},i,\sigma,g} = 1$. Note that a well-formed expression φ of type a exists at indexes i, σ and assignment g in $\mathcal{M}$ if and only if $[\![\varphi]\!]_{\mathcal{M},i,\sigma,g}$ is a member of $T_{i,a}(B_{i,a})$. Since the existence of constants and other closed expressions does not depend on the variable assignment, and we may omit the variable assignments when considering existence in these cases; and likewise since the existence of expressions not containing the Vlach operators does not depend on the storage indexes σ, we may omit them in such cases.

One of the primary uses of the Vlach operators is that they allow expression of the possibilist quantifiers, under the hypothesis that the values of the variables possibly exist. This result holds in this semantics, as we can quickly verify:

Proposition 3.2. *Suppose that v is a variable of type a and suppose that φ has v free. Then the following are valid on a model:*

- $[(\Diamond \mathsf{E}_a(v)) \wedge \varphi] \rightarrow [\uparrow \Diamond \exists\, v \downarrow \varphi]$

- $[\uparrow \Diamond \exists\, v \downarrow \varphi] \rightarrow [\Sigma\, v\, \varphi]$

Hence, if $\Diamond \mathsf{E}_a(v)$ is valid on a model, then $[\Sigma\, v\, \varphi] \leftrightarrow [\uparrow \Diamond \exists\, v \downarrow \varphi]$ is valid on the model.

Proof. For the first item, suppose that $[\![(\Diamond \mathsf{E}_a(v)) \wedge \varphi]\!]_{\mathcal{M},i,\sigma,g} = 1$. By the first conjunct, choose index j in $T_{i,s}(B_{i,s}) \cap R[i]$ with $g(v)$ in $T_{j,a}(B_{j,a})$. Then by the second conjunct we have $[\![\downarrow \varphi]\!]_{\mathcal{M},j,\sigma i,g} = 1$. Since g is a v-variant of itself with $g(v)$ in $T_{j,a}(B_{j,a})$, we have $[\![\exists\ v \downarrow \varphi]\!]_{\mathcal{M},j,\sigma i,g} = 1$. Since j in $T_{i,s}(B_{i,s}) \cap R[i]$ we have $[\![\Diamond\ \exists\ v \downarrow \varphi]\!]_{\mathcal{M},i,\sigma i,g} = 1$. Then $[\![\uparrow \Diamond\ \exists\ v \downarrow \varphi]\!]_{\mathcal{M},i,\sigma,g} = 1$.

Conversely, suppose that $[\![\uparrow \Diamond\ \exists\ v \downarrow \varphi]\!]_{\mathcal{M},i,\sigma,g} = 1$. Then $[\![\Diamond\ \exists\ v \downarrow \varphi]\!]_{\mathcal{M},i,\sigma i,g} = 1$. Then there is index j in $T_{i,s}(B_{i,s}) \cap R[i]$ with $[\![\exists\ v \downarrow \varphi]\!]_{\mathcal{M},j,\sigma i,g} = 1$. Then there is v-variant h of g with $h(v)$ in $T_{j,a}(B_{j,a})$ such that $[\![\downarrow \varphi]\!]_{\mathcal{M},j,\sigma i,h} = 1$. Then $[\![\varphi]\!]_{\mathcal{M},i,\sigma,h} = 1$. Then since g is a v-variant of h, we have $[\![\Sigma\ v\ \varphi]\!]_{\mathcal{M},i,\sigma,g} = 1$. □

The question of whether $\Diamond \mathsf{E}_{et}(v)$ is the question of whether every property $g(v) = F \in D_{et}$ of individuals possibly exists. This often is closely tied to the question of the existence of properties holding only of individuals that cannot coexist. That is, the interesting case is when there are e.g. x_1, x_2 in D_e such that $F(x_1) = 1$ and $F(x_2) = 1$ and $F(x) = 0$ for $x \neq x_1, x_2$, and such that x_1 and x_2 do not both exist in any common index.

3.4 Ultraproduct construction

We describe an ultraproduct construction on frames. It is the natural combination of the usual ultraproduct construction in propositional modal logic with the superstructure construction from nonstandard analysis.[4]

3.4.1 Preliminaries on superstructures

In this subsection we present some preliminaries on superstructures, following the treatment of Chang-Keisler, along with some refinements which will be needed in what follows.[5] For

[4]See Blackburn et al. (2001, pp. 104-106) and Chang and Keisler (1990, §4.4 pp. 262 ff).
[5]See Chang and Keisler (1990, §4.4 p. 262 ff.).

any set X, we define the *superstructure* $V(X)$ over X by:

$$V_0(X) = X, \quad V_{n+1}(X) = V_n(X) \cup P(V_n(X)), \quad V(X) = \bigcup_n V_n(X) \tag{3.3}$$

That is, $V(X)$ is just the ω-stages over the cumulative hierarchy over X. We say that X is a *base set* if: it is non-empty, and the emptyset is not in X, and $x \cap V(X)$ is empty for each element x of X. Being a base set can be ensured by choosing the elements of X to all have the same non-zero set-theoretic rank.[6] One can show the following by induction on $n \geq 0$:[7]

$$V_n(X) \subseteq V_{n+1}(X), \qquad V_{n+1}(X) \setminus X = P(V_n(X)) \tag{3.4}$$

For the latter, the left-to-right inclusion uses the fact that it is a base set. Note that the former quickly implies that any non-empty subset of a base set is also a base set. Likewise this implies that

$$(y \in V(X) \;\&\; y \in x \in V_n(X)) \Longrightarrow (n > 0 \;\&\; y \in V_{n-1}(X)) \tag{3.5}$$

In sum, the hypothesis that X forms a base set is a way to implement the idea of X being urelemente.[8]

Hence, the natural set-theoretic language in which to view $V(X)$ is as a two-sorted language with one sort for urelemente which range over X and one for sets, which range over $V(X) \setminus X$. We write such formulas as $\varphi(\overline{x}, \overline{y})$ with all free variables and parameters displayed wherein $\overline{x}$ is reserved for urelements and $\overline{y}$ are reserved for sets. The *bounded formulas* are the smallest class of such formulas which contains $x_i \in y_j$ and $y_i \in y_j$, which is closed under the propositional connectives, and which is closed under bound quantifiers which are bound to sets: that is $\exists\, x_i \in y_j\; \varphi(\overline{x}, \overline{y})$ and $\exists\, y_i \in y_j\; \varphi(\overline{x}, \overline{y})$ are in the class if $\varphi(\overline{x}, \overline{y})$ is, and similarly

[6]See Jech (2006, p. 250, Lemma 15.47).
[7]Equations (3.4)-(3.5) are from Chang and Keisler (1990, p. 264).
[8]See Jech (2006, p. 250).

for the universal quantifier. For each bounded formula in this two-sorted language we define a Δ_0-formula $\varphi^X(\overline{x},\overline{y}) \equiv (\varphi(\overline{x},\overline{y}))^X$ in the one-sorted language of ordinary set theory which has parameter X and which is defined by the identity map on atomics, compositionally on the propositional connectives, and by $(\exists\ x_i \in y_j\ \varphi(\overline{x},\overline{y}))^X \equiv \exists\ x_i \in (y_j \cap X)\ \varphi^X(\overline{x},\overline{y})$, and by $(\exists\ y_i \in y_j\ \varphi(\overline{x},\overline{y}))^X \equiv \exists\ y_i \in (y_j \setminus X)\ \varphi^X(\overline{x},\overline{y})$. That is, the map $\cdot^X$ is a flattening map from the two-sorted to the one-sorted language which simply replaces the bound urelemente quantifiers by elements of X, and replaces the bound set quantifiers by elements of $V(X)\setminus X$. Then one can show that for all bounded formulas $\varphi(\overline{x},\overline{y})$ and all $\overline{x}$ from X and all $\overline{y}$ from $V(X) \setminus X$ one has:

$$V(X) \vDash \varphi(\overline{x},\overline{y}) \text{ iff } V \vDash \varphi^X(\overline{x},\overline{y}) \tag{3.6}$$

On the left-hand side, by V is meant the set-theoretic universe itself.

This result is in (3.6) is in general false if one replaces bounded formulas with Δ_0-formulas. For instance, $V(X)$ would model that any two urelements x_1, x_2 are subsets of one another, whereas V need not.[9] The relevant formula $\forall\ x_3 \in x_1,\ x_3 \in x_2$ expresssing that urelement x_1 is a subset of urelemente x_2 violates the constraints on bounded formulas in that it places urelements on the right-hand side of the membership relation. That is, it is not boundedly-expressible because its satisfaction conditions require checking about the members of urelemente, which are hidden from view in $V(X)$. The Δ_0-formulas which are boundedly-expressible are those which never query membership in urelemente and whose quantifiers are always bound to sets from $V(X)$ rather than urelemente from $V(X)$.

Using this rule of thumb, one can quickly consult standard lists of Δ_0-expressible concepts and find instances of these which are boundedly-expressible.[10] To describe these we again use the convention that x-variables, including subscripted versions, are reserved for urele-

[9] This is the example mentioned before Lemma 4.4.4 in Chang and Keisler (1990, p. 265). The result in (3.6) is our preferred way of generalizing this lemma.

[10] See Kunen (1980, pp. 120-122). Compare Chang and Keisler (1990, Lemma 4.4.3).

mente and y-variables, including subscripted versions, are reserved for sets. And in this, the expressibility of a functional notion is understood as the expressibility of its graph:

- *Boolean algebra structure on sets*: $y_1 \cup y_2$, $y_1 \cap y_2$, $y_1 \cup y_2$, $y_1 \setminus y_2$.
- *Pairing*: $\{x\}$, $\{x_1, x_2\}$, $\langle x_1, x_2 \rangle = \{x_1, \{x_1, x_2\}\}$; y is a unordered pair of two urelemente or sets; y is an ordered pair of two urelemente or sets. Ordered triples $\langle x_1, x_2, x_3 \rangle = \langle \langle x_1, x_2 \rangle, x_3 \rangle$ and similarly for other n-tuples.
- *Relations and functions*: y is an binary relation on urelemente or sets; y is such a binary relation which is the graph of a function; y_1 is the domain of such a function y; y_2 is the range of such a function y; y is the graph of a function $y : y_1 \to y_2$; y is the result of composing two such graphs (when the range of the one is a subset of the range of the other); x_2 is the result of applying the graph of such a function y to an input x_1 within its domain.

In these items, one can also replace the urelemente variables by set variables and still retain a boundedly-expressible notion. A helpful calculation related to pairing is the following:

$$y_1, \ldots, y_m \in V_n(X) \Longrightarrow \langle y_1, \ldots, y_m \rangle \in V_{n+2m}(X) \tag{3.7}$$

And similarly when we replace set variable y_i by urelemente variable x_i, we get a bound of $V_{2m}(X)$ for the n-tuple.

The result in (3.6) has important consequences for the above list. For instance, if $\varphi(y, y_1, y_2)$ is the bounded formula expressing that $y : y_1 \to y_2$, then one has that for all sets y, y_1, y_2 in $V(X)$:

$$V(X) \vDash \varphi(y, y_1, y_2) \text{ iff } V \vDash y : y_1 \to y_2$$

Results of this form are typically called *absoluteness* results.[11] It is the formal justification for claims that $V(X)$ correctly models certain notions. There are limits to absoluteness, as the ultraproduct construction can help us to see.

3.4.2 The bounded ultraproduct

Let J be an index set, and let U be an ultrafilter over J. That is, U is a subset of the powerset of J which is closed under finite intersections, closed upwards under subsets of J, does not contain $\varnothing$, and contains each subset of J or its complement. Intuitively, U is a collection of large subsets of J, while its complement is a collection of small subsets of J. We say that a property Φ holds for *U-many j* if $\{j \in J : \Phi(j)\}$ is an element of U. This is the analogue of the measure-theoretic notion of a property holding almost everywhere or almost surely.

We consider subsets of the ultraproduct $\prod_U V(X)$ which are bounded in a certain sense. Recall that we obtain the ultraproduct by considering maps $x : J \to V(X)$ modulo the equivalence relation which declares $x, x' : J \to V(X)$ to be equivalent iff they agree on a large set, that is, $\{j \in J : x(j) = x'(j)\} \in U$. We denote the equivalence class of x as $[x]_U$. Hence the ultraproduct $\prod_U V(X)$ is defined to consist of all equivalence classes $[x]_U$ as x ranges over maps $x : J \to V(X)$. Without loss of generality, we may assume that all elements of $\prod_U V(X)$ have the same set-theoretic rank, which is above the set-theoretic ranks of the elements of $V(X)$.[12] This implies that $\prod_U V(X)$ is a base set, as too are any of its non-empty subsets.

We define auxiliary sets $A_n(X)$ which collect together the equivalence class of maps $x : J \to V_n(X)$. That is, $A_n(X) = \{[x]_U \in \prod_U V(X) : x : J \to V_n(X)\}$, and we likewise set

[11] See Kunen (1980, Chapter IV).

[12] One could do this by choosing ordinal γ above all the elements of X and replacing the usual set-theoretic construction of the equivalence classes $[x]_U$ by $\{\gamma\} \cup [x]_U$.

$A(X) = \bigcup_n A_n(X)$. Then one can check that we have the following analogue of (3.4):

$$A_n(X) \subseteq A_{n+1}(X), \quad A_{n+1}(X) \setminus A_0(X) = \{[x]_U : \mathrm{rng}(x) \subseteq P(V_n(X))\} \tag{3.8}$$

Let $Y = A_0(X)$, so that Y is a base set. We define the Mostowski collapse function $\pi : A(X) \to V(Y)$ such that $\pi = \bigcup_n \pi_n$ where $\pi_n : A_n(X) \to V_n(Y)$. For $n = 0$, define π_n to be the identity function. For the induction step, we define π_{n+1} on $A_{n+1}(X)$ by $\pi_{n+1}([x]_U) = [x]_U$ if $[x]_U \in A_0(X)$, and otherwise by

$$\pi_{n+1}([x]_U) = \{\pi_n([y]_U) : [y]_U \in A_n(X) \text{ \& for } U\text{-many } j,\, y(j) \in x(j)\} \tag{3.9}$$

Like in the proof of the Mostowski collapse theorem,[13] one shows by induction on $n \geq 0$ that

- $\pi_n : A_n(X) \to V_n(Y)$,
- $\pi_n : A_n(X) \to V_n(Y)$ is an injection,
- For $m < n$, π_m and π_n agree on $A_m(X)$,
- $\pi_n : (A_n(X) \setminus (\bigcup_{m<n} A_m(X))) \to (V_n(Y) \setminus (\bigcup_{m<n} V_m(Y)))$.

Note that the fourth item implies that if $[x]_U$ in A, then there are U-many j such that $x(j)$ is an urelement of $V(X)$ (resp. set of $V(X)$) if and only if $[x]_U$ is in $A_0(X)$ (resp. in $A(X) \setminus A_0(X)$), which in turn happens if and only if $\pi([x]_U)$ is an urelement of $V(Y)$ (resp. set of $V(Y)$).

As in the previous subsection, we write bounded formulas as $\varphi(\overline{x}, \overline{y})$ with all free variables and parameters displayed wherein $\overline{x}$ is reserved for urelements and $\overline{y}$ are reserved for sets. We

[13] As one can see, the hypotheses of set-like and extensional from that theorem are satisfied in this context. The third hypothesis of the theorem, namely, well-foundedness, is taken care of by the fact that we are only going ω-stages up in the hierarchy above a set of urelemente. See Jech (2006, p. 69, Theorem 6.15) for additional details.

write $[\overline{x}]_U$ as an abbreviation for the tuple $[\overline{x}_1]_U, \ldots, [x_n]_U$, and similarly for $[\overline{y}]_U$. Likewise, we write $\overline{x}(j)$ as an an abbreviation for $x_1(j), \ldots, x_n(j)$, and similarly for $\overline{y}(j)$. Then we have the following, for every bounded formula $\varphi(\overline{x}, \overline{y})$ and $[\overline{x}]_U, [\overline{y}]_U$ in $A(X)$ such that for U-many j one has that $\overline{x}(j)$ are urelemente of $V(X)$ and $\overline{y}(j)$ are sets of $V(X)$:

$$\big(\text{ for } U\text{-many } j,\, V(X) \vDash \varphi(\overline{x}(j), \overline{y}(j))\,\big) \text{ iff } V(Y) \vDash \varphi(\pi([\overline{x}]_U), \pi([\overline{y}]_U)) \tag{3.10}$$

The proof is identical to the proof of the Łos theorem. With two applications of (3.6) we obtain:

$$\big(\text{ for } U\text{-many } j,\, V \vDash \varphi^X(\overline{x}(j), \overline{y}(j))\,\big) \text{ iff } V \vDash \varphi^Y(\pi([\overline{x}]_U), \pi([\overline{y}]_U)) \tag{3.11}$$

Finally, we record information about the saturation of bounded ultraproducts. The following property is *countable bounded saturation*:

- Suppose $n_0 \geq 0$. Suppose $\varphi_i(\overline{x}, \overline{y}, \overline{u}_i, \overline{v}_i)$ is a countable sequence of bounded formulas with all free variables displayed, where $\overline{x}, \overline{u}_i$ are urelemente variables and $\overline{y}, \overline{v}_i$ are set variables. Suppose that $[\overline{p}_i]_U$ in $A_0(X)$ and $[\overline{q}_i]_U$ in $A_{n_0}(X)$, and that for every $m > 0$ there is $[\overline{x}_m]_U$ in $A_0(X)$ and $[\overline{y}_m]_U$ in $A_{n_0}(X)$ such that $V(Y) \vDash \bigwedge_{i \leq m} \varphi_i(\pi([\overline{x}_m]_U), \pi([\overline{y}_m]_U), \pi([\overline{p}_i]_U), \pi([\overline{q}_i]_U))$. Then there is $[\overline{x}]_U$ in $A_0(X)$ and $[\overline{y}]_U$ in $A_{n_0}(X)$ such that one has $V(Y) \vDash \bigwedge_m \varphi_m(\pi([\overline{x}_m]_U), \pi([\overline{y}_m]_U), \pi([\overline{p}_m]_U), \pi([\overline{q}_m]_U))$.

Following the familiar argument, a sufficient condition for this is the following property of the ultraproduct, which is called *countable incompleteness*:

- There is a countable sequence X_i in the ultrafilter U such that $\bigcap_i X_i$ is not in U.

For any infinite index set J, there are always countably incomplete ultrafilters.[14]

[14] See Keisler (1964).

3.4.3 Frames within superstructures and bounded ultraproducts

To apply the bounded ultraproduct to frames, we proceed as follows. First, given a frame whose base elements are subsets of a base set X, we show that equivalence classes of functions which take these elements as values lie within one of the sets $A_n(X)$ to which the Łos-like result (3.10) applies. Second, we show that the property of being a frame is expressible by a bounded formula. In the next section we proceed similarly for models, but in this section we tackle frames.

Hence, suppose that $\langle I, B, D, T\rangle$ is a frame, and suppose that X is a base set which is a superset of I, B_a, D_a for all j from J and basic extended types a. Then we show by induction on $\deg(c)$ that for each extended type c:

- For each $i \in I$ the set $B_{i,c}$ is in $V_{4\cdot\deg(c)}(X)$
- The function $B_{\cdot,c} : I \to V_{4\cdot\deg(c)}(X)$ is in $V_{8\cdot\deg(c)}(X)$.
- The set D_c is in $V_{4\cdot\deg(c)}(X)$
- For each $i \in I$ the set $T_{i,c} : B_{i,c} \to D_c$ is an element of $V_{12\cdot\deg(c)}(X)$
- The $T_{\cdot,c} : I \to V_{12\cdot\deg(c)}(X)$ is an element of $V_{16\cdot\deg(c)}(X)$.

It suffices to prove the item for D_c by induction on $n = \deg(c)$, since the proof of the other items just iterate this proof. For $n = 1$ this is because $D_c \subseteq X$ implies $D_c \in V_1(X) \subseteq V_4(X)$. Suppose it holds for n; to show it holds for $n + 1$, first consider type $c = \langle a, b\rangle$ where a, b are types. Then every f in $D_{\langle a,b\rangle}$ is a function $f : D_a \to D_b$. Then by (3.7) one has that for each z in D_a one has that $\langle z, f(z)\rangle$ is in $V_{4n+2}(X)$. Then when identified with its graph, one has that f is a subset of $V_{4n+2}(X)$ and so an element of $V_{4n+3}(X)$. Hence $D_{\langle a,b\rangle}$ is a subset of $V_{4n+3}(X)$ and so an element of $V_{4(n+1)}(X)$. The same argument works for $c = \langle s, b\rangle$.

Now we turn to the bounded-expressibility of being a frame, where we deploy the strict partial order $<$ and partial order $\leq$ on basic extended types from §3.2.1. As mentioned there, the set $\mathrm{pred}_{\leq}(c)$ of $\leq$-predecessors of c has size $m_c \leq 2^n - 1$ when c has degree n. We define a countable sequence of $4m_c$-place bounded formulas $\varphi_c(I, \overline{B}_{\cdot,b}, \overline{D}_b, \overline{T}_{\cdot,b})$, whose places are reserved for the m_c-many 4-tuples $I, B_{\cdot,b}, D_b, T_{\cdot,b}$ as b ranges $\leq c$. Then we show the following, where c ranges over basic extended types:

$$\{I, B_{\cdot,c}, D_c, T_{\cdot,c} : c\} \text{ defines a frame iff } \forall\, c\, V(X) \vDash \varphi_c(I, \overline{B}_{\cdot,b}, \overline{D}_b, \overline{T}_{\cdot,b}) \tag{3.12}$$

That is, a countable sequence $I, B_{\cdot,c}, D_c, T_{\cdot,c}$ as c ranges over basic extended types defines a frame if and only if $V(X) \vDash \varphi_c(I, \overline{B}_{\cdot,b}, \overline{D}_b, \overline{T}_{\cdot,b})$ for all basic extended types c. By "define a frame" we mean the natural notion: if we define $\langle I, B, D, T\rangle$ by setting them equal at basic extended type c to $\langle I, B_{\cdot,c}, D_c, T_{\cdot,c}\rangle$, then the result is a frame. The result in (3.12) will hold by the inspection of the definition of φ_c, which we will produce shortly, together with the absoluteness results described in §3.4.1.

The definition of φ_c is by induction on basic extended type c, and simply involves carefully writing out the definition of frame in §3.2.2 to make sure that it is boundedly-expressible. For the basic extended type e, we define $\varphi_e(I, B_{\cdot,e}, D_e, T_{\cdot,e})$ to say that: I is a non-empty set of urelemente, and $B_{\cdot,e}$ is a function with domain I, and D_e is a non-empty set of urelemente, and $T_{\cdot,e}$ is a function with domain I, and for all i in I one has that $T_{i,e} : B_{i,e} \to D_e$. For the functional notions (e.g. domain, range, and application), we use the bounded-expressibility set out in §3.4.1. The other basic extended types are similar. For basic extended type $\langle a, b\rangle$ we define $\varphi_{\langle a,b\rangle}(I, \overline{B}_{\cdot,\langle a,b\rangle}, \overline{D}_{\langle a,b\rangle}, \overline{T}_{\cdot,\langle a,b\rangle})$ to be:

- The conjunction of $\varphi_a(I, \overline{B}_{\cdot,c}, \overline{D}_c, \overline{T}_{\cdot,c})$ where the entries are indexed by the c ranging $\leq a$, and $\varphi_b(I, \overline{B}_{\cdot,c}, \overline{D}_c, \overline{T}_{\cdot,c})$ where the entries are indexed by the c ranging $\leq b$.

- $B_{\cdot,\langle a,b\rangle}$ is a function with domain I; and for all i in I and all f in $B_{i,\langle a,b\rangle}$, one has that $f : B_{i,a} \to B_{i,b}$ such that for all x_0, x_1 from $B_{i,a}$ if $T_{i,a}(x_0) = T_{i,a}(x_1)$ then $T_{i,b}(f(x_0)) = T_{i,b}(f(x_1))$

- For all f in $D_{\langle a,b\rangle}$ one has that $f : D_a \to D_b$

- $T_{\cdot,\langle a,b\rangle}$ is a function with domain I; and for all i in I one has $T_{i,\langle a,b\rangle} : B_{i,\langle a,b\rangle} \to D_{\langle a,b\rangle}$ such that for all x in $B_{i,a}$ one has $(T_{i,\langle a,b,\rangle}(f))(T_{i,a}(x)) = T_{i,b}(f(x))$.

Of course, the second and fourth items simply record (*frame:1*) and (*frame:2*) in a way that makes their bounded-expressibility vivid. The definition for the basic extended type $\langle s, b\rangle$ is very similar and does the same for (*frame:1s*) and (*frame:2s*).

This allows us to define what we shall call the ultraproduct of a sequence of frames. Suppose that J is an index set and $\langle I^j, B^j, D^j, T^j\rangle$ is a sequence of frames. By replacing by copies, we may assume that there is a base set X which contains $I^j, B^j_{i,a}, D^j_a$ for each basic extended type a and each i in I^j, and we may further assume that $D^j_t = D^{j'}_t = \{0, 1\}$ for each j, j' from j. Further suppose that U is an ultrafilter on J. Then define the following maps from J to $V(X)$, where c ranges over extended types, :

$$\iota(j) = I^j, \quad \beta_c(j) = B^j_{\cdot,c}, \quad \delta_c(j) = D^j_c, \quad \tau_c(j) = T^j_{\cdot,c} \tag{3.13}$$

Then $[\iota]_U$ is in $A_0(X) = Y$ and $[\beta_c]_U, [\delta_c]_U, [\tau_c]_U$ are members of $A_{16\cdot\deg(c)}(X)$. Then define their image under $\pi : A(X) \to V(Y)$ as follows

$$I = \pi([\iota]_U), \quad B_{\cdot,c} = [\beta_c]_U, \quad D_c = [\delta_c]_U, \quad T_{\cdot,c} = [\tau_c]_U \tag{3.14}$$

Since $\langle I^j, B^j, D^j, T^j\rangle$ are frames for each j in J, by the left-to-right direction of (3.12), we have that for all extended types c that $V(X) \models \varphi_c(I^j, \overline{B}^j_{\cdot,b}, \overline{D}^j_b, \overline{T}^j_{\cdot,b})$, where again b ranges over extended types $\leq c$. Then by the definitions in (3.13), this implies that $V(X) \models$

$\varphi_c(\iota(j), \overline{\beta(j)}_{\cdot,b}, \overline{\delta(j)}_b, \overline{\tau(j)}^j_{\cdot,b})$. Then by the left to right direction of (3.10) and the definitions in (3.14) we have that $V(Y) \vDash \varphi_c(I, \overline{B}_{\cdot,b}, \overline{D}_b, \overline{T}_{\cdot,b})$. Then by the right-to-left direction of (3.12), we have that $\langle I, B, D, T\rangle$ is a frame. We hence define $\langle I, B, D, T\rangle$ to be *the ultra-product of the sequence of frames* $\langle I^j, B^j, D^j, T^j\rangle$ indexed by j from J, over the ultrafilter U, and we denote this as $\prod_U \langle I^j, B^j, D^j, T^j\rangle$.

Suppose that U is countably incomplete. Suppose that a is an extended type such that for each $n > 0$ there are U-many j with D^j_a having size $> n$. Then we show that: $D_{\langle a,t\rangle}$ is a proper subset of $D_t^{D_a}$. To see this, let $<_j$ be a well-order on D^j_a. For each $n > 0$ define a map $x_n : J \to V(X)$ by $x_n(j)$ is the $(n+1)$-st element of D^j_a under the ordering $<_j$ if D^j_a has size $> n$, and let it be least element of D^j_a under $<_j$ otherwise. Then $[x_n]_U$ are distinct elements of $A_0(X)$ and for each $n > 0$ for U-many j we have $x_n(j) \in D_j(a)$. Consider $\{\pi([x_n]_U) : n > 0\}$, which is then a subset of D_a. Let $f : D_a \to D_t$ be its characteristic function. Then we claim that f is not in $D_{\langle a,t\rangle} = \pi([\delta_{\langle a,t\rangle}]_U)$. For, if it were, then $f = \pi([\xi]_U)$ for some $[\xi]_U$ in $A(X)$. Then consider the bounded formula $\psi_i(x, y, u_1, \ldots, u_i)$ which says that y is a function and $y(x) = 1$ and x is distinct from $u_1, \ldots, u_i$. Then for each $m > 0$ one has that $V(Y) \vDash \bigwedge_{i\leq m} \psi_i(\pi([x_{m+1}]_U), \pi([\xi]_U), \pi([x_1]_U), \ldots, \pi([x_i]_U))$. Hence by countable bounded saturation, there is $[\zeta]_U$ in $A_0(X)$ such that $V(Y) \vDash \bigwedge_i \psi_i(\pi([\zeta]_U), \pi([\xi]_U), \pi([x_1]_U), \ldots, \pi([x_i]_U))$. But then $f(\pi([\zeta]_U)) = 1$, which implies that $\pi([\zeta_U]) = \pi([x_i]_U)$ for some i, a contradiction. Hence indeed $D_{\langle a,t\rangle}$ is a proper subset of $D_t^{D_a}$, so that the frame is not full. The same argument works for a types $\langle a, b\rangle, \langle s, b\rangle$ under the additional assumption that there are U-many j such that D^j_b has more than two elements. For, in this case just put well-orders on D^j_b and identify its least element with zero and its next element with one, and use these elements in lieu of the role that zero and one play in characteristic functions.

3.4.4 Models within superstructures and bounded ultraproducts

We now turn to bounded ultraproducts of models. Like with frames, we begin with a single model $\mathcal{M}$, in two steps. First, given a model whose base elements are subsets of a base set X, we show that that the equivalence classes of the functions which take the interpretation of the denotation functions of well-formed expressions lie within one of the sets $A_n(X)$ to which the Łos-like result (3.10) applies. Second, we show that these denotation functions are definable.

Hence, suppose that $\mathcal{M} = \langle I, B, D, T\rangle$ is a model in signature L, and suppose that X is a base set which is a superset of I, B_a, D_a for all j from J and basic extended types a. Then for well-formed expression α of type c whose free variables are exactly $v_1 : c_1, \ldots, v_n : c_n$, and for each $k \geq 0$, the partial denotation function

$$[\![\alpha]\!]_{\mathcal{M},\cdot,\cdot,\cdot} : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c \tag{3.15}$$

is an element $V_{\rho(\alpha,k)}(X)$, where $\rho(\alpha, k) = 4\cdot\max(\deg(\overline{c}_i), \deg(c)) + 2(n+k+2)+1$. (Note that c, n are determined by α). In (3.15) we view the three inputs in the subscript as respectively reserved for: the index of evaluation, a k-tuple of indexes used for evaluation of the Vlach operators, and the inputs from the D_{c_i} as supplying values for the free variables and hence playing the role of the variable assignment. The denotation is partial only in that it depends on $k \geq 0$, and to better mark this in the notation we mark it with a superscript:

$$[\![\alpha]\!]^k_{\mathcal{M},\cdot,\cdot,\cdot} : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c \tag{3.16}$$

The calculation that $[\![\alpha]\!]^k_{\mathcal{M},\cdot,\cdot,\cdot}$ is an element of $V_{\rho(\alpha,k)}(X)$ simply follows from the calculation regarding ordered tuples in (3.7) and the results about the entry of D_c-sets into $V(x)$ from §3.4.3 as a function of the degree of c.

Now we show that for each well-formed expression α and each $k \geq 0$ there is a bounded formula $\psi_{\alpha,k}(y, \cdot, \cdot, \cdot, \cdot)$ such that for all superstructures $V(X)$ one has:

$$\begin{aligned}
&\{I, B_{\cdot,c}, D_c, T_{\cdot,c} : c\} \text{ is a frame } \& \ \{[\![\alpha]\!]^k_{\mathcal{M}} : \alpha, \geq 0\} \text{ is denotation function of } \mathcal{M} \\
&\qquad\qquad \text{iff } \forall\, c\, V(X) \vDash \varphi_c(I, \overline{B}_{\cdot,b}, \overline{D}_b, \overline{T}_{\cdot,b}) \\
&\qquad\qquad\quad \& \ \forall \alpha\ \forall\ k \geq 0\ V(X) \vDash \psi_{\alpha,k}([\![\alpha]\!]_{\mathcal{M}}, I, \overline{B}_{\cdot,a}, \overline{D}_a, \overline{T}_{\cdot,a})
\end{aligned} \tag{3.17}$$

In this we use the formulas and notation as in (3.12). The parameters in $\psi_{\alpha,k}$ are four-tuples of $\overline{B}_{\cdot,a}, \overline{D}_a, \overline{T}_{\cdot,a}$ as a ranges over the extended types $\leq$ the type of any subexpression of α. To enhance readability, we suppress these parameters when defining the formulas.

We define these formulas $\psi_{\alpha,k}$ by induction on complexity of α, handling all $k \geq 0$ simultaneously:[15]

- For α a variable $v : c$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I \times I^k \times D_c \to D_c$ and that for all (i, σ, x) in $I \times I^k \times D_c$ one has $y(i, \sigma, x) = x$.

- For $\alpha : c$ is a constant, the bounded formula $\psi_{\alpha,k}(y)$ simply says that $y : I \times I^k \to D_c$ and for all $(i, \sigma), (i, \tau)$ in $I \times I^k$ one has $y(i, \sigma) = y(i, \tau)$.

- For α an instance of application $\gamma(\beta)$ where $\alpha : c$ and $\gamma : \langle b, c\rangle$ and $\beta : b$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c$ such that for all $(i, \sigma, u_1, \ldots, u_n)$ in $I \times I^k \times D_{c_1} \times \cdots \times D_{c_n}$, there is y_1 in $V_{\rho(\beta,k)}$ and z_1 in D_b such that $\psi_{\beta,k}(y_1)$ and $y_1(i, \sigma, u_1, \ldots, u_n) = z_1$, and there is y_2 in $V_{\rho(\gamma,k)}$ and z_2 in $D_{\langle b,c\rangle}$ such that $\psi_{\beta,k}(y_2)$ and $y_2(i, \sigma, u_1, \ldots, u_n) = z_2$, and $y(i, \sigma, u_1, \ldots, u_n) = z_2(z_1)$.

- For α an instance of lambda abstraction $\lambda v_{n+1}.\beta$ where $v_{n+1} : c_{n+1}$ and $\beta : b$ and $\alpha : c$ where $c = \langle c_{n+1}, b\rangle$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c$ such that for all $(i, \sigma, u_1, \ldots, u_n)$ in $I \times I^k \times D_{c_1} \times \cdots \times D_{c_n}$ one has that $y(i, \sigma, u_1, \ldots, u_n)$

[15]Note that it is only the clauses related to the Vlach operators that depend on k.

is a function in $D_{\langle a,b\rangle}$ such that there is y_1 in $V_{\rho(\beta,k)}$ with $\psi_{\beta,k}(y_1)$ and for all u_{n+1} from $D_{c_{n+1}}$ one has $(y(i,\sigma,u_1,\ldots,u_n))(u_{n+1}) = y_1(i,\sigma,u_1,\ldots,u_n,u_{n+1})$.

- For α an instance of equality $\beta = \gamma$ where $\beta : b, \gamma : b$ and $c = t$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c$ such that for all $(i,\sigma,u_1,\ldots,u_n)$ in $I \times I^k \times D_{c_1} \times \cdots \times D_{c_n}$ there is y_1 in $V_{\rho(\beta,k)}$ with $\psi_{\beta,k}(y_1)$ and there is y_2 in $V_{\rho(\gamma,k)}$ with $\psi_{\gamma,k}(y_2)$ and $y(i,\sigma,u_1,\ldots,u_n) = 1$ if $y_1(i,\sigma,u_1,\ldots,u_n) = y_2(i,\sigma,u_1,\ldots,u_n)$, while $y(i,\sigma,u_1,\ldots,u_n) = 0$ if $y_1(i,\sigma,u_1,\ldots,u_n) \neq y_2(i,\sigma,u_1,\ldots,u_n)$.

- For α an instance of degree comparison $\beta \prec_d \gamma$ where $\beta : d, \gamma : d$ and $c = t$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c$ such that for all $(i,\sigma,u_1,\ldots,u_n)$ in $I \times I^k \times D_{c_1} \times \cdots \times D_{c_n}$ there is y_1 in $V_{\rho(\beta,k)}$ with $\psi_{\beta,k}(y_1)$ and there is y_2 in $V_{\rho(\gamma,k)}$ with $\psi_{\gamma,k}(y_2)$ and $y(i,\sigma,u_1,\ldots,u_n) = 1$ if $y_1(i,\sigma,u_1,\ldots,u_n) \prec_d y_2(i,\sigma,u_1,\ldots,u_n)$, while $y(i,\sigma,u_1,\ldots,u_n) = 0$ if not $y_1(i,\sigma,u_1,\ldots,u_n) \prec_d y_2(i,\sigma,u_1,\ldots,u_n)$.

- For α a propositional connective, say $\beta \wedge \gamma$, where $\beta : t, \gamma : t$ and $c = t$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c$ such that for all $(i,\sigma,u_1,\ldots,u_n)$ in $I \times I^k \times D_{c_1} \times \cdots \times D_{c_n}$ there is y_1 in $V_{\rho(\beta,k)}$ with $\psi_{\beta,k}(y_1)$ and there is y_2 in $V_{\rho(\gamma,k)}$ with $\psi_{\gamma,k}(y_2)$ and $y(i,\sigma,u_1,\ldots,u_n) = 1$ if both $y_1(i,\sigma,u_1,\ldots,u_n) = 1$ and $y_2(i,\sigma,u_1,\ldots,u_n) = 1$, while $y(i,\sigma,u_1,\ldots,u_n) = 0$ if not both $y_1(i,\sigma,u_1,\ldots,u_n) = 1$ and $y_2(i,\sigma,u_1,\ldots,u_n) = 1$. The other propositional connectives are handled similarly.

- For α a quantifier, say $\exists\, v_{n+1}\, \beta$, where $v_{n+1} : c_{n+1}$ and $\beta : t$ and $c = t$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c$ such that for all $(i,\sigma,u_1,\ldots,u_n)$ in $I \times I^k \times D_{c_1} \times \cdots \times D_{c_n}$ there is y_1 in $V_{\rho(\beta,k)}$ with $\psi_{\beta,k}(y_1)$ such that $y(i,\sigma,u_1,\ldots,u_n) = 1$ iff there is u_{n+1} in $D_{c_{n+1}} \cap T_{i,c_{n+1}}(B_{i,c_{n+1}})$ such that $y_1(i,\sigma,u_1,\ldots,u_n,u_{n+1}) = 1$. The other quantifiers (both actualist and possibilist) are handled similarly.

- For α a modal operator, say $\Diamond\beta$, where $\beta : t$ and $c = t$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c$ such that for all $(i,\sigma,u_1,\ldots,u_n)$ in $I \times I^k \times D_{c_1} \times \cdots \times D_{c_n}$

there is y_1 in $V_{\rho(\beta,k)}$ with $\psi_{\beta,k}(y_1)$ such that $y(i,\sigma,u_1,\ldots,u_n) = 1$ iff there is j in $D_s \cap \cap T_{i,s}(B_{i,s})$ such that $y_1(j,\sigma,u_1,\ldots,u_n) = 1$. The other modal operators (both actualist and possibilist) are handled similarly.

- For α the intension operation $\frown\beta$, where $\beta : b$ and $c = \langle s,b\rangle$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c$ such that for all $(i,\sigma,u_1,\ldots,u_n)$ in $I \times I^k \times D_{c_1} \times \cdots \times D_{c_n}$ one has that $y(i,\sigma,u_1,\ldots,u_n)$ is a function in $D_{\langle s,b\rangle}$ such that there is y_1 in $V_{\rho(\beta,k)}$ with $\psi_{\beta,k}(y_1)$ and for all j from D_s one has $(y(i,\sigma,u_1,\ldots,u_n))(j) = y_1(j,\sigma,u_1,\ldots,u_n)$.

- For α the extension operation $\smile\beta$, where $\beta : \langle s,c\rangle$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I\times I^k\times D_{c_1}\times\cdots\times D_{c_n} \to D_c$ such that for all $(i,\sigma,u_1,\ldots,u_n)$ in $I\times I^k\times D_{c_1}\times\cdots\times D_{c_n}$ there is y_1 in $V_{\rho(\beta,k)}$ with $\psi_{\beta,k}(y_1)$ and $y(i,\sigma,u_1,\ldots,u_n) = y_1(i,\sigma,u_1,\ldots,u_n)(i)$.

- For α the upwards Vlach operator $\uparrow\beta$ where $\beta : c$, the bounded formula $\psi_{\alpha,k}(y)$ says that $y : I\times I^k\times D_{c_1}\times\cdots\times D_{c_n} \to D_c$ such that for all $(i,\sigma,u_1,\ldots,u_n)$ in $I\times I^k\times D_{c_1}\times\cdots\times D_{c_n}$ there is y_1 in $V_{\rho(\beta,k)}$ with $\psi_{\beta,k+1}(y_1)$ and $y(i,\sigma,u_1,\ldots,u_n) = y_1(i,\sigma,i,u_1,\ldots,u_n)$.

- For α the downwards Vlach operator $\downarrow\beta$ where $\beta : c$, the bounded formula $\psi_{\alpha,0}(y)$ simply is $\psi_{\beta,0}(y)$, while the bounded formula $\psi_{\alpha,k+1}(y)$ says that $y : I \times I^k \times D_{c_1} \times \cdots \times D_{c_n} \to D_c$ such that for all $(i,\sigma,u_1,\ldots,u_n)$ in $I\times I^k\times D_{c_1}\times\cdots\times D_{c_n}$ there is y_1 in $V_{\rho(\beta,k)}$ with $\psi_{\beta,k}(y_1)$ and if $\sigma = \tau j$ then $y(i,\sigma,u_1,\ldots,u_n) = y_1(j,\tau,u_1,\ldots,u_n)$.

The result in (3.17) holds by the inspection of the definition of the bounded formulas ψ_c, together with the absoluteness results described in §3.4.1.

Suppose, as in the last section, that we have an index set J and a sequence of models $\mathcal{M}_j = \langle I^j, B^j, D^j, T^j\rangle$, and suppose that U is an ultrafilter on J. In the last section we described the ultraproduct of frames $\prod_U \langle I^j, B^j, D^j, T^j\rangle$. We now define the ultraproduct of models. For each well-formed expression α and $k \geq 0$, define $\Sigma_{\alpha,k} : J \to V(X)$ by $\Sigma_{\alpha,k}(j) = [\![\alpha]\!]^k_{\mathcal{M}_j}$.

Then $[\Sigma_{\alpha,k}]_U$ is in $A_{\rho(\alpha,k)}(X)$. Then define its image under $\pi : V(X) \to V(Y)$ as follows:

$$[\![\alpha]\!]^k_{\mathcal{M}} = \pi([\Sigma_{\alpha,k}]_U) \tag{3.18}$$

Since each $\mathcal{M}_j$ is a model, for each j in J we have that $V(X) \vDash \psi_{\alpha,k}([\![\alpha]\!]^k_{\mathcal{M}_j}, \overline{B}^j_{\cdot,a}, \overline{D}^j_a, \overline{T}^j_{\cdot,a})$. Then by the left-to-right direction of (3.10) we have that $V(Y) \vDash \psi_{\alpha,k}([\![\alpha]\!]^k_{\mathcal{M}}, \overline{B}_{\cdot,a}, \overline{D}_a, \overline{T}_{\cdot,a})$. Then by the right-to-left direction of (3.17), we have that $[\![\alpha]\!]^k_{\mathcal{M}}$ as defined in (3.18) does indeed define a model $\mathcal{M}$, and in particular we can set:

$$[\![\alpha]\!]_{\mathcal{M},i,\sigma,g} = [\![\alpha]\!]^{|\sigma|}_{\mathcal{M}}(i, \sigma, g(v_1), \ldots, g(v_n)) \tag{3.19}$$

We call this *the ultraproduct of the models*, and write it as $\mathcal{M} = \prod_U \mathcal{M}_j$.

We say that *theory* in the intensional theory of types is a collection of well-formed expressions whose main connective is equality. We say that a model $\mathcal{M}$ *satisfies* a theory if for each α in the theory we have $[\![\alpha]\!]_{\mathcal{M},i,\sigma,g} = 1$. A theory is *finitely satisfiable* if each finite subset of it is satisfiable. Then we have, as a result of our work in this section:

Theorem 3.3. *(Compactness) Every finitely satisfiable theory is satisfiable.*

Proof. This is by induction on the cardinality of the theory. It holds by hypothesis when the theory is itself finite. Suppose that the theory has infinite cardinality κ. Enumerate it as α_γ for $\gamma < \kappa$. Let U be an ultrafilter extending the generalized Fréchet filter $\{X \subseteq \kappa : |\kappa \setminus X| < \kappa\}$. For each $\theta < \kappa$ the theory $\{\alpha_\gamma : \gamma < \theta\}$ has cardinality $< \kappa$ and hence by induction hypothesis has a model $\mathcal{M}_\theta$. Consider the ultraproduct $\prod_U \mathcal{M}_\theta$. Then for each $\gamma < \kappa$ one has that $\mathcal{M}_\theta$ satisfies α_γ for all $\theta > \gamma$. Hence for U-many θ one has that $\mathcal{M}_\theta$ satisfies α_γ. Hence $\prod_U \mathcal{M}_\theta$ satisfies α_γ, and hence satisfies the theory. $\square$

3.5 Stalnaker-like frames and models

The aim of this section is to view Stalnaker's (2012) approach to modeling contingent propositions through the lens of the semantics described above. While Stalnaker himself worked at the level of propositional modal logic, our approach is of course type-theoretic. Following Montague, the intension operator $^\frown\varphi$ gives us a well-formed expression for the proposition expressed by φ. Part of our focus in what follows is on the extent to which $^\frown\varphi$ exists at a world, in the models of the kind envisioned by Stalnaker. On an austere rendering of the Stalnakrian idea, what we find is that this existence depends heavily on the absence of occurrences of the Montagovian extension operator $\smile$ inside the expression φ. Hence, in this austere setting, the only explicitly namable contingent propositions are those which involve the Montagovian extension operator.

Throughout §3.5, we work with universal accessibility relations, and so omit explicit mention of them.

3.5.1 Stalnaker-like frames and models

A base frame $\langle I, B, D, S\rangle$ is *Stalnaker-like* if for each i in I there is an equivalence relation $E_i \subseteq I \times I$ such that $B_{i,s} = {}^{I}\!/_{E_i}$ and $S_{i,s} : B_{i,s} \to D_s$ just selects a representative of each equivalence class. Hence the base frame is necessarily injective on type s.

Suppose that a frame $\langle I, B, D, S, T\rangle$ is such that its underlying base frame $\langle I, B, D, S\rangle$ is Stalnaker-like with associated equivalence relations E_i. Then note that the condition in (*frame:1s*) is trivially met. For, suppose b is a type and $f : B_{i,s} \to D_b$ and x_0, x_1 are from $B_{i,s}$. Since the x_0, x_1 from $B_{i,s}$ are equivalence classes $[y_0]_{E_i}$ and $[y_1]_{E_i}$ and the map $T_{i,s} = S_{i,s}$ just chooses representatives, the supposition that $T_{i,s}([y_0]_{E_i}) = T_{i,s}([y_1]_{E_i})$ implies by injectivity that $[y_0]_{E_i} = [y_1]_{E_i}$, and so f necessarily acts the same on them. Hence, in the setting of

Stalnaker-like frames, condition (*frame:1s*) simply says that $B_{i,\langle s,b\rangle}$ is some subset of the functions $f : B_{i,s} \to D_b$. In the case of full Stalnaker-like frames, it is of course the set of all such functions.

A frame $\langle I, B, D, S, T\rangle$ is *Stalnaker-like* if the base frame $\langle I, B, D, S\rangle$ is Stalnaker-like with associated equivalence relations E_i and for all types b and all $f : B_{i,s} \to D_b$, one has that $T_{i,\langle s,b\rangle}(f)(j) = f([j]_{E_i})$. This is just to insist that the way that $f : {}^{I}/_{E_i} \to D_b$ is extended to $T_{i,\langle s,b\rangle}(f) : I \to D_b$ is as follows: on an input of an index j look to how f acted on the index's equivalence class $[j]_{E_i}$. This goes above and beyond (*frame:2s*) since this only stipulates how $T_{i,\langle s,b\rangle}(f)$ must act on indexes which are representatives of their equivalence classes, while of course many indexes do not get selected to be representatives.

In what follows, we write Stalnaker-like base frames (resp. frames) as $\langle I, E, B, D, S\rangle$ (resp. $\langle I, E, B, D, S, T\rangle$) where E is the family of equivalence relations E_i associated to the base frame (resp. frame). The equivalence relations E_i should not be confused with the accessibility relation, which again throughout §3.5 we assume to be universal.

We can extend Proposition 3.1 as follows:

Proposition 3.4.

- *Every Stalnaker-like base frame can be extended to a Stalnaker-like full minimal frame.*
- *If the Stalnaker-like base frame injective (resp. and/or cohesive), then it can be extended to an injective (resp. and/or cohesive) Stalnaker-like full frame.*
- *No Stalnaker-like base frame and hence no Stalnaker-like frame is constant if the associated equivalence relation is not the identity function.*

Proof. We simply replace (3.2) in the proof of Proposition 3.1 with

$$T_{i,\langle s,b\rangle}(f)(y) = f([y]_{E_i}) \tag{3.20}$$

This is minimal since if $T_{i,\langle s,b\rangle}(f)$ and $T_{i,\langle s,b\rangle}(g)$ agree on $T_{i,s}(B_{i,s})$, then they agree on a representative from each E_i-equivalence class. This takes care of the argument for the first part of the proposition and the injectivity aspect of the second part. The cohesiveness aspect follows by a minor modification of the argument in Proposition 3.1.

Second, suppose that f is an element of both $B_{i,\langle s,b\rangle}$ and $B_{j,\langle s,d\rangle}$. We must show that both $T_{i,\langle s,b\rangle}(f)$ and $T_{j,\langle s,d\rangle}(f)$ as defined in (3.20) are equal. Since f is an element of both $B_{i,\langle s,b\rangle}$ and $B_{j,\langle s,d\rangle}$, we know that its domain is $B_{i,s} = B_{j,s}$. But this is just to say $I/E_i = I/E_j$. Hence we have that $E_i = E_j$ and thus

$$T_{i,\langle s,b\rangle}(f)(y) = f([y]_{E_i}) = f([y]_{E_j}) = T_{j,\langle s,b\rangle}(f)(y) \tag{3.21}$$

□

A model $\mathcal{M}$ is *Stalnaker-like* if its underlying frame is Stalnaker-like.

The following two propositions give simple sufficient conditions for the existence of very elementary kinds of intensions in Stalnaker-like models:

Proposition 3.5. *Suppose that $\mathcal{M}$ is a Stalnaker-like model with underlying full Stalnaker-like frame. Suppose that a is a type and i is an index. Let $F : I \to D_a$ be a function such that for all indexes j, k with jE_ik one has $F(j) = F(k)$. Then F is in $T_{i,\langle s,a\rangle}(B_{i,\langle s,a\rangle})$. Hence for any indexes i, σ and variable assignment g and any well-formed expression φ of type $\langle s,a\rangle$ with $[\![\varphi]\!]_{\mathcal{M},i,\sigma,g} = F$, one has that φ exists at i, σ, g in $\mathcal{M}$.*

Proof. Define $f : B_{i,s} \to D_a$ by $f([j]_{E_i}) = F(j)$, which is well-defined by hypothesis and which is in $B_{i,\langle s,a\rangle}$ by fullness of the frame. Then since the frame is Stalnaker-like one has that $T_{i,\langle s,a\rangle}(f)(j) = f([j]_{E_i}) = F(j)$ for all indexes j. Hence $T_{i,\langle s,a\rangle}(f) = F$. Hence F is in $T_{i,\langle s,a\rangle}(B_{i,\langle s,a\rangle})$. □

Proposition 3.6. *Suppose that $\mathcal{M}$ is a Stalnaker-like model with underlying full Stalnaker-like frame. Suppose that a is a type, Φ is a well-formed expression of type $\langle a,b\rangle$ and φ is a well-formed expression of type a. Suppose that i,σ are indexes and g is a variable assignment. Suppose that $^\frown\Phi$ exists at i,σ,g and $^\frown\varphi$ exists at i,σ,g. Then $^\frown\Phi(\varphi)$ exists at i,σ,g.*

Proof. Suppose that $[\![{}^\frown\Phi]\!]_{\mathcal{M},i,\sigma,g} = T_{i,\langle s,\langle a,b\rangle\rangle}(\Gamma)$ where $\Gamma : B_{i,s} \to D_{\langle a,b\rangle}$. Suppose that $[\![{}^\frown\varphi]\!]_{\mathcal{M},i,\sigma,g} = T_{i,\langle s,a\rangle}(\gamma)$ where $\gamma : B_{i,s} \to D_a$. Define $\Delta : B_{i,s} \to D_b$ by $\Delta(y) = (\Gamma(y))(\gamma(y))$. Then one has the following for all indexes j, where we abbreviate $y = [j]_{E_i}$:

$$
\begin{aligned}
&[\![{}^\frown\Phi(\varphi)]\!]_{\mathcal{M},i,\sigma,g}(j) && \\
&= [\![\Phi(\varphi)]\!]_{\mathcal{M},j,\sigma,g} && \text{by semantics of } {}^\frown \\
&= ([\![\Phi]\!]_{\mathcal{M},j,\sigma,g})([\![\varphi]\!]_{\mathcal{M},j,\sigma,g}) && \text{by semantics for application} \\
&= (([\![{}^\frown\Phi]\!]_{\mathcal{M},i,\sigma,g})(j))(([\![{}^\frown\varphi]\!]_{\mathcal{M},i,\sigma,g})(j)) && \text{by semantics of } {}^\frown \\
&= (T_{i,\langle s,\langle a,b\rangle\rangle}(\Gamma)(j))(T_{i,\langle s,a\rangle}(\gamma)(j)) && \text{by assumptions on } \Gamma,\gamma \\
&= (\Gamma(y))(\gamma(y)) && \text{by Stalnaker-like and } y = [j]_{E_i} \\
&= \Delta(y) && \text{by defn of } \Delta \\
&= T_{i,\langle s,b\rangle}(\Delta)(j) && \text{by Stalnaker-like}
\end{aligned}
$$

Hence $^\frown\Phi(\varphi)$ exists at i,σ,g. □

By contrast, the following is a simple sufficient condition for the non-existence of a very elementary kind of proposition. In this result, we are using the elementary set-theory notation 2^κ for the cardinality of the powerset of κ, and we are using $|\cdot|$ for the cardinality operation taking a set to its cardinality. The hypothesis of this result is essentially the supposition that the equivalence relation E_i is very coarse, and cuts down the cardinality of the worlds from the large value $\kappa = |I|$ to the small value $\lambda = |B_{i,s}| = |{}^I/_{E_i}|$ in such a way that the number of propositions is thereby also lessened.

Proposition 3.7. *Suppose that one has a full Stalnaker-like frame with index i such that $2^{|B_{i,s}|} < 2^{|I|}$. Then for any type a, and any constant symbols C of type $\langle a,t\rangle$ and c of type a, there is a Stalnaker-like model $\mathcal{M}$ with this underlying frame such that $^\frown(Cc)$ does not exist at i and yet $c, {}^\frown c$ exists at i.*

Proof. Since we are dealing with closed well-formed expressions C and c with no Vlach operators, we omit the storage indexes and the variable assignments in this proof.

Let i be a fixed index, which we are thinking about as the index mentioned in the hypothesis, although we will not invoke the constraint $2^{|B_{i,s}|} < 2^{|I|}$ until the last paragraph. First let us indicate how to interpret c. Choose fixed element γ_i in $B_{i,a}$ and set $\Gamma_i = T_{i,a}(\gamma_i)$, so that Γ_i is in $D_a \cap T_{i,a}(B_{i,a})$ and set $c^{\mathcal{M}}(j) = \Gamma_i$. This has the effect that $c^{\mathcal{M}}(j)$ is constant and does not depend on j. Then $[\![c]\!]_{\mathcal{M},i} = c^{\mathcal{M}}(i) = \Gamma_i \in T_{i,a}(B_{i,a})$ and hence c exists at i. Define δ_i in $B_{i,\langle s,a\rangle}$ by $\delta_i([j]_{E_i}) = \Gamma_i$, so that for all j in I we have:

$$T_{i,\langle s,a\rangle}(\delta_i)(j) = \delta_i([j]_{E_i}) = \Gamma_i = c^{\mathcal{M}}(j) = [\![c]\!]_{\mathcal{M},j} = [\![{}^\frown c]\!]_{\mathcal{M},i}(j)$$

Hence $T_{i,\langle s,a\rangle}(\delta_i) = [\![{}^\frown c]\!]_{\mathcal{M},i}$ and thus $^\frown c$ exists at i.

With this same i fixed, choose f_i, g_i in $B_{i,\langle a,t\rangle}$ such that $f_i(\gamma_i) = 1$ and $g_i(\gamma_i) = 0$, and then set $F_i = T_{i,\langle a,t\rangle}(f_i)$ and $G_i = T_{i,\langle a,t\rangle}(g_i)$, noting that

$$\begin{aligned} F_i(\Gamma_i) &= (T_{i,\langle a,t\rangle}(f_i))(T_{i,a}(\gamma_i)) = (T_{i,t})(f_i(\gamma_i)) = f_i(\gamma_i) = 1 \\ G_i(\Gamma_i) &= (T_{i,\langle a,t\rangle}(g_i))(T_{i,a}(\gamma_i)) = (T_{i,t})(g_i(\gamma_i)) = g_i(\gamma_i) = 0 \end{aligned} \tag{3.22}$$

In these equations, the antepenultimate identity is just due to (*frame:2*), and the penultimate identity is just the fact that $T_{i,t}$ is the identity.

Second, let us say how to interpret C of type $\langle a,t\rangle$. Since our hypothesis on our fixed index i is that $2^{|B_{i,s}|} < 2^{|I|}$, the set $T_{i,\langle s,t\rangle}(B_{i,\langle s,t\rangle})$ is a proper subset of $D_{\langle s,t\rangle}$. Choose $H : I \to \{0,1\}$ which is in the latter but not the former. Then set $C^{\mathcal{M}}(j) = F_i$ if $H(j) = 1$, and $C^{\mathcal{M}}(j) = G_i$ if $H(j) = 0$. These definitions and (3.22) suffice to obtain the last identity in the following:

$$[\![{}^\frown Cc]\!]_{\mathcal{M},i}(j) = [\![Cc]\!]_{\mathcal{M},j} = [\![C]\!]_{\mathcal{M},j}([\![c]\!]_{\mathcal{M},j}) = (C^{\mathcal{M}}(j))(\Gamma_i) = H(j)$$

Then by the choice of H, one has that ${}^\frown Cc$ does not exist at index i. □

The following two propositions illustrate that the presence of the extension operator is an obstacle to the existence of the intension of a well-formed expression. Moreover, the hypotheses of the below result are so weak that in many natural models one will have that these intensions are necessarily non-existent.

In the below proposition, the intensions are literally are the semantic values of the determiners "some," "all" and "the" in the traditional Montagovian framework:[16]

Proposition 3.8. *Suppose Y, X are variables of type $b = \langle s, \langle e,t\rangle\rangle$ and x is variable of type e. Suppose α is $\lambda Y \lambda X\ \exists\ x\ ({}^\smile Y(x) \wedge {}^\smile X(x))$, and β is $\lambda Y \lambda X\ \forall\ x\ ({}^\smile Y(x) \to {}^\smile X(x))$, and γ is $\lambda Y \lambda X\ \exists\,!\ x\ {}^\smile Y(x) \wedge \forall\ x\ ({}^\smile Y(x) \to {}^\smile X(x))$. Suppose $\mathcal{M}$ is a Stalnaker-like model, and that there are indexes i,j,k with jE_ik and $j \neq k$. Then ${}^\frown\alpha$ and ${}^\frown\beta$ and ${}^\frown\gamma$ do not exist at i in $\mathcal{M}$.*

Proof. These three well-formed expressions have type $a = \langle b, \langle b, e\rangle\rangle$, and hence the type of their intension is $\langle s, a\rangle$. For B, A in D_b, one has $([\![{}^\frown\alpha]\!]_{\mathcal{M},i}(j))(B)(A) = 1$ iff $B(j), A(j)$ are not disjoint. Likewise, one has $([\![{}^\frown\beta]\!]_{\mathcal{M},i}(j))(B)(A) = 1$ iff $B(j) \subseteq A(j)$, while $([\![{}^\frown\gamma]\!]_{\mathcal{M},i}(j))(B)(A) = 1$ iff both $|B(j)| = 1$ and $B(j) \subseteq A(j)$. Here, we are viewing elements of type $D_{\langle e,t\rangle}$ as subsets of D_e, in the familiar way.

[16]See Gamut (1991, p. 114).

We focus on α, since the arguments for β and γ are similar. Suppose that $^\frown\alpha$ exists at i in $\mathcal{M}$. Then $[\![^\frown\alpha]\!]_{\mathcal{M},i} = T_{i,\langle s,a\rangle}(f)$ where f is in $B_{i,\langle s,a\rangle}$. But then since jE_ik we have

$$[\![^\frown\alpha]\!]_{\mathcal{M},i}(j) = T_{i,\langle s,a\rangle}(f)(j) = f([j]_{E_i}) = f([k]_{E_i}) = T_{i,\langle s,a\rangle}(f)(k) = [\![^\frown\alpha]\!]_{\mathcal{M},i}(k)$$

But then we would have that $B(j), A(j)$ are not disjoint iff $B(k), A(k)$ are not disjoint. But since $j \neq k$ it is easy to choose $A, B : I \to D_{\langle e,t\rangle}$ such that $B(j), A(j)$ are not disjoint while $B(k), A(k)$ are disjoint. □

In this proposition, we show the same is true of the composition of the intension and extension operators on which the operators do not cancel out:

Proposition 3.9. *Suppose that b is a type, and that c is a constant of type $\langle s,b\rangle$. Suppose that $\langle I, E, B, D, S, T, \rangle$ is a full Stalnaker-like frame with indexes i, j, k with jE_ik and $j \neq k$. Then there is an Stalnaker-like model above this frame such that $^\frown{}^\smile c$ does not exist at i.*

Proof. Pick a function $H : I \to D_b$ with $H(j) \neq H(k)$. Define $c^{\mathcal{M}} : I \to D_b^I$ so that $c^{\mathcal{M}}$ always outputs H.

Note that while $c^{\mathcal{M}}(j) = H = c^{\mathcal{M}}(k)$, we nonetheless have $c^{\mathcal{M}}(j)(j) = H(j) \neq H(k) = c^{\mathcal{M}}(k)(k)$. Since $[\![^\frown{}^\smile c]\!]_{\mathcal{M},i}(j) = [\![^\smile c]\!]_{\mathcal{M},j} = [\![c]\!]_{\mathcal{M},j}(j) = c^{\mathcal{M}}(j)(j)$, and similarly for k, we then have $[\![^\frown{}^\smile c]\!]_{\mathcal{M},i}(j) \neq [\![^\frown{}^\smile c]\!]_{\mathcal{M},i}(k)$.

Suppose that $^\frown{}^\smile c$ does exist at i. Then there is f in $B_{i,\langle s,b\rangle}$ with $[\![^\frown{}^\smile c]\!]_{\mathcal{M},i}(j) = T_{i,\langle s,a\rangle}(f)(j) = f([j]_{E_i})$. But then we would have $[\![^\frown{}^\smile c]\!]_{\mathcal{M},i}(j) = f([j]_{E_i}) = f([k]_{E_i}) = [\![^\frown{}^\smile c]\!]_{\mathcal{M},i}(k)$, contradicting the conclusion of the previous paragraph. □

3.5.2 Austerely Stalnaker-like frames and models

A Stalnaker-like base frame $\langle I, E, B, D, S\rangle$ is *austere* if for all indexes i, j, k, if jE_ik then both of the following hold

$$E_j = E_k \qquad (\textit{austere:1})$$

$$S_{j,a}(B_{j,a}) = S_{k,a}(B_{k,a}) \qquad (\textit{austere:2})$$

where a ranges over basic extended types. A Stalnaker-like frame $\langle I, E, B, D, S, T\rangle$ is *austere* if its base frame is austere and if (*austere:2*) holds for all extended types a, wherein the S-maps are replaced by the T-maps.

Proposition 3.10.

- *Every austerely Stalnaker-like base frame can be extended to an austerely Stalnaker-like full frame.*
- *If the austerely Stalnaker-like base frame injective (resp. and/or cohesive), then it can be extended to an injective (resp. and/or cohesive) austerely Stalnaker-like full frame.*

Proof. We simply verify that austerity holds for the construction in Proposition 3.1 and Proposition 3.4. In the proof of Proposition 3.1, we fixed well-orders on the outer domains as we went along; in this proof, we also fix well-orderings of the inner domains as we go along.

Suppose that austerity holds for type a, b; we show it holds for type $\langle a, b\rangle$. Suppose that jE_ik. The induction hypothesis is that $T_{j,a}(B_{j,a}) = T_{k,a}(B_{k,a})$ and $T_{j,b}(B_{j,b}) = T_{k,b}(B_{k,b})$; and we show that $T_{j,\langle a,b\rangle}(B_{j,\langle a,b\rangle}) = T_{k,\langle a,b\rangle}(B_{k,\langle a,b\rangle})$. It suffices to show that any F in $T_{j,\langle a,b\rangle}(B_{j,\langle a,b\rangle})$ is also in $T_{k,\langle a,b\rangle}(B_{k,\langle a,b\rangle})$. So suppose that $F = T_{j,\langle a,b\rangle}(f)$ as defined in (3.1), for some f in $B_{j,\langle a,b\rangle}$. We define g in $B_{k,\langle a,b\rangle}$ such that $F = T_{k,\langle a,b\rangle}(g)$. It is helpful to then keep the following diagram in mind: by hypothesis we have the left-most commuting square as in (*frame:2*), and we want to build the right-most square so that it too commutes and comports with (*frame:2*):

$$\begin{tikzcd} B_{j,a} \arrow[r, "T_{j,a}"] \arrow[d, "f"'] & D_a \arrow[d, "F"] & B_{k,a} \arrow[l, "T_{k,a}"'] \arrow[d, "g"] \\ B_{j,b} \arrow[r, "T_{j,b}"] & D_b & B_{k,b} \arrow[l, "T_{k,b}"'] \end{tikzcd} \tag{3.23}$$

To define $g : B_{k,a} \to B_{k,b}$, suppose that x in $B_{k,a}$ is given. Let $y = T_{k,a}(x)$ which is in $T_{k,a}(B_{k,a}) = T_{j,a}(B_{j,a})$. Then $F(y)$ is defined as in the first case break in (3.1), so that it is a member of $T_{j,b}(B_{j,b}) = T_{k,b}(B_{k,b})$. Using the well-ordering of $B_{k,b}$, choose $g(x)$ to be an element of $T_{k,b}^{-1}(F(y))$. Note that the definition of $g(x)$ depends only on $F(y) = F(T_{k,a}(x))$, and so g meets the condition in (*frame:1*) for being an element of $B_{k,\langle a,b\rangle}$. Further, by the first case break in the definition of $T_{k,\langle a,b\rangle}(g)$ in (3.1), we see that by construction $T_{k,\langle a,b\rangle}(g)(y) = F(y)$ for y in $T_{k,a}(B_{k,a})$, so that the right-most square in (3.23) commutes. This implies that

$$T_{j,b}(f(B_{j,a})) = F(T_{j,a}(B_{j,a})) = F(T_{k,a}(B_{k,a})) = T_{k,b}(g(B_{k,a}))$$

Then the definition of $T_{k,\langle a,b\rangle}(g)$ on the second case break in (3.1) will be the same as the definition of $T_{k,\langle a,b\rangle}(f)$.

Suppose that austerity holds for type b; we show it holds for type $\langle s,b\rangle$. Suppose that jE_ik. Then the austerity of the base frame requires that $E_j = E_k$ and hence $B_{j,s} = B_{k,s}$ and thus $B_{j,\langle s,b\rangle} = B_{k,\langle s,b\rangle}$. Suppose that f is in this set. Then for all indexes k' we have, per the construction in Proposition 3.4, that:

$$T_{j,\langle s,b\rangle}(f)(k') = f([k']_{E_j}) = f([k']_{E_k}) = T_{k,\langle s,b\rangle}(f)(k')$$

Hence we have $T_{j,\langle s,b\rangle}(B_{j,\langle s,b\rangle}) = T_{k,\langle s,b\rangle}(B_{k,\langle s,b\rangle})$. ☐

Suppose that $\mathcal{M}$ is a model with underlying frame $M = \langle I, E, B, D, S, T\rangle$ which is austerely Stalnaker-like. Then we say that $\mathcal{M}$ is *austerely Stalnaker-like* if for all indexes i, j, k one has that jE_ik implies the following, for all constant symbols c:

$$c^{\mathcal{M}}(j) = c^{\mathcal{M}}(k) \tag{\textit{austere:3}}$$

As one can see by the second part of the following proposition, the model constructed in the Proposition 3.7 does not satisfy the (*austere:3*) condition. The first part of the following proposition is needed to establish the second part, and in the first part we use the equivalence relation $\equiv_{j,k}$ on finite sequences of indexes from the end of §3.3.2 to handle the case of the Vlach operators.

Proposition 3.11. *Suppose that $\mathcal{M}$ is austerely Stalnaker-like model with $\langle I, E, B, D, S, T\rangle$ a full Stalnaker-like frame.*

- *For all well-formed expressions φ of type a which do not include the extension operator, and for all indexes i, j, k with jE_ik, and for all finite sequences of indexes $j'\sigma, k'\tau$ with $j'\sigma \equiv_{j,k} k'\tau$, and for all variable assignments g, one has $[\![\varphi]\!]_{\mathcal{M},j',\sigma,g} = [\![\varphi]\!]_{\mathcal{M},k',\tau,g}$.*
- *For every well-formed expression φ of type a which does not include the extension operator and all indexes i, σ and every variable assignment g, one has that $^\frown\varphi$ exists at i, σ, g in $\mathcal{M}$.*

Proof. The first item is by induction on the complexity of φ, with the induction statement having universal quantifiers over the indexes and the assignments.

First suppose that φ is a variable of type a. Then one has $[\![\varphi]\!]_{\mathcal{M},j',\sigma,g} = [\![v]\!]_{\mathcal{M},j',\sigma,g} = g_a(v) = [\![v]\!]_{\mathcal{M},k',\tau,g} = [\![\varphi]\!]_{\mathcal{M},k',\tau,g}$.

Second suppose that φ is a constant of type a. Then it follows directly from (*austere:3*).

Third suppose that φ is $\alpha(\beta)$, where α is of type $\langle a, b\rangle$ and β is of type a. Further suppose that the result holds for α, β as an induction hypothesis. Then we show that the result holds for $\alpha(\beta)$. But one has $[\![\varphi]\!]_{\mathcal{M},j',\sigma,g} = [\![\alpha]\!]_{\mathcal{M},j',\sigma,g}([\![\beta]\!]_{\mathcal{M},j',\sigma,g}) = [\![\alpha]\!]_{\mathcal{M},k',\tau,g}([\![\beta]\!]_{\mathcal{M},k',\tau,g}) = [\![\varphi]\!]_{\mathcal{M},k',\tau,g}$.

The induction steps for the lambda terms follows since the induction statement has a universal quantifier over assignments. Many induction steps, such as those of identity, the ordering on degrees, and the propositional connectives, are trivial. The induction steps for the quantifiers and the modalities follow since the induction statement has a universal quantifier over assignments, and by recourse to (*austere:2*).

The induction step for the intension operator follows since $[\![{}^\frown\varphi]\!]_{\mathcal{M},j',\sigma,g}(i') = [\![\varphi]\!]_{\mathcal{M},i',\sigma,g} = [\![\varphi]\!]_{\mathcal{M},i',\tau,g} = [\![{}^\frown\varphi]\!]_{\mathcal{M},k',\tau,g}(i')$, where the middle identity follows from the induction hypothesis and the fact that the equivalence $j'\sigma \equiv_{j,k} k'\tau$ trivially implies the equivalence $i'\sigma \equiv_{j,k} i'\tau$.

For the Vlach operators, we have $[\![\uparrow \varphi]\!]_{\mathcal{M},j',\sigma,g} = [\![\varphi]\!]_{\mathcal{M},j',\sigma j',g} = [\![\varphi]\!]_{\mathcal{M},k',\tau k',g} = [\![\uparrow \varphi]\!]_{\mathcal{M},k',\tau,g}$, where the middle identity follows from the induction hypothesis and the fact that the equivalence $j'\sigma \equiv_{j,k} k'\tau$ trivially implies $j'\sigma j' \equiv_{j,k} k'\tau k'$. The argument for the other Vlach operator is similar.

The proof of the second item is by induction on complexity of φ, again with the induction statement having universal quantifiers over the indexes and the assignments.

First suppose that v is a variable of type a. Then note that $[\![{}^\frown v]\!]_{\mathcal{M},i,\sigma,g}(j) = [\![v]\!]_{\mathcal{M},j,\sigma,g} = g_a(v)$, and so $[\![{}^\frown v]\!]_{\mathcal{M},i,\sigma,g}$ is a constant function and we are done by Proposition 3.5.

Second suppose that φ is a constant symbol c of type a. Define $f : B_{i,s} \to D_a$ by $f([j]_{E_i}) = c^{\mathcal{M}}(j)$, which is well-defined by (*austere:3*), and which is in $B_{i,\langle s,a\rangle}$ by the fullness condition. Then by definition of a Stalnaker-like frame we have that $T_{i,\langle s,a\rangle}(f)(j) = f([j]_{E_i}) = c^{\mathcal{M}}(j) = [\![{}^\frown c]\!]_{\mathcal{M},i,\sigma,g}(j)$, and since j was an arbitrary index we have that $T_{i,\langle s,a\rangle}(f) = [\![{}^\frown c]\!]_{\mathcal{M},i,\sigma,g}$. Suppose that α is a well-formed expression of type $\langle a, b\rangle$ and β is a well-formed expression of type a. Suppose by induction that the result holds for α, β. We show that the result holds for $\alpha(\beta)$, the well-formed expression of type b. Since $[\![{}^\frown\alpha]\!]_{\mathcal{M},i,\sigma,g}$ is in $T_{i,\langle s,\langle a,b\rangle\rangle}(B_{i,\langle s,\langle a,b\rangle\rangle})$, and since $[\![{}^\frown\beta]\!]_{\mathcal{M},i,\sigma,g}$ is in $T_{i,\langle s,a\rangle}(B_{i,\langle s,a\rangle})$, choose f' in $B_{i,\langle s,\langle a,b\rangle\rangle}$ and f in $B_{i,\langle s,b\rangle}$ such that $T_{i,\langle s,\langle a,b\rangle\rangle}(f') = [\![{}^\frown\alpha]\!]_{\mathcal{M},i,\sigma,g}$ and $T_{i,\langle s,a\rangle}(f) = [\![{}^\frown\beta]\!]_{\mathcal{M},i,\sigma,g}$. Define f'' in $B_{i,\langle s,b\rangle}$ by $f''(y) = (f'(y))(f(y))$ where y is an element of $B_{i,s}$; note by the fullness condition, we have that f'' is in $B_{i,\langle s,b\rangle}$.

We show that for all j in I, we have $T_{i,\langle s,a\rangle}(f'')(j) = [\![{}^\frown\alpha(\beta)]\!]_{\mathcal{M},i,\sigma,g}(j)$. Hence let j in I be given, and let $y = [j]_{E_i}$ and let $k = T_{i,s}(y)$, noting that since we are in a Stalnaker-like frame we have jE_ik. Then we argue as follows:

$T_{i,\langle s,b\rangle}(f'')(j)$	
$= f''(y)$	by Stalnaker-like and defn of y
$= (f'(y))(f(y))$	by defn of f''
$= (T_{i,\langle s,\langle a,b\rangle\rangle}(f')(T_{i,s}(y)))((T_{i,\langle s,a\rangle}(f))(T_{i,s}(y)))$	by (*frame:2s*)
$= (T_{i,\langle s,\langle a,b\rangle\rangle}(f')(k))((T_{i,\langle s,a\rangle}(f))(k))$	by defn of k
$= ([\![{}^\frown\alpha]\!]_{\mathcal{M},i,\sigma,g}(k))([\![{}^\frown\beta]\!]_{\mathcal{M},i,\sigma,g}(k))$	by defn f', f
$= ([\![\alpha]\!]_{\mathcal{M},k,\sigma,g})([\![\beta]\!]_{\mathcal{M},k,\sigma,g})$	by semantics of ${}^\frown$
$= ([\![\alpha(\beta)]\!]_{\mathcal{M},k,\sigma,g})$	by semantics for application
$= ([\![\alpha(\beta)]\!]_{\mathcal{M},j,\sigma,g})$	by first item
$= ([\![{}^\frown\alpha(\beta)]\!]_{\mathcal{M},i,\sigma,g})(j)$	by semantics of ${}^\frown$

This completes the inductive step for application.

For the lambda terms, suppose that v is a variable of type a and that φ is a well-formed expression of type b, and suppose for the induction hypothesis that the result holds for φ. Suppose that i is an index and that g is a variable assignment. Suppose that j, k are indexes with jE_ik. By the first part of the proposition, we have that $[\![\varphi]\!]_{\mathcal{M},j,\sigma,g[v/x]} = [\![\varphi]\!]_{\mathcal{M},k,\sigma,g[v/x]}$ for all x in D_a. Thus the function $f : B_{i,s} \to D_{\langle a,b\rangle}$ given by $f([j]_{E_i})(x) = [\![\varphi]\!]_{\mathcal{M},j,\sigma,g[v/x]}$ is well-defined. Then since the model is Stalnaker-like, we have that for all j in I and x in D_a

$$\begin{aligned}(T_{i,\langle s,\langle a,b\rangle\rangle}(f)(j))(x) &= f([j]_{E_i})(x) = [\![\varphi]\!]_{\mathcal{M},j,\sigma,g[v/x]} \\ &= ([\![\lambda v.\varphi]\!]_{\mathcal{M},j,\sigma,g})(x) = ([\![{}^\frown\lambda v.\varphi]\!]_{\mathcal{M},i,\sigma,g}(j))(x)\end{aligned}$$

Since this held for all j in I and x in D_a, we have then that $T_{i,\langle s,\langle a,b\rangle\rangle}(f) = [\![{}^\frown\lambda v.\varphi]\!]_{\mathcal{M},i,\sigma,g}$, and so indeed ${}^\frown\lambda v.\varphi$ exists at i.

For identity and the linear ordering on the degrees, the argument is similar to the propositional connectives, and for these we just give the argument for conjunction since it is representative. Suppose the result holds for φ and ψ of type t. By the induction hypothesis we have that $[\![{}^\frown\varphi]\!]_{\mathcal{M},i,\sigma,g} = T_{i,\langle s,t\rangle}(f_1)$ and $[\![{}^\frown\psi]\!]_{\mathcal{M},i,\sigma,g} = T_{i,\langle s,t\rangle}(f_2)$ for functions $f_1, f_2 : B_{i,s} \to D_t$. We want to find f_3 with $[\![{}^\frown(\varphi \wedge \psi)]\!]_{\mathcal{M},i,\sigma,g} = T_{i,\langle s,t\rangle}(f_3)$. Take $f_3(y) = \inf(f_1(y), f_2(y))$, so that $f_3 : B_{i,s} \to D_t$ and f_3 in $B_{i,\langle s,t\rangle}$ by fullness of the frame. Let j in I be given and let $y = [j]_{E_i}$. We have the following chain of identities:

$$\begin{aligned}
&(T_{i,\langle s,t\rangle}(f_3))(j) && \\
&= f_3(y) && \text{by Stalnaker-like and defn of } y\\
&= \inf(f_1(y), f_2(y)) && \text{by defn of } f_3\\
&= \inf((T_{i,\langle s,t\rangle}(f_1))(j), (T_{i,\langle s,t\rangle}(f_2))(j)) && \text{by Stalnaker-like and defn of } y\\
&= \inf([\![{}^\frown\varphi]\!]_{\mathcal{M},i,\sigma,g}(j), [\![{}^\frown\psi]\!]_{\mathcal{M},i,\sigma,g}(j)) && \text{by defn of } f_1, f_2\\
&= \inf([\![\varphi]\!]_{\mathcal{M},j,\sigma,g}, [\![\psi]\!]_{\mathcal{M},j,\sigma,g}) && \text{by semantics of } {}^\frown\\
&= [\![(\varphi \wedge \psi)]\!]_{\mathcal{M},j,\sigma,g} && \text{by semantics for conjunction}\\
&= [\![{}^\frown(\varphi \wedge \psi)]\!]_{\mathcal{M},i,\sigma,g}(j) && \text{by semantics of } {}^\frown
\end{aligned}$$

For the quantifiers, we do only the existential case. Suppose that v is a variable of type a and that ψ is a well-formed expression of type t which does not include the extension operator. Let φ be $\exists\ v\ \psi$, which is also of type t. The induction hypothesis is that the result holds for ψ. For the remainder of this paragraph, fix indexes i, σ and variable assignment g. Then by induction hypothesis, for all variable assignments h one has that $[\![{}^\frown\psi]\!]_{\mathcal{M},i,\sigma,h}$ is in $T_{i,\langle s,t\rangle}(B_{i,\langle s,t\rangle})$, and so we can write it as $[\![{}^\frown\psi]\!]_{\mathcal{M},i,\sigma,h} = T_{i,\langle s,t\rangle}(f_h)$ for some choice of $f_h : B_{i,s} \to D_t$. Define $f : B_{i,s} \to D_t$ by $f([j]_{E_i}) = 1$ iff there is v-variant h of g with $f_{i,\sigma,h}([j]_{E_i}) = 1$ and $h_a(v)$ in $T_{j,a}(B_{j,a})$. Note that f is well-defined by (*austere:2*); and f is in $B_{i,\langle s,t\rangle}$ by fullness of the frame. It suffices now to show that $[\![{}^\frown\varphi]\!]_{\mathcal{M},i,\sigma,g}(j) = 1$ iff $T_{i,\langle s,t\rangle}(f)(j) = 1$, for all indexes j. We do the left-to-right direction since the other direction is similar. Suppose that $[\![{}^\frown\varphi]\!]_{\mathcal{M},i,\sigma,g}(j) = 1$. Then there is v-variant h of g with both $[\![\psi]\!]_{\mathcal{M},j,\sigma,h} = 1$ and $h_a(v)$ in $T_{j,a}(B_{j,a})$. Then $[\![{}^\frown\psi]\!]_{\mathcal{M},i,\sigma,h}(j) = 1$ (noting that the subscript on the $[\![\ ]\!]$ is index i and not index j). Then by the choice of f_h we have that $T_{i,\langle s,t\rangle}(f_h)(j) = 1$. Then since the frame is Stalnaker-like we have that $f_h([j]_{E_i}) = 1$. Then by the definition of f we have that $f([j]_{E_i}) = 1$. Then since the frame is Stalnaker-like we have $T_{i,\langle s,t\rangle}(f)(j) = 1$, which is what we wanted to show.

For the modal operators, we do only the diamond case. Suppose that ψ is a well-formed expression of type t which does not include the extension operator. Let φ be $\Diamond\psi$, which is also of type t. The induction hypothesis is that the result holds for ψ. For the remainder of this proof, fix indexes i, σ and variable assignment g. Hence by the induction hypothesis for all indexes j' one has that $[\![{}^\frown\psi]\!]_{\mathcal{M},j',\sigma,g}$ is in $T_{j',\langle s,t\rangle}(B_{j',\langle s,t\rangle})$, and so we can write it as $[\![{}^\frown\psi]\!]_{\mathcal{M},j',\sigma,g} = T_{j',\langle s,t\rangle}(f_{j'})$ for some choice of $f_{j'} : B_{j',s} \to D_t$. Define $f : B_{i,s} \to D_t$ by $f([j]_{E_i}) = 1$ iff there is j' in $T_{j,s}(B_{j,s})$ with $f_{j',\sigma,g}([j']_{E_{j'}}) = 1$. Note that f is well-defined by (*austere:2*); and f is in $B_{i,\langle s,t\rangle}$ by fullness of the frame. It suffices now to show that $[\![{}^\frown\varphi]\!]_{\mathcal{M},i,\sigma,g}(j) = 1$ iff $T_{i,\langle s,t\rangle}(f)(j) = 1$ for all indexes j. We do the left-to-right direction since the other direction is similar. Suppose that $[\![{}^\frown\varphi]\!]_{\mathcal{M},i,\sigma,g}(j) = 1$, or what is the same that $[\![\Diamond\psi]\!]_{\mathcal{M},j,\sigma,g} = 1$. Then there is j' in $T_{j,s}(B_{j,s})$ such that $[\![\psi]\!]_{\mathcal{M},j',\sigma,g} = 1$. Then

$[\![{}^\frown\psi]\!]_{\mathcal{M},j',\sigma,g}(j') = 1$. Then by choice of $f_{j'}$ we have $T_{j',\langle s,t\rangle}(f_{j',g})(j') = 1$. Then since the frame is Stalnaker-like we have $f_{j',g}([j']_{E_{j'}}) = 1$. Then by definition of f, we have $f([j]_{E_i}) = 1$. Then since the frame is Stalnaker-like, we have $T_{i,\langle s,t\rangle}(f)(j) = 1$, which is what we wanted to show.

Finally, we show that the result holds when φ is ${}^\frown\psi$. Suppose that ψ has type a, so that ${}^\frown\psi$ has type $b = \langle s,a\rangle$. By induction hypothesis, $[\![{}^\frown\psi]\!]_{\mathcal{M},i,\sigma,g}$ in $T_{i,\langle s,a\rangle}(B_{i,\langle s,a\rangle})$. Choose γ_i in $B_{i,\langle s,a\rangle}$ such that $[\![{}^\frown\psi]\!]_{\mathcal{M},i,\sigma,g} = T_{i,\langle s,a\rangle}(\gamma_i)$, and set Γ_i equal to this quantity, which is an element of D_b. But then we have $\Gamma_i = \Gamma_{i'}$ for all indexes i, i', since

$$\Gamma_i(j) = [\![{}^\frown\psi]\!]_{\mathcal{M},i,\sigma,g}(j) = [\![\psi]\!]_{\mathcal{M},j,\sigma,g} = [\![{}^\frown\psi]\!]_{\mathcal{M},i',\sigma,g}(j) = \Gamma_{i'}(j)$$

Hence, let $\Gamma = \Gamma_i$ for any index i. Now, as noted in the proof of Proposition 3.1 and hence Proposition 3.4, $B_{i,\langle s,b\rangle}$ contains all constant functions, and so let f in $B_{i,\langle s,b\rangle}$ be a function with constant value Γ. Then one has

$$T_{i,\langle s,b\rangle}(f)(j) = f([j]_{E_i}) = \Gamma = \Gamma_j = [\![{}^\frown\psi]\!]_{\mathcal{M},j,\sigma,g} = [\![{}^\frown\varphi]\!]_{\mathcal{M},i,\sigma,g}(j)$$

For the Vlach operators, suppose that φ is an expression of type a. Then by the first part of the proposition one has that if jE_ik then

$$[\![{}^\frown\!\uparrow\varphi]\!]_{\mathcal{M},i,\sigma,g}(j) = [\![\varphi]\!]_{\mathcal{M},j,\sigma j,g} = [\![\varphi]\!]_{\mathcal{M},k,\sigma k,g} = [\![{}^\frown\!\uparrow\varphi]\!]_{\mathcal{M},i,\sigma,g}(k)$$

Hence, the function $f : B_{i,s} \to D_a$ defined by $f([j]_{E_i}) = [\![{}^\frown\!\uparrow\varphi]\!]_{\mathcal{M},i,\sigma,g}(j)$ is well-defined. Hence for all indexes j we have

$$T_{i,\langle s,a\rangle}(f)(j) = f([j]_{E_i}) = [\![{}^\frown\!\uparrow\varphi]\!]_{\mathcal{M},i,\sigma,g}(j)$$

And thus $T_{i,\langle s,a\rangle}(f) = [\![{}^\frown\!\uparrow\varphi]\!]_{\mathcal{M},i,\sigma,g}$.

For the other Vlach operator, first note that we have the following, for any indexes $i, \sigma k$:

$$[\![{}^\frown\!\downarrow\varphi]\!]_{\mathcal{M},i,\sigma k,g}(j) = [\![\downarrow\varphi]\!]_{\mathcal{M},j,\sigma k,g} = [\![\varphi]\!]_{\mathcal{M},k,\sigma,g}$$

$$[\![{}^\frown\!\downarrow\varphi]\!]_{\mathcal{M},i,\varnothing,g}(j) = [\![\downarrow\varphi]\!]_{\mathcal{M},j,\varnothing,g} = [\![\varphi]\!]_{\mathcal{M},j,\varnothing,g} = [\![{}^\frown\varphi]\!]_{\mathcal{M},i,\varnothing,g}(j)$$

In the first case, once $i, \sigma k$ are fixed we have a constant function, and are done by Proposition 3.5. In the second case, we can appeal directly to the induction hypothesis for φ. □

But note that by Proposition 3.8, this problem is not specific to austere frames and models, but pertains to all Stalnaker-like models.

3.6 Direct limits

3.6.1 Relevant aspects of the direct limit construction

Recall that a *directed partial order* is a partial order $(I, \leq)$ which has the directedness property: for any two elements i, j in I there is element k in I with $k \geq i, j$. These are well-known from propositional modal logic since they correspond to the system $\mathsf{S4.2}$.[17]

But directed partial orders are also used to index directed families of objects. In particular, given a category, recall that a *directed family* in a category is a family of objects B_i indexed by elements i of I and maps $h_{ij} : B_i \to B_j$ for $i \leq j$ such that $h_{ii} : B_i \to B_i$ is the identity map and if $i \leq j \leq k$ then $h_{ik} = h_{jk} \circ h_{ij}$. Finally, a *direct limit* is an object D such that there are maps $h_i : B_i \to D$ which commute with the h_{ij}-maps for $i \leq j$ by $h_j \circ h_{ij} = h_i$, and such that for any other objects D' and maps $h'_i : B_i \to D'$ with this property there is a unique map

[17] See Hughes and Cresswell (1996, p. 134).

$h' : D \to D'$ such that $h' \circ h_i = h'_i$.[18] One typically records this information in the diagram shown in Figure 3.2.

$$\begin{tikzcd}
 & D' & \\
 & D \arrow[u, dotted, "\exists\,!\,h'"'] & \\
B_i \arrow[ur, "h_i"] \arrow[uur, bend left, "h'_i"] \arrow[rr, "h_{ij}", "i \leq j"'] & & B_j \arrow[ul, "h_j"'] \arrow[uul, bend right, "h'_j"']
\end{tikzcd}$$

Figure 3.2: The direct limit of directed family B_i

In the category of first-order structures in a given first-order signature L with homomorphisms in that signature, direct limits of directed systems exist.[19] The underlying set of the direct limit is simply the equivalence classes $[x]_E$ of elements of the first-order domains (which we assume to be disjoint) under the equivalence relation of being "eventually identical," that is xEy for x from B_i and y from B_j if and only if there is $k \geq i, j$ such that $h_{ik}(x) = h_{jk}(y)$. The maps $h_i : B_i \to D$ are then given by $h_i(a) = [a]_E$. Further, the interpretation of the signature is given as follows on the direct limit

- For an L-constant symbol c, since the maps are L-homomorphisms, one has that c^{B_i} is eventually identical to c^{B_j}, for any indexes i, j, and hence c^D is just their equivalence class.

- For an n-place relation symbol P, one simply declares P^D to be the n-tuples of equivalence classes of elements $x_1 \in B_{i_1}, \ldots, x_n \in B_{i_n}$ such that there is $j \geq i_1, \ldots, i_n$ with $(h_{i_1 j}(x_1), \ldots, h_{i_n, j}(x_n)) \in P^{B_j}$.

[18] The direct limit occurs throughout contemporary mathematics – see e.g. Lang (2002, pp. 159 ff) for algebra and Sakai (2013, pp. 55 ff) for topology.

[19] See Hodges (1993, p. 50) and Chang and Keisler (1990, p. 322).

- For an n-place function symbol f, on input of an n-tuple of equivalence classes of elements $x_1 \in B_{i_1}, \ldots, x_n \in B_{i_n}$ with $j \geq i_1, \ldots, i_n$, it outputs the equivalence class of $f^{B_j}(h_{i_1 j}(x_1), \ldots, h_{i_n,j}(x_n))$, which implies the identity:

$$f^D([x_1]_E, \ldots, [x_n]_E) = [f^{B_j}(h_{i_1 j}(x_1), \ldots, h_{i_n j}(x_n))]_E \tag{3.24}$$

These are well-defined using the properties of the directed family. A classical part of the construction is that if h_{ij} are L-embeddings, then so is h_i.[20]

3.6.2 Associating frames and models to direct limits

Given such a direct limit of a directed system, we can form the base frame $\langle I, \leq, B, D, S\rangle$ simply by setting $S_{i,e} : B_{i,e} \to D_e$ equal to $h_i : B_i \to D$, and by setting $S_{i,s} : B_{i,s} \to D_s$ equal to the identity map on I.[21] Note that the maps $h_{ij} : B_i \to B_j$ for $i \leq j$ are technically not part of this base frame, despite their being a constitutive part of the directed family. By Proposition 3.1, let $\langle I, \leq, B, D, S, T\rangle$ be a full minimal frame extending the base frame.

Let us now define a signature $\widetilde{L}$ in the sense of §3.3.3 from the first-order signature L. We identify an L-constant symbol c with an $\widetilde{L}$-constant symbol $\widetilde{c}$ of type e, and we identify a binary L-function symbol f with an $\widetilde{L}$-constant symbol $\widetilde{f}$ of type $\langle e, \langle e, e\rangle\rangle$ (by Currying), and similarly for n-ary function symbols for larger arities. Finally, using characteristic functions, we identify an L-binary relation symbol R with an $\widetilde{L}$-constant symbol $\widetilde{R}$ of type $\langle e, \langle e, t\rangle\rangle$, and similarly for n-ary relation symbols for larger arities.

In order to define a model L, we need to take a preliminary step and to introduce a further constraint. In particular, in the case where L has no relations, we argue that for any $\widetilde{L}$-

[20]See Hodges (1993, p. 51).

[21]It might also be natural to consider setting $B_{i,s} = R[i]$ and let $S_{i,s} : B_{i,s} \to I$ be the restriction of the identity map.

constant symbol $\widetilde{C}$ of type a, one has that the naturally defined element $\widetilde{C}^{B_i}$ is an element of $B_{i,a}$. Further, we can extend this to the case where L has relations if we further assume that the h_{ij} are L-embeddings. Formally, for any $\widetilde{L}$-constant symbol $\widetilde{C}$ of type a and any index i, we are in the next paragraphs defining $\widetilde{C}^{B_i}$ and showing that it is an element of $B_{i,a}$. After we have done this, the idea will be to define a model by pushing forward under the transfer maps.

First consider the case of an L-constant symbol c. We set $\widetilde{c}^{B_i} = c^{B_i}$, which is obviously an element of $B_i = B_{i,e}$.

Second consider the case of an L-unary function symbol f. One has that $\widetilde{f}$ has type $\langle e,e\rangle$, and we set $\widetilde{f}^{B_i}(u) = f^{B_i}(u)$. We must check that $\widetilde{f}^{B_i}$ is in $B_{i,\langle e,e\rangle}$. Since we are working with a full frame, it then suffices to check that (*frame:1*) holds. Hence, suppose that $T_{i,e}(u) = T_{i,e}(v)$, where u, v are from $B_{i,e} = B_i$. Then $S_{i,e}(u) = S_{i,e}(v)$. Then $[u]_E = [v]_E$. Then there is $j \geq i$ such that $h_{ij}(u) = h_{ij}(v)$. Then $f^{B_j}(h_{ij}(u)) = f^{B_j}(h_{ij}(v))$. Then since h_{ij} is an L-homomorphism, we have that $h_{ij}(f^{B_i}(u)) = f^{B_j}(h_{ij}(u)) = f^{B_j}(h_{ij}(v)) = h_{ij}(f^{B_i}(v))$ and hence $[f^{B_i}(u)]_E = [f^{B_i}(v)]_E$. Then $S_{i,e}(f^{B_i}(u)) = S_{i,e}(f^{B_i}(v))$. Then $T_{i,e}(\widetilde{f}^{B_i}(u)) = T_{i,e}(\widetilde{f}^{B_i}(v))$, which is what we wanted to show.

Let us now consider the slightly more complicated case of an L-binary function symbol g. Then $\widetilde{g}$ has type $a = \langle e, \langle e,e\rangle\rangle$, and we set $(\widetilde{g}^{B_i}(u))(y) = g^{B_i}(u,y)$. We must show that $\widetilde{g}^{B_i}$ is an element of $B_{i,a}$, that is, that it satisfies (*frame:1*). So suppose that u, v are from $B_{i,e} = B_i$ with $T_{i,e}(u) = T_{i,e}(v)$; we must argue that $T_{i,\langle e,e\rangle}(\widetilde{g}^{B_i}(u)) = T_{i,\langle e,e\rangle}(\widetilde{g}^{B_i}(v))$. Hence let x from some B_k be given. We must show that $T_{i,\langle e,e\rangle}(\widetilde{g}^{B_i}(u))[x]_E = T_{i,\langle e,e\rangle}(\widetilde{g}^{B_i}(v))[x]_E$. By minimality, it suffices to consider the case where xEy and y is from B_i. Then what we must show is that $T_{i,\langle e,e\rangle}(\widetilde{g}^{B_i}(u))[y]_E = T_{i,\langle e,e\rangle}(\widetilde{g}^{B_i}(v))[y]_E$. This is the same as $T_{i,\langle e,e\rangle}(\widetilde{g}^{B_i}(u))(T_{i,e}(y)) = T_{i,\langle e,e\rangle}(\widetilde{g}^{B_i}(v))(T_{i,e}(y))$. By appeal to (*frame:2*),[22] this is the same as $T_{i,e}(\widetilde{g}^{B_i}(u)(y)) =$

[22]Note that the appeal to (*frame:2*) in showing (*frame:1*) is permissible, since we are really appealing to the hypothesis that the frame satisfies (*frame:2*) to show that a particular function, namely $\widetilde{g}^{B_i}$, satisfies the defining condition (*frame:1*) of being a member of $B_{i,a}$.

$T_{i,e}(\widetilde{g}^{B_i}(v)(y))$. This in turn is equivalent to $S_{i,e}(\widetilde{g}^{B_i}(u)(y)) = S_{i,e}(\widetilde{g}^{B_i}(v)(y))$ and $[g^{B_i}(u,y)]_E = [g^{B_i}(v,y)]_E$. As in the last paragraph, the hypothesis that $T_{i,e}(u) = T_{i,e}(v)$ gives us $j \geq i$ such that $h_{ij}(u) = h_{ij}(v)$. Then as in the last paragraph, we have $h_{ij}(g^{B_i}(u,y)) = h_{ij}(g^{B_i}(v,y))$, which implies $[g^{B_i}(u,y)]_E = [g^{B_i}(v,y)]_E$, which is what we needed to show. The argument for higher arities is similar.

Turning to the relation symbols, we again extend the hypothesis so that h_{ij} are L-embeddings. But then the $T_{i,e}$ maps are injective implies that conditions (*frame:1*) is trivially satisfied and so $B_{i,\langle a,b\rangle}$ just consists of all functions from $B_{i,a}$ to $B_{i,b}$. Hence if P is an L-unary relation symbol, then $\widetilde{P}$ is an $\widetilde{L}$-constant of type $\langle e,t\rangle$ and we set $\widetilde{P}^{B_i}(u) = 1$ iff $P^{B_i}(u)$, and this is then trivially an element of $B_{i,\langle e,t\rangle}$. And similarly for higher arities.

All this in place, we can now define an $\widetilde{L}$-model. We do this by pushing forward the interpretation of each of the symbols to form an $\widetilde{L}$-model $\mathcal{M}$ with frame $\langle I, \leq, B, D, S, T\rangle$. In particular, for any $\widetilde{L}$-constant symbol $\widetilde{C}$ of type a, we define $\widetilde{C}^{\mathcal{M}}(i) = T_{i,a}(\widetilde{c}^{B_i})$, which we can do since we have just shown that $\widetilde{C}^{B_i}$ is an element of $B_{i,a}$.

First let us show:

Proposition 3.12. *Suppose that u is a variable of type e. Then $\Diamond \mathsf{E}_e(u)$ is valid on the model $\mathcal{M}$.*

Proof. For the sake of definiteness, let us specify that $\mathsf{E}_e(u) \equiv \exists\, v\; u = v$, where v is a variable of type e distinct from u. Then let i, σ be indexes and let g be a variable assignment with $g_e(u) = [x]_E$ where x is from B_j. Then we must show that $[\![\Diamond \exists\, v\; v = u]\!]_{\mathcal{M},i,\sigma,g} = 1$. By directedness, choose $k \geq i, j$ and let $y = h_{jk}(x)$, so that $[x]_E = [y]_E$ (taking care to note that we are using u, v for variables of type e and equivalence classes of x, y for elements of the underlying domain $D = D_e$). Let h be the v-variant of g such that $h_e(v) = [y]_E$. Then we have

$$[\![u]\!]_{\mathcal{M},k,\sigma,h} = h_e(v) = [y]_E = [x]_E = g_e(u) = h_e(u) = [\![v]\!]_{\mathcal{M},i,\sigma,g}$$

Then $[\![v = u]\!]_{\mathcal{M},k,\sigma,h} = 1$, and since $[y]_E \in h_k(B_k) = T_{k,e}(B_{k,e}$ since $y = h_{jk}(x) \in B_k$, we have $[\![\exists\, v\; v = u]\!]_{\mathcal{M},k,\sigma,g} = 1$ and hence $[\![\Diamond\exists\, v\; v = u]\!]_{\mathcal{M},i,\sigma,g} = 1$. $\square$

We can use this to show that all of the first-order structure on D is expressible in the modal structure on $\mathcal{M}$:

Theorem 3.13. *For every first-order formula φ in first-order signature L with n distinct free first-order variables there is a well-formed expression $\widetilde{\varphi}$ in the signature $\widetilde{L}$ with n-distinct free variables $u_1, \ldots, u_n$ of type e such that*

$$D \vDash \varphi([x_1]_E, \ldots, [x_n]_E) \textit{ iff } [\![\widetilde{\varphi}]\!]_{\mathcal{M},i,\sigma,g} = 1$$

for all elements $[x_1]_E, \ldots, [x_n]_E$ of D and all indexes i, σ and all variable assignments g such that $g_e(u_1) = [x_1]_E, \ldots, g_e(u_n) = [x_n]_E$. Further, $\widetilde{\varphi}$ can be chosen so that all quantifiers are of type e.

(As with the previous discussion, if there are relations in the signature L, we further suppose that the maps h_{ij} are L-embeddings).

Proof. By the previous proposition and Proposition 3.2, it suffices to show the result for the atomics. And indeed by familiar considerations it suffices to consider unnested atomics φ.[23]

[23]See Hodges (1993, 58 ff.).

First consider the L-constant c. Then, $c^{\mathcal{M}}(i) = T_{i,e}(\widetilde{c}^{B_i}) = [c^{B_i}]_E = c^D$. Hence if variable assignment g assigns variable u_1 to the element $[x_1]_E$ of $D_e = D$, then we have

$$c^D = [x_1]_E \text{ iff } [\![c = u_1]\!]_{\mathcal{M},i,\sigma,g} = 1$$

Second, consider the L-function symbol f, which for the sake of simplicity we assume to be binary, so that the $\widetilde{L}$-constant symbol $\widetilde{f}$ has type $a = \langle e, \langle e, e\rangle\rangle$. By the previous argument, $\widetilde{f}^{B_i}$ is an element of $B_{i,a}$ and hence $\widetilde{f}^{\mathcal{M}}(i) = T_{i,a}(\widetilde{f}^{B_i})$ is an element of $D_{i,a}$. Hence, suppose that $x_1 \in B_{i_1}$ and $x_2 \in B_{i_2}$ and $y \in B_j$, so that $[x_1]_E$ and $[x_2]_E$ and $[y]_E$ are elements of $D_e = D$. Suppose that g is a variable assignment with $g(u_1) = [x_1]_E$ and $g(u_2) = [x_2]_E$ and $g(v) = [y]_E$ (and take care to note that the subscripted u, v are the variables while equivalences classes of the subscripted x, y are elements of the model). Then we claim:

$$f^D([x_1]_E, [x_2]_E) = [y]_E \text{ iff } [\![\Diamond(\mathsf{E}(u_1) \wedge \mathsf{E}(u_2) \wedge \widetilde{f}(u_1)(u_2) = v)]\!]_{\mathcal{M},i,\sigma,g} = 1 \tag{3.25}$$

First suppose that $f^D([x_1]_E, [x_2]_E) = [y]_E$. It suffices then to show that $[\![(\mathsf{E}(u_1) \wedge \mathsf{E}(u_2) \wedge f(u_1)(u_2) = v)]\!]_{\mathcal{M},k,\sigma,g} = 1$, where $k \geq i_1, i_2, j, i$. For the first two conjuncts, note that

$$g(u_1) = [x_1]_E = [h_{i_1,k}(x_1)]_E = T_{k,e}(h_{i_1k}(x_1)) \in T_{k,e}(B_{k,e})$$
$$g(u_2) = [x_2]_E = [h_{i_2,k}(x_2)]_E = T_{k,e}(h_{i_2k}(x_2)) \in T_{k,e}(B_{k,e})$$

For the third, we use the identities in the two previous lines and we apply (*frame:2*) twice:

$$\begin{aligned}
[\![\widetilde{f}(u_1)(u_2)]\!]_{\mathcal{M},k,\sigma,g} &= \Big((T_{k,\langle e,\langle e,e\rangle\rangle}(\widetilde{f}^{B_k}))(T_{k,e}(h_{i_1k}(x_1)))\Big)(T_{k,e}(h_{i_2k}(x_2))) \\
&= \Big(T_{k,\langle e,e\rangle}(\widetilde{f}^{B_k}(h_{i_1,k}(x_1)))(T_{k,e}(h_{i_2k}(x_2))\Big) \\
&= T_{k,e}\Big((\widetilde{f}^{B_k}(h_{i_1,k}(x_1)))(h_{i_2k}(x_2))\Big) \\
&= T_{k,e}(f^{B_k}(h_{i_1,k}(x_1),h_{i_2,k}(x_2))) = [f^{B_k}(h_{i_1,k}(x_1),h_{i_2,k}(x_2))]_E \\
&= f^D([x_1]_E,[x_2]_E) = [y]_E = [\![v]\!]_{\mathcal{M},k,\sigma,g}
\end{aligned}$$

Conversely, suppose that $[\![\Diamond(\mathsf{E}(u_1)\wedge\mathsf{E}(u_2)\wedge\widetilde{f}(u_1)(u_2)=v)]\!]_{\mathcal{M},i,\sigma,g}=1$. Choose $k\geq i$ such that $[\![(\mathsf{E}(u_1)\wedge\mathsf{E}(u_2)\wedge f(u_1)(u_2)=v)]\!]_{\mathcal{M},k,\sigma,g}=1$. Then by the first two conjuncts one has

$$\begin{aligned}
&[x_1]_E = g(u_1)\in T_{k,e}(B_{k,e}) = h_k(B_k) \\
&[x_2]_E = g(u_2)\in T_{k,e}(B_{k,e}) = h_k(B_k)
\end{aligned}$$

Hence choose $z_1,z_2\in B_k$ such that $[x_1]_E=[z_1]_E$ and $[x_2]_E=[z_2]_E$. Then applying (*frame:2*) twice again, we obtain:

$$\begin{aligned}
&[y]_E = g(v) = [\![\widetilde{f}(u_1)(u_2)]\!]_{\mathcal{M},k,\sigma,g} = \Big((T_{k,\langle e,\langle e,e\rangle\rangle}(\widetilde{f}^{B_k}))(g(u_1))\Big)(g(u_2)) \\
&= \Big((T_{k,\langle e,\langle e,e\rangle\rangle}\widetilde{f}^{B_k})(T_{k,e}(z_1))\Big)(T_{k,e}(z_2)) \\
&= \Big(T_{k,\langle e,e\rangle}(\widetilde{f}^{B_k}(z_1))\Big)(T_{k,e}(z_2)) \\
&= T_{k,e}((\widetilde{f}^{B_k}(z_1))(z_2)) = [f^{B_k}(z_1,z_2)]_E = f^D([z_1]_E,[z_2]_E) = f^D([x_1]_E,[x_2]_E)
\end{aligned}$$

Finally, under the supposition that h_{ij} are L-embeddings, a similar argument shows that for a binary relation P, we have

$$([x_1]_E,[x_2]_E)\in P^D \text{ iff } [\![\Diamond(\mathsf{E}(u_1)\wedge\mathsf{E}(u_2)\wedge\widetilde{P}(u_1)(u_2)=1)]\!]_{\mathcal{M},i,\sigma,g}=1$$

and similarly for higher arities. $\square$

3.6.3 Options for the degree type

In the above, we have not explicitly tended to type d, and per the remarks in §3.2.2, in the above construction we can take $S_{i,d} = S_{i,t}$ and $B_{i,d} = D_d = \{0,1\}$, with the natural ordering; that is, we just treat the domains associated to type d as a redundant copy of the domains associated to type t. However, there are two natural variations in which one might want something different. First, consider the case where the directed family is a directed family of expansions of linear orders, with distinguished binary relation $\leq$, where the maps are L-embeddings. Then the direct limit is itself a linear order. Hence, in this case we can rather just treat treat the domains associated to type d as a redundant copy of the domains associated to type e, and set $S_{i,d} = S_{i,e}$ and $B_{i,d} = B_{i,e}$ and $D_d = D_e$.

Second, consider the case where the direct family again has $\leq$ in its signature, but that this satisfies only the axioms of reflexivity, transitivity, and linearity, and not anti-symmetry. Then consider the equivalence relation $\cong_i$ on $B_{i,e}$ given by $x \cong_i y$ iff $B_i \vDash ((x \leq y) \wedge (y \leq x))$. Then if we set $B_{i,d} = {}^{B_{i,e}}/_{E_i}$ we have that $B_{i,d}$ has a natural linear order induced by $\leq$, namely $[x]_{\cong_i} \leq [y]_{\cong_i}$ iff $B_i \vDash x \leq y$. The equivalence relation $\cong$ defined on the direct limit D similarly induces a linear order on $D_d = {}^{D_e}/_{E}$, and there is a map $S_{i,d} : B_{i,d} \to D_d$ given by sending $S_{i,d}([x]_{\cong_i}) = [h_i(x)]_{\cong}$. It is easy to check that this is injective (since we are assuming the maps h_{ij} are L-embeddings) and a homomorphism of linear orders and so an embedding of linear orders. Likewise, the maps $H_{ij} : B_{i,d} \to B_{j,d}$ given by sending $H_{ij}([x]_{\cong_i}) = [h_{ij}(x)]_{\cong_j}$ is an embedding of linear orders for the same reason, and D_d is the direct limit of this directed family $B_{i,d}$ of linear orders.

The above examples illustrate how to commensurate comparative structures in the semantics. Fundamentally, this serves to demonstrate that degrees can be integrated into the variable domain framework without issue. But the second example also shows that we can emulate some amount of 'scale synthesis' in our models. The suggestion that the scales upon which measure functions and degrees are situated can be generated from primitive relations or quasiorders is standard in linguistics.[24] However, because of its psychologistic nature, this construction is typically handled antecedently to the linguistic model, and the resulting scale is taken as given. Our second example situates the complete process into the model with the desired scale structure arising from the semantics. This scale rendering is similar to that of Bale (2008), and if one were to proliferate the degree type in our setting over a number of different primary scales, this would evoke the universal scale Bale postulates.

Additionally, with a degree type in our variable domain setting, we can transparently model some novel semantic possibilities. For instance, the standard treatment for the positive form of a gradable adjective, as in "Mary is tall", is to view it as a special case of the comparative construction where Mary's degree of tallness is compared with a contextually-supplied standard degree of tallness.[25] We can model this by including a constant symbol of type d whose varying value across worlds supplies the contextual value of the standard of comparison. Now, if we examine a modal extension of the previous proposition, "Mary could have been tall", we have some options for our translation and model. One could simply apply a diamond operator to the previous analysis, in effect quantifying over the different contexts and so Mary achieves possible-tallness by there being a context of low standards. We could also permit Mary's height to vary across worlds, and so Mary achieves possible-tallness by her height possibly being exceptional. The former option seems comparatively coarse in illustrating how Mary is possibly-tall. We can vivify this with our semantics if we take the standard degree to be provided by some function involving, say, a generalized quantifier over (the degrees of tallness of) individuals with the property in question. The variable domain semantics then give an example where Mary achieves possible-tallness by there being contexts

[24]See Cresswell (1976, p. 281) for an early statement on this. Luce et al. (2007) provides a thorough and general view of such constructions across disciplines.

[25]See, e.g., von Stechow (1984, 59ff.) and Kennedy and McNally (2005, p. 348–351).

in which there are few exceedingly taller individuals, with neither anyone's height varying nor the method of determining tallness changing.

3.6.4 Barcan and related principles

As noted in Proposition 3.12, the principle $\diamond \mathsf{E}_e(v)$ is valid in direct limit models. As for Barcan, it fails on type e for the usual reasons related to anti-monotonicity.[26] To see a failure at higher-type, consider type $\langle e,e\rangle$ and consider $I = \{1,2\}$ with the natural ordering in which $B_1 = \{1\}$ has one element and $B_2 = \{1,2\}$ has two elements and where $h_{12}(1) = 1$. Since the direct limit is isomorphic to B_2 itself, we can set $D_e = B_2$ with the embedding $h_1 = h_{12}$. By Proposition 3.1, let $\langle I, \leq, B, D, S, T\rangle$ be a full injective minimal frame. We further require that it satisfies condition (3.2) from the proof of that proposition. This condition entails that $T_{1,\langle e,e\rangle} : B_{1,\langle e,e\rangle} \to D_{\langle e,e\rangle}$ is the function sending the identity function on B_1 to the constant function sending both $1,2$ to 1. Further, the frame conditions guarantee that $B_{2,\langle e,e\rangle} \to D_{\langle e,e\rangle}$ and $T_{2,\langle e,e\rangle} : B_{2,\langle e,e\rangle} \to D_{\langle e,e\rangle}$ is the constant function. Obviously every function from B_1 to itself is the identity, but the same is not so on B_2, and this observation allows one to establish the following, where $x : e$ and $y : \langle e,e\rangle$ are variables:

$$[\![\diamond(\exists\, y\ \forall\, x\ y(x) \neq x) \wedge \neg(\exists\, y\ \diamond\ \forall\, x\ y(x) \neq x)]\!]_{\mathcal{M},1,\varnothing,g} = 1$$

To see this, let h be the y-variant of g such that $h(y)$ is the function which permutes 1 and 2. Let k be an x-variant of h such that $k(x) \in T_{2,e}(B_{2,e}) = \{1,2\}$. If $k(x) = 1$ then $[\![y(x)]\!]_{\mathcal{M},2,\varnothing,k} = 2$ and if $k(x) = 2$ then $[\![y(x)]\!]_{\mathcal{M},2,\varnothing,k} = 1$, and hence in either case $[\![y(x) \neq x]\!]_{\mathcal{M},2,\varnothing,k} = 1$. Generalizing over all such x-variants, we have $[\![\forall\, x\ y(x) \neq x]\!]_{\mathcal{M},2,\varnothing,h} = 1$. Since h is a y-variant of g with $h(y) \in T_{2,\langle e,e\rangle}(B_{2,\langle e,e\rangle})$ we have $[\![\exists\, y\ \forall\, x\ y(x) \neq x]\!]_{\mathcal{M},2,\varnothing,g} = 1$. Since $2 \geq 1$ we have $[\![\diamond(\exists\, y\ \forall\, x\ y(x) \neq x)]\!]_{\mathcal{M},1,\varnothing,g} = 1$. For the second conjunct, suppose

[26] Fitting and Mendelsohn (1998, p. 182)

for reductio that $[\![\exists\ y\ \ \Diamond\ \ \forall\ x\ y(x) \neq x)]\!]_{\mathcal{M},1,\varnothing,g} = 1$. Let h be a variable assignment such that $h(y) \in T_{1,\langle e,e\rangle}(B_{1,\langle e,e\rangle})$ and there is $j \in \{1,2\}$ with $[\![\forall\ x\ y(x) \neq x)]\!]_{\mathcal{M},j,\varnothing,h} = 1$. Then $h(y)$ is the the constant function sending both $1,2$ to 1. If k is the x-variant of h with $k(x) = 1 \in T_{j,e}(B_{j,e})$ then $[\![y(x)]\!]_{\mathcal{M},j,\varnothing,g} = 1 = [\![x]\!]_{\mathcal{M},j,\varnothing,k}$ a contradiction.

Converse Barcan holds on type e, either in the form of the axiom $\mathsf{E}_e(v) \to \Box\mathsf{E}_e(v)$ or the schema $(\exists\ v\ \ \Diamond\ \varphi(v)) \to (\Diamond\exists\ v\ \ \varphi(v))$, where v is a variable of type e.[27] This is just because $i \leq j$ implies $T_{i,e}(B_{i,e}) \subseteq T_{j,e}(B_{j,e})$, and then the usual monotonicity consideration surrounding Converse Barcan.[28] We have the same result at all types in certain frames:

Proposition 3.14. *Suppose that $\langle I, \leq, B, D, S, T\rangle$ is a full injective bounded frame. Then for all extended types a and indexes $i \leq j$ we have $T_{i,a}(B_{i,a}) \subseteq T_{j,a}(B_{j,a})$.*

Proof. It suffices, by induction on type, to construct for each pair of indexes $i \leq j$ an injection $H_{ij,a} : B_{i,a} \to B_{j,a}$ such that $T_{j,a} \circ H_{ij,a} = T_{i,a}$. Recall that since the frame is injective, the conditions (*frame:1*)-(*frame:2*) are satisfied trivially and hence the sets $B_{i,a}$ are determined by the recursion:

$$B_{i,a} = B_i, \quad B_{i,s} = I, \quad B_{i,t} = \{0,1\}, \quad B_{i,\langle a,b\rangle} = B_{i,b}^{B_{i,a}}, \quad B_{i,\langle s,b\rangle} = D_b^{B_{i,s}}$$

For the basic extended types, there are three case to consider. For $a = e$, we simply set $H_{i,a} = h_i$. For the basic extended types $a = t$ and $a = s$, it follows trivially since the maps are all the identity in these cases. For the induction step, the type $\langle s,b\rangle$ is trivial since $B_{i,s} = B_{j,s} = W$ and hence $B_{i,\langle s,b\rangle} = D_b^{B_{i,s}} = D_b^{B_{j,s}} = B_{j,\langle s,b\rangle}$. That is, one can just take $H_{ij,\langle s,b\rangle}$ to be the identity map. And the inclusion $T_{i,\langle s,b\rangle}(B_{i,\langle s,b\rangle}) \subseteq T_{j,\langle s,b\rangle}(B_{j,\langle s,b\rangle})$ follows trivially since in this case these two sets are identical since $T_{i,\langle s,b\rangle}$ and $T_{i,\langle s,b\rangle}$ are the identity maps.

[27] Fitting and Mendelsohn (1998, pp. 181-182).
[28] Fitting and Mendelsohn (1998, p. 181).

However the induction step associated to $\langle a, b\rangle$ is complex. Hence, suppose the result holds for types a, b, and we show it holds for $\langle a, b\rangle$. For $i \leq j$ and f in $B_{i,\langle a,b\rangle}$ we define $H_{ij,\langle a,b\rangle}(f) = g$ in $B_{j,\langle a,b\rangle}$ as follows:

$$\begin{tikzcd}
& B_{j,a} \arrow[dr, "T_{j,a}"] \arrow[dd, "g"] & \\
B_{i,a} \arrow[ur, "H_{ij,a}"] \arrow[rr, "T_{i,a}"] \arrow[dd, "f"] & & D_a \arrow[dd, "T_{i,\langle a,b\rangle}(f)=T_{j,\langle a,b\rangle}(g)"] \\
& B_{j,b} \arrow[dr, "T_{j,b}"] & \\
B_{i,b} \arrow[ur, "H_{ij,b}"] \arrow[rr, "T_{i,b}"] & & D_b
\end{tikzcd}$$

$$g(y) = \begin{cases} (H_{ij,b} \circ f \circ H_{ij,a}^{-1})(y) & \text{if } T_{j,a}(y) \in T_{i,a}(B_{i,a}), \\ T_{j,b}^{-1}((T_{i,\langle a,b\rangle}(f))(T_{j,a}(y))) & \text{otherwise} \end{cases}$$

This is well-defined on the first clause since if $T_{j,a}(y) = T_{i,a}(x)$ for some x in $B_{i,a}$ then by induction hypothesis $T_{j,a}(y) = T_{i,a}(x) = T_{j,a}(H_{ij,a}(x))$ and since T_j is an injection we have $y = H_{ij,a}(x)$. It is well-defined on the second since the boundedness constraint gives that for all values of u, one has that $(T_{i,\langle a,b\rangle}(f))(u)$ is an element of $T_{i,b}(f(B_{i,a}))$, which is trivially a subset of $T_{i,b}(B_{i,b})$, and which by injunction hypothesis is a subset of $T_{j,b}(B_{j,b})$.

First we show that $T_{i,b}(f(B_{i,a})) = T_{j,b}(g(B_{j,a}))$. For the left to right inclusion, suppose that z is in $T_{i,b}(f(B_{i,a}))$, with $z = T_{i,b}(f(x))$ for some x in $B_{i,a}$. Let $y = H_{ij,a}(x)$ in $B_{j,a}$. Then $T_{j,a}(y) = T_{i,a}(H_{ij,a}(x)) = T_{i,a}(x)$, so that $T_{j,a}(y)$ is also in $T_{i,a}(B_{i,a})$. Then by the first clause in the definition of g, we have that $T_{j,b}(g(y)) = T_{j,b}(H_{ij,b}(f(H_{ij,a}^{-1}(y)))) = T_{j,b}(H_{ij,b}(f(x))) = T_{i,b}(f(x)) = z$. For the right to left inclusion, suppose that z is in $T_{j,b}(g(B_{j,a}))$, with $z = T_{j,b}(g(y))$ for some y in $B_{j,a}$. There are two cases. First suppose that $T_{j,a}(y)$ is in $T_{i,a}(B_{i,a})$. Then by the first clause of the definition of g, we have $z = T_{j,b}(g(y)) = (T_{j,b} \circ H_{ij,b} \circ f \circ H_{ij,a}^{-1})(y) = T_{i,b}(f(H_{ij,a}^{-1}(y)))$, which is an element of $T_{i,b}(f(B_{i,a}))$. Second suppose that

$T_{j,a}(y)$ is not in $T_{i,a}(B_{i,a})$. Then by the second clause of the definition of g, we have $z = T_{j,b}(g(y)) = ((T_{i,\langle a,b\rangle}(f))(T_{j,a}(y)))$, which by the boundedness condition is an element of $T_{i,b}(f(B_{i,a}))$.

We show now that for all z in D_a, we have $(T_{i,\langle a,b\rangle}(f))(z) = (T_{j,\langle a,b\rangle}(g))(z)$. By the previous paragraph and strong minimality it suffices to show this for z in the union $T_{i,a}(B_{i,a}) \cup T_{j,a}(B_{j,a})$, where recall that the former set is a subset of the latter set by induction hypothesis. First suppose that z is in $T_{i,a}(B_{i,a})$. Choose x in $B_{i,a}$ such that $T_{i,a}(x) = z$. Let $y = H_{ij,a}(x)$ so that y is in $B_{j,a}$, and $T_{j,a}(y) = T_{j,a}(H_{ij,a}(x)) = T_{i,a}(x) = z$, and so $T_{j,a}(y)$ is an element of $T_{i,a}(B_{i,a})$. Then by two applications (*frame:2*) with an application of the first clause in definition of g in the middle, we have $(T_{j,\langle a,b\rangle}(g))(z) = T_{j,b}(g(y)) = (T_{j,b} \circ H_{ij,b} \circ f \circ H_{ij,a}^{-1})(y) = T_{i,b}(f(x)) = (T_{i,\langle a,b\rangle}(f))(z)$. Second suppose that z is in $T_{j,a}(B_{j,a}) \setminus T_{i,a}(B_{i,a})$. Choose y in $B_{j,a}$ such that $T_{j,a}(y) = z$. Then by two applications of (*frame:2*) with an application of second clause in the definition of g in the middle, we have $(T_{j,\langle a,b\rangle}(g))(z) = T_{j,b}(g(y)) = (T_{i,\langle a,b\rangle}(f))(T_{j,a}(y)) = (T_{i,\langle a,b\rangle}(f))(z)$.

Finally, we show that $H_{ij,\langle a,b\rangle}$ is injective. Suppose that $H_{ij,\langle a,b\rangle}(f) = H_{ij,\langle a,b\rangle}(f')$. We must show that $f = f'$. Suppose that x is in $B_{i,a}$; we must show that $f(x) = f'(x)$. Let $y = H_{ij,a}(x)$ in $B_{j,a}$. Then since $T_{j,a}(y) = T_{j,a}(H_{ij,a}(x)) = T_{i,a}(x)$ we have that $T_{j,a}(y)$ is in $T_{i,a}(B_{i,a})$. Then we have

$$\begin{aligned} H_{ij,b}(f(x)) &= H_{ij,b}(f(H_{ij,a}^{-1}(y))) = (H_{ij,\langle a,b\rangle}(f))(y) \\ &= (H_{ij,\langle a,b\rangle}(f'))(y) = H_{ij,b}(f'(H_{ij,a}^{-1}(y))) = H_{ij,b}(f'(x)) \end{aligned}$$

Then since $H_{ij,b}$ is an injection we are done. □

3.6.5 Direct limits and situations

Let L be a first-order signature, and let B be a first-order L-structure. Let J be an index set which enumerates all of the L-substructures of B, where B_j is the j-th such L-substructure. We put a partial order on J by declaring that $j \leq k$ iff B_j is a substructure of B_k. Since the original model is a substructure of itself, this is directed.

Let I be any directed subset of J. (For instance, if L is relational and B is infinite, then any subset of B is an L-substructure by restriction, and one could consider I to be the set of L-substructures whose domain is finite – this would be directed simply because the union of two finite sets is finite). If $j \leq k$, then let $h_{jk} : B_j \to B_k$ be the identity map, and let D be the L-substructure generated by the B_i as i ranges over I. By directedness, this is just the union of the B_i with i in I, in the case where L is relational. Let $h_i : B_i \to D$ be given by the identity map. Then by the previous paragraphs we have that D is the direct limit of the direct family B_i as i varies across I. (For instance, if B was the real numbers with just less-than, and I was the set of finite L-substructures all of whose members were rationals, then D would be the rational numbers). The construction earlier in this section then shows how to associate to this a base frame $\langle I, \leq \restriction I, B, D, S \rangle$, and then we can extend to a frame, and then to a model by pushing forward.

This construction can be extended as follows. Let L be a first-order signature, and let Γ be an index set which indexes a collection of L-structures B^γ for γ in Γ, where they all have the same first-order domain D. For each γ in Γ, let J^γ be an index set which enumerates all of the L-substructures of B^γ, and where B^γ_j is the j-th element in the list, when j comes from J^γ. We assume without loss of generality that these index sets are disjoint as γ varies, and we put the partial order $\leq^\gamma$ on J^γ as before. Then we have direct limits D^γ as γ varies but we can take them to have the same underlying domain D as the original structures, since we enumerated all of the L-substructures. And then we can use the union construction from §3.2.4 to consider the base frame formed from their union, which will have as its set of indexes $I = \bigsqcup_{\gamma\in\Gamma} J^\gamma$. Then we can extend and push forward to obtain a model $\mathcal{M}$, as before. We will no longer have (3.25) per se, but rather the following

$$i \in J^\gamma \text{ implies: } f^{D^\gamma}(x_1, x_2) = y \text{ iff } [\![\Diamond(\mathsf{E}(u_1) \wedge \mathsf{E}(u_2) \wedge \widetilde{f}(u_1)(u_2) = v)]\!]_{\mathcal{M},i,\sigma,g} = 1$$

In this, again g assigns u_1 to x_1, u_2 to x_2 and v to y. (But note that the E subscript has dropped out since in these direct limits we can just take this equivalence relation to be identity).

Let us note what the minimal and maximal elements of the unioned indexed sets under the unioned ordering are. The maximal elements are the indexes corresponding to the original models B^γ, and these are incomparable. The minimal elements are simply the minimal substructures, but in the case where the language L is relational, the minimal elements will be the singletons. It is natural to consider a couple of other modalities in this setting. First, we use $\Box^*$ and $\Diamond^*$ for the downward looking variants of $\Box$ and $\Diamond$. For instance, $\Box^*$ records truth at all indexes below the index of all evaluation, and $\Diamond^*$ records truth at some index below the index of evaluation. Second, we use $\blacksquare$ and $\blacklozenge$ as the modalities associated to the equivalence relation which j in J^γ bears to k in J^δ precisely when $\gamma = \delta$. Then, the necessity operator $\blacksquare$ records truth in all substructures of B^γ when evaluated at an index j in J^γ. Third, we use $\boxdot$ and $\diamondsuit$ for the modality associated to the relation which j in J^γ bears to k in J^δ precisely when B^γ_j and B^δ_k have the same underlying domain. Fourth, we use $\boxdot$ and $\diamondsuit$ for the modal operator which j in J^γ bears to k in J^δ precisely when both $\gamma = \delta$ and k is the index associated B^δ itself (since this relation is functional, note that $\boxdot$ and $\diamondsuit$ are equivalent).

It can be worthwhile to look at a concrete example. Suppose that our original first-order signature L just consists of a unary relation symbol P which we draw in orange, and a symmetric binary relation Q which we draw with red lines between the two nodes. Suppose that we start with three L-structures $B^\alpha, B^\beta, B^\gamma$ displayed in Figure 3.3, which have underlying domain just consisting of the three distinct numbers $1, 2, 3$.

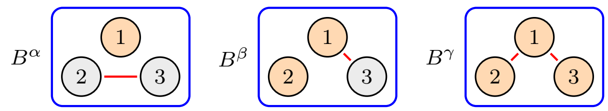

Figure 3.3: Three first-order L-structures

Then for each $\delta \in \{\alpha, \beta, \gamma\}$ we have that there are seven non-empty subsets of B^δ, and hence seven L-substructures (since, again, L is relational). Then we can draw these substructures as follows, where the partial order $\leq^\alpha \sqcup \leq^\beta \sqcup \leq^\gamma$ is the transitive closure of the drawn blue arrows in Figure 3.4. For each $\delta \in \{\alpha, \beta, \gamma\}$, we take these substructures $B^\delta_1, \ldots, B^\delta_7$ to be enumerated as the words are on a page, moving from left to right and then down row by row, so that B^δ_1 is identical qua L-structure to B^δ, and so that B^δ_7 has domain consisting of the singleton $\{3\}$. To do just do some simple examples, consider the statement "v is orange" (or, more formally $P(v)$). Suppose that we evaluate relative to a variable assignment which sends v to 2. Then evaluated at any B^α_i, we have that "v is orange" is false and so "$\blacklozenge$ (v is orange)" is likewise false at any index B^β_i. However, at any B^β_i, we have that "v is orange" is true and so back at an index B^α_i we have that "◈ (v is orange)" is true. Likewise, consider the statement "u is red-related" where this is an abbreviation for "there is a something with a red line from u to it." Suppose that we evaluate relative to a variable assignment which sends y to 1 and z to 3. Then at B^γ_5 (the world with domain $\{1\}$) we have that "y is red-related" is false but it is true at B^γ_2 (the world with domain $\{1, 2\}$) since 2 exists at that world. Hence "$\blacklozenge$ (y is not red-related)" is true at B^γ_2, and "$\Diamond$ (y is red-related)" is true at B^γ_5.

This general approach seems to model well Kratzer's idea of thinking about Barwise's situations as small parts of worlds and Davidsonian events as minimal parts of worlds.[29] In

[29]See Kratzer (1989), Kratzer (2012, Chapter 5), and Kratzer (2019, §9).

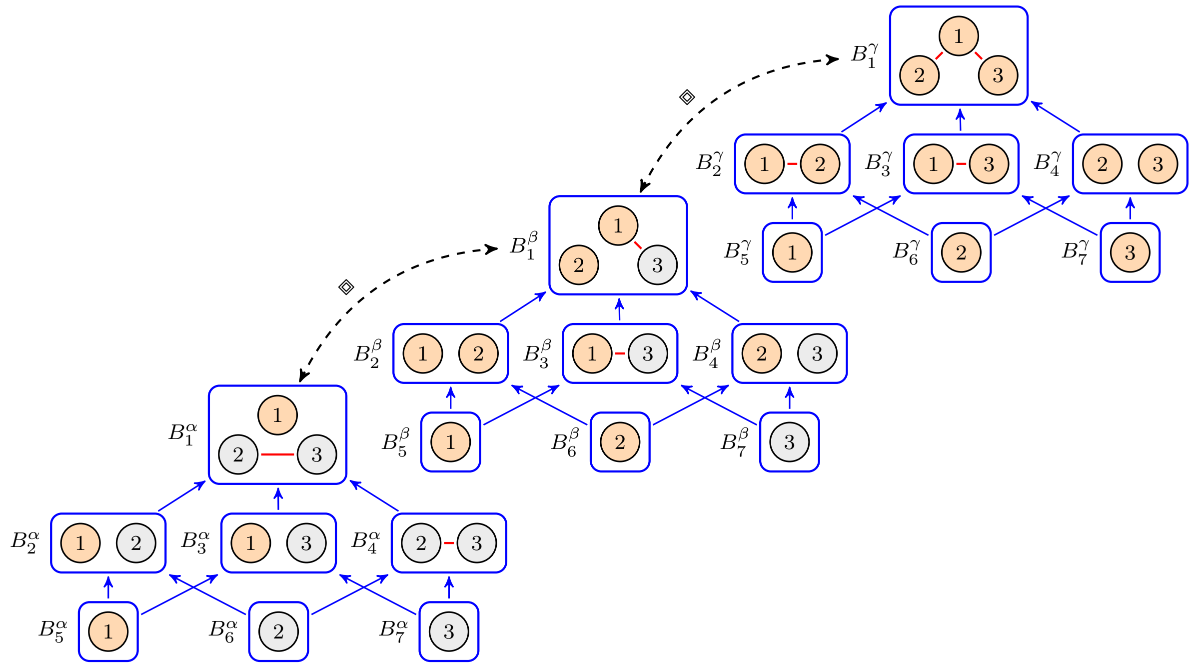

Figure 3.4: The directed families of L-substructures, unioned together

Kratzer's own formal semantics, this is only handled at the propositional level, with a fixed set of situations equipped with a partial order, on which the maximal elements correspond to worlds and the minimal elements correspond to events.[30] However, often when this framework is applied – such as by Elbourne in his treatment of definite descriptions[31] – one inevitably starts using a higher-order predicate logic structure.

One difference between our semantics and that of Kratzer and Elbourne is that our semantics, following Montague, hides the indexes from quantification in the object-language. Kratzer and Elbourne are convinced, due to arguments of Cresswell, that natural language contains sentences whose expressive power is equal to that of quantification over worlds.[32] Some of Cresswell's arguments pertain to sentences which are ostensibly recalcitrant to formalization in the absence of quantification over worlds. The standard examples pertain to variants of

[30]See Kratzer (1989, pp. 614-616), Kratzer (2012, p. 117).

[31]See Elbourne (2013). See, e.g., the semantic value of the definite article in Elbourne (2013, p. 47), and his use of the Kratzerian framework of situations and worlds on Elbourne (2013, p. 32).

[32]Compare Kratzer (2019, §5) and Elbourne (2013, p. 36).

"It might have been that everyone actually rich was poor."[33] But, as is familiar, this can be treated with the aid of the Vlach operators as "$\uparrow \Diamond \ \forall \ x \ ((\downarrow \mathsf{Rich}(x)) \rightarrow \mathsf{Poor}(x))$."

This then raises the question of whether, given the semantics on offer here, any of the other expressions which Kratzer and Elbourne are interested in require quantification over situations. Much of what is distinctive in Elbourne's work goes through a postulated Q-morpheme whose semantic value can be written in the metalanguage of our semantics as follows, where p is a variable of type $\langle s,t\rangle$ (i.e. the type of propositions):[34]

$$[\![\mathtt{Q}p]\!]_{\mathcal{M},i,\sigma k,g} = 1 \text{ iff there is } j \text{ such that } i \leq j \leq k \text{ and } [\![\smile p]\!]_{\mathcal{M},j,\sigma,g} = 1$$

But these kinds of operators are familiar from propositional temporal logics where the indexes are intervals of time $[i,k] = \{j \in \mathbb{R} : i \leq j \leq k\}$. And a natural part of the study of these propositional logics is that operators corresponding to geometric operations on intervals are not expressible in more elementary propositional temporal logics, with just forward looking and backwards looking operators.[35]

Here is a simple proof that the Q-operator is inexpressible in propositional modal logic with a Vlach operator. For the moment, let us simply take models $\mathcal{M}$ of modal propositional logic to be given by a quadruple $\langle W, \leq, \geq, V\rangle$ where W is the set of worlds, one accessibility relation is given by a partial order $\leq$, the other is given by its mirror $\geq$, and the valuation of basic propositional letters if given by V. We take $\leq$ to be governing $\Box$ and $\Diamond$, while we take $\geq$ to be governing $\blacksquare$ and $\blacklozenge$, as above. Then consider the sentence θ given by $\uparrow(\blacksquare(p \rightarrow \mathtt{Q}(q)))$.

[33]See Cresswell (1990, pp. 34 ff).

[34]See Elbourne (2013, p. 35). As we have written it, Q takes inputs p of type $\langle s,t\rangle$ and outputs an element Qp of type t, and so Q itself has type $\langle\langle s,t\rangle,t\rangle$. Elbourne's Q-morpheme takes inputs x of type e and X of type $\langle s,\langle e,t\rangle\rangle$ respectively, and outputs an element of type t. But given such x and X, one can consider $p = {}^\frown((\smile X)(x))$. Further, Elbourne's Q-morpheme takes two situation inputs i,k, but we have written these respectively as the index of evaluation and the first stored index. Such a move would not be justifiable in all contexts, but in Elbourne's paradigmatic applications, these are evidentally the roles occupied by the indexes he employs. A final added complication is that Elbourne uses the type of minimality operator familiar from the received semantics on counterfactuals. But this could be added onto our semantics too with little extra difficulty.

[35]See Halpern and Shoham (1991) and Marx and Venema (1997, Chapter 4).

This is a propositional schema associated to a pattern in predicate logic which Elbourne frequently uses, where the antecedent stands in for an instance of the subject-phrase and the consequent stands in for an instance of the predicate-phrase– e.g. for "All F's are G's", one might have p being $F(x)$ and q being $G(x)$, for some given x whose value is fixed by the variable assignment (since we are presently in the propositional fragment, there is no variable assignment). The sentence θ has the following truth conditions, as one can easily check:

$$\mathcal{M}, s, \sigma \vDash \theta \text{ iff } \forall\ s' \leq s\ (s' \in V(p) \Rightarrow \exists\ s'' \in [s', s]\ s'' \in V(q))$$

In this, we again use the familiar notation $[s', s] = \{t : s' \leq t \leq s\}$ for intervals with endpoints.

To see that the Q-operator is inexpressible in propositional bimodal logic with a Vlach operator, we show that θ is not preserved under bisimulations. Consider the following two models, $\mathcal{A}$ to the left and $\mathcal{B}$ to the right, shown in Figure 3.5.

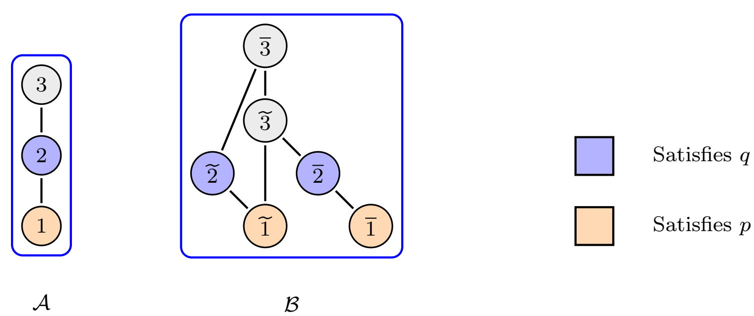

Figure 3.5: Countermodel to the expressibility of Q

The ordering $\leq$ on $\mathcal{A}$ is the same as the natural ordering on the natural numbers, with higher numbers displayed higher in the diagram. The ordering $\leq$ on $\mathcal{B}$ is transitive closure of the

Hasse diagram displayed. Note that in $\mathcal{B}$, we have the following facts about the ordering:

$$\overline{n} \leq \widetilde{m} \text{ iff } m = 3 \tag{3.26}$$

$$\widetilde{n} \leq \overline{m} \text{ iff } n = 1 \text{ and } m = 3 \tag{3.27}$$

The valuation on $\mathcal{A}$ and $\mathcal{B}$ are as displayed: the 3's do not satisfy any atomics, while the 2's satisfy only q and the 1's satisfy only p.

To finish the argument, note that $\mathcal{A}, 3 \vDash \theta$ but that $\mathcal{B}, \overline{3} \nvDash \theta$. For, in $\mathcal{A}, 3$, one has that θ is true since the only thing which has p is 1, and there is something between it and 3 which has q, namely 2. However, in $\mathcal{B}, \overline{3}$, one has that $\widetilde{1} \leq \overline{3}$ has p, but there is nothing in the interval $[\widetilde{1}, \overline{3}]$ which has q.

Despite this, the surjection $f : B \to A$ given by $f(\widetilde{n}) = n$ and $f(\overline{n}) = n$ is a bounded morphism, so that $\mathcal{B}$ and $\mathcal{A}$ are bisimilar via this function.[36] Clearly the condition on atomics is satisfied, by the above remarks on the valuations. Further, (3.26)-(3.27) implies that $f : \mathcal{B} \to \mathcal{A}$ respects the ordering. Finally, for the back condition, note that $f(\widetilde{n}) \leq m$ implies $n \leq m$ and hence $\widetilde{n} \leq \widetilde{m}$ and $f(\widetilde{m}) = m$. (A similar argument works with $\widetilde{\cdot}$ replaced by $\overline{\cdot}$, and with $\leq$ replaced by $\geq$).

3.7 Inverse Limits

3.7.1 Relevant aspects of the inverse limit construction

Now we turn to inverse families and inverse limits. Given a category, recall that an *inverse family* in a category is a family of objects B_i indexed by elements i of the directed partial order I and maps $h_{ji} : B_j \to B_i$ for $j \geq i$ such that $h_{ii} : B_i \to B_i$ is the identity map and if

[36] See Blackburn et al. (2001, p. 59 Definition 60; p. 66, Proposition 2.19.(iv)).

$k \geq j \geq i$ then $h_{ki} = h_{ji} \circ h_{kj}$. Finally, an *inverse limit* is an object D such that there are maps $h_i : D \to B_i$ which commute with the h_{ji}-maps for $j \geq i$ by $h_{ji} \circ h_j = h_i$, and such that for any other object D' and maps $h'_i : D' \to B_i$ with this property there is a unique map $h' : D' \to D$ such that $h_i \circ h' = h'_i$.[37] One typically records this information in the diagram shown in Figure 3.6.

$$\begin{tikzcd}
& D' \arrow[ddl, bend right, "h'_j"'] \arrow[d, dotted, "\exists\,!\,h'"'] \arrow[ddr, bend left, "h'_i"] & \\
& D \arrow[dl, "h_j"'] \arrow[dr, "h_i"] & \\
B_j \arrow[rr, "h_{ji}", "j\geq i"'] & & B_i
\end{tikzcd}$$

Figure 3.6: The inverse limit of the inverse family B_i

In the category of first-order structures in a given first-order signature L with homomorphisms in that signature, inverse limits of inverse families exist, provided the language contains a constant symbol. In this case, the inverse limit D has the following underlying set:

$$D = \{x \in \prod_i B_i : \text{for all } i, j \text{ in } I \text{ with } j \geq i, \text{ one has } h_{ji}(x_j) = x_i\} \tag{3.28}$$

and the maps $h_i : D \to B_i$ are given by the projection functions $h_i(x) = x_i$. Context will indicate which variables are being reserved for elements of the product $\prod_i B_i$, and we will be using the i-subscripts for the i-th projection. That is, in what follows if we are using x, y, z etc. for elements of the product, then x_i, y_i, z_i etc. refer to the i-th projection function applied to these elements. Note that the underlying set D from (3.28) is non-empty if there

[37]The inverse limit is likewise ubiquitous– see e.g. (Lang, 2002, pp. 49ff, 159 ff) for algebra, and (Sakai, 2013, pp. 204 ff) for topology.

is a constant symbol c in L since $h_{ji}(c^{B_j}) = c^{B_i}$ by definition of an L-homomorphism. In all canonical examples, both h_{ji} and h_i are surjective, and we assume this in what follows.

The structure on the inverse limit is given by restriction on the natural structure on the product $\prod_i B_i$ as an L-structure for constant and function symbols:

- For constant symbols, it is $(c^D)_i = c^{B_i}$.
- For binary relation symbols, it is defined so that $P^D(x,y)$ iff $\forall\, i\ P^{B_i}(x_i, y_i)$ and similarly for other arities.
- For binary function symbols f, one has that $(f^D(x,y))_i = f^{B_i}(x_i, y_i)$ and similarly for other arities.

If D' is an L-substructure of D, then its interpretation of the L-signature will just be by restriction of that of D. Further, in this case, we denote $h'_i = h_i \upharpoonright D$. In what follows, we restrict attention to D' such that $h'_i : D' \to B_i$ is surjective for all i in I. This is because the construction of the next section only requires this much of the inverse limit, and since the inverse limit is itself a special case where $D' = D$.

3.7.2 Associating frames and models to inverse limits

Let us now define a signature $\widetilde{L}$ in the sense of §3.3.3 from the first-order signature L. For each L-symbol $\mathbf{s}$ we create a copy $\widetilde{\mathbf{s}}$ which serves as the graph of $\mathbf{s}$, modulo currying. For a constant symbol c in L, we define $\widetilde{c}$ in $\widetilde{L}$ to be of type $\langle e, t\rangle$. For a binary relation P of type of L, we define $\widetilde{P}$ to be of type $\langle e, \langle e, t\rangle\rangle$, and similarly for other arities. For a binary function symbol f, we define $\widetilde{f}$ to be of type $\langle e, \langle e, \langle e, t\rangle\rangle\rangle$, and similarly for other arities. We also add to $\widetilde{L}$ a constant symbol $\widetilde{=}$ and $\approx$ of type $\langle e, \langle e, t\rangle\rangle$.

Since $h_{ji} : B_j \to B_i$ and $h'_i : D' \to B_i$ are surjective, we build a base frame by choosing injections $S_i : B_i \to D'$ such that $h'_i \circ S_i = id_{B_i}$ and by setting $S_{i,e} = S_i$. As before, we set $S_{i,s} : B_{i,s} \to D'_s$ equal to the identity map on I.[38] Hence our base frame is $\langle I, \leq, B, D', S \rangle$, and we extend it to a full, minimal, injective frame $\langle I, \leq, B, D', S, T \rangle$ using Proposition 3.1.

Now we build a model $\mathcal{M}$ in signature $\widetilde{L}$. If c is a constant symbol of L, then we define $\widetilde{c}^{\mathcal{M}}(i)(x) = 1$ iff $h'_i(x) = c^{B_i}$. If P is a binary relation symbol of L, then we define $\widetilde{P}^{\mathcal{M}}(i)(x,y) = 1$ iff $P^{B_i}(h'_i(x), h'_i(y))$, and similarly for other arities. If f is a binary function symbol of L, then we define $\widetilde{f}^{\mathcal{M}}(i)(x)(y)(z) = 1$ iff $f^{B_i}(h'_i(x), h'_i(y)) = h'_i(z)$, and similarly for other arities. Finally, we define $\cong^{\mathcal{M}}(i)(x)(y) = 1$ iff $h'_i(x) = h'_i(y)$, and $\approxeq^{\mathcal{M}}(i)(x)(y) = 1$ iff $h'_i(x), h'_i(y)$ satisfy the same L-formulas without parameters in B_i. Since these are identity-like relations, we write $u \cong v$ and $u \approxeq v$. Hence, if assignment g assigns variable u of type e to element x of D' and assigns variable v of type e to element y of D', then one has $[\![u \cong v]\!]_{\mathcal{M},i,\sigma,g} = 1$ iff x, y project down to the same object in B_i under h'_i. Likewise, under this assignment, one has $[\![u \cong v]\!]_{\mathcal{M},i,\sigma,g} = 1$ iff x, y project down to L-indiscernible objects in B_i under h'_i.

Note that since the model is full, for every symbol c of $\widetilde{L}$ we have that the constant symbol $\widetilde{c}$ of type a is such that $\widetilde{c}^{\mathcal{M}}$ is an element of $D'_{\langle s,a \rangle}$.

Now we show that the L-structure on the inverse limit D is definable in the modal language $\widetilde{L}$ in $\mathcal{M}$.

Note that when g assigns u to x, we have the following, for all i:

$$c^{D'} = x \text{ iff } [\![\Box(\widetilde{c}(u) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$$

[38]Again, it might also be natural to consider setting $B_{i,s} = R[i]$ and let $S_{i,s} : B_{i,s} \to I$ be the restriction of the identity map.

First, suppose that $c^{D'} = x$. Then for all j, we have $x_j = c^{B_j}$ and hence $h'_j(x) = c^{B_j}$ for all $j \geq i$ and thus $\widetilde{c}^{\mathcal{M}}(j)(x) = 1$ for all $j \geq i$ and thus $[\![(\widetilde{c}(u) = 1)]\!]_{\mathcal{M},j,\sigma,g} = 1$ for all $j \geq i$, and thus $[\![\Box(\widetilde{c}(u) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$. Conversely, suppose that $[\![\Box(\widetilde{c}(u) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$. Then $\widetilde{c}^{\mathcal{M}}(j)(x) = 1$ for all $j \geq i$. Then $h'_j(x) = c^{B_j}$ for all $j \geq i$. Suppose that k is given. Choose $j \geq k, i$. Then $x_j = c^{B_j}$. Then $h'_j(x) = c^{B_j}$. Then $h'_k(x) = h_{jk}(h'_j(x)) = h_{jk}(c^{B_j}) = c^{B_k}$. Hence for all k, we have $x_k = c^{B_k}$. Thus $c^{D'} = x$.

Likewise, note that when g assigns u to x and v to y, we have the following, for a binary relation P from L, and for all i:

$$P^{D'}(x,y) \text{ iff } [\![\Box(\widetilde{P}(u,v) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$$

First suppose that $P^{D'}(x,y)$. Then for all j one has $P^{B_j}(x_j,y_j)$. Then for all $j \geq i$ one has $P^{B_j}(x_j,y_j)$. Then for all $j \geq i$ one has $[\![\widetilde{P}(u)(v) = 1]\!]_{\mathcal{M},j,\sigma,g} = 1$. Then $[\![\Box(\widetilde{P}(u,v) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$. Conversely, suppose $[\![\Box(\widetilde{P}(u,v) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$. Then for all $j \geq i$ one has $[\![\widetilde{P}(u)(v) = 1]\!]_{\mathcal{M},j,\sigma,g} = 1$. Then for all $j \geq i$ one has $\widetilde{P}^{\mathcal{M}}(j)(x)(y) = 1$, which happens iff $P^{B_j}(h'_j(x), h'_j(y))$. Suppose that k is given. Choose $j \geq k, i$. Then $P^{B_j}(h'_j(x), h'_j(y))$. Then $P^{B_k}(h_{jk}(h'_j(x)), h_{jk}(h'_j(y)))$. Then $P^{B_k}(h'_k(x), h'_k(y))$. Hence, for all k we have $P^{B_k}(x_k, y_k)$. Then $P^{D'}(x,y)$.

Likewise, note that when g assign u to x and v to y and w to z, we have the following, for a binary function f from L, for all i:

$$f^{D'}(x,y) = z \text{ iff } [\![\Box(\widetilde{f}(u,v,w) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$$

First suppose that $f^{D'}(x,y) = z$. Then for all j one has $f^{B_j}(x_j,y_j) = z_j$. Then for all $j \geq i$ one has that $f^{B_j}(h'_j(x), h'_j(y)) = h'_j(z)$. Then for all $j \geq i$ one has that $[\![\widetilde{f}(u,v,w) = 1]\!]_{\mathcal{M},j,\sigma,g} = 1$. Then $[\![\Box(\widetilde{f}(u,v,w) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$. Conversely, suppose $[\![\Box(\widetilde{f}(u,v,w) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$. Then for all $j \geq i$ one has that $f^{B_j}(h'_j(x), h'_j(y)) = h'_j(z)$. Let k be given. Choose $j \geq i, k$.

Then $f^{B_j}(h'_j(x), h'_j(y)) = h'_j(z)$. Then $f^{B_k}(h'_k(x), h'_k(y)) = f^{B_k}(h_{jk}(h'_j(x)), h_{jk}(h'_j(y))) = h_{jk}(f^{B_j}(h'_j(x), h'_j(y))) = h_{jk}(h'_j(z)) = h'_k(z)$. Hence for all k we have $f^{B_k}(h'_k(x), h'_k(y)) = h'_k(z)$. Thus $f^{D'}(x, y) = z$.

Note that when g assigns u to x and v to y, we have the following, for all i:

$$x = y \text{ iff } [\![\Box(\widetilde{=}(u, v) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$$

First suppose that $x = y$. Then for all j one has $x_j = y_j$, or what is the same: $h'_j(x) = h'_j(y)$. Then for all $j \geq i$ one has $h'_j(x) = h'_j(y)$. Then $[\![\widetilde{=}(u, v) = 1]\!]_{\mathcal{M},j,\sigma,g} = 1$ for all $j \geq i$. Then $[\![\Box(\widetilde{=}(u, v) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$. Conversely, suppose that $[\![\Box(\widetilde{=}(u, v) = 1)]\!]_{\mathcal{M},i,\sigma,g} = 1$. Then for all $j \geq i$ one has $h'_j(x) = h'_j(y)$. Let k be given. Let $j \geq i, k$. Then $h'_j(x) = h'_j(y)$. Then $h'_k(x) = h_{kj}(h'_j(x)) = h_{kj}(h'_j(y)) = h'_k(y)$. Then $x_k = y_k$ for all k, and hence $x = y$.

In this section, we have shown that, for all indexes i, an atomic $\mathbf{s}$ in D' holds iff $\Box\widetilde{\mathbf{s}}$ holds at i. However this can fail even for negated atomics, as can be seen by a simple example, displayed in Figure 3.7. Suppose that we have just three indexes i, j, k, ordered by $k \geq i, j$ with i, j being incomparable. If B_i, B_j, B_k consist of the numbers as displayed in the diagram, then elements of the inverse limit D are triples $x = \langle x_i, x_j, x_k \rangle$ such that $h_{ki}(x_k) = x_i$ and $h_{kj}(x_j) = x_i$, where the maps h_{ki} and h_{kj} are as drawn in the diagram. Further, we assume that we have a unary predicate P which is drawn in orange in the diagram. Consider an assignment which assigns v to $x = \langle 6, 9, 2 \rangle$. Then we have $D \vDash \neg P(v)$ since 2 is not orange. But we have $[\![P(v)]\!]_{\mathcal{M},j,\sigma,g} = 1$ since 9 is orange, and so $[\![\Box\neg P(v)]\!]_{\mathcal{M},j,\sigma,g} = 0$.

3.7.3 Barcan and related principles

We follow the enumeration of principles from the analogous discussion in §3.6.4 about direct limits.

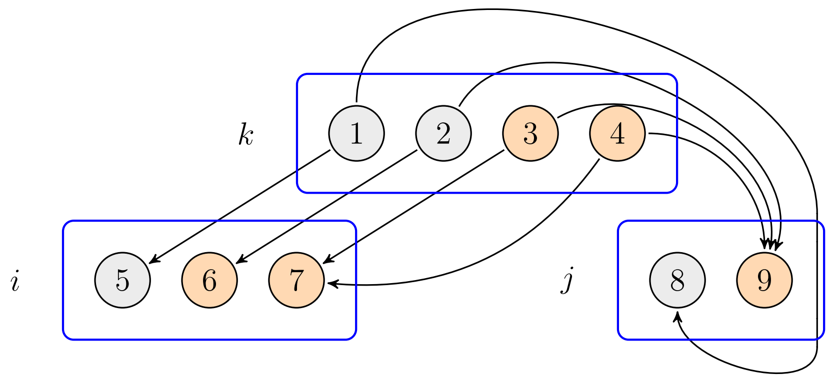

Figure 3.7: The non-necessity of non-atomics

First, consider the validity of $\Diamond \mathsf{E}_e(v)$, where v is a variable of type e. This is invalid precisely when there is an upwards-closed non-empty subset I_0 of indexes such that D' is a proper superset of $\bigcup_{j\in I_0} T_{i,e}(B_{i,e})$. This can happen for cardinality reasons when e.g. $I_0 = I$ is countable and D' is uncountable and each B_i is countable. By contrast, we have:

Proposition 3.15. *Suppose that I is countably infinite and D' is countable. Then there is a choice of maps $S_i : B_i \to D'$ such that $h'_i \circ S_i = id_{B_i}$ such that for every upwards-closed non-empty subset I_0 of I, one has that $D' = \bigcup_{j\in I_0} S_i(B_i)$.*

Proof. Fix an enumeration $i_1, i_2, \ldots$ of I without repetition, and fix an enumeration $y_1, y_2, \ldots$ of D' so that each element of D' is repeated infinitely many times in this enumeration. At stage n of the construction, we use directedness and the infinitude of I to choose index $j_n > i_1, \ldots, i_n$, and we use choice to pick $S_{j_n} : B_{j_n} \to D'$ such that $h'_{j_n} \circ S_{j_n} = id_{B_{j_n}}$ and such that y_n is in $S_{j_n}(B_{j_n})$. The latter can be done simply by choosing y_n to be the value of $x_n = h_{j_n}(y_n)$ under the map S_{j_n}. For indexes i not equal to any j_n, we use choice to select $S_i : B_i \to D'$ such that $h'_i \circ S_i = id_{B_i}$.

Suppose that I_0 is an upward-closed non-empty subset of I. It suffices to show that each y in D' is in the set $\bigcup_{j\in I_0} S_i(B_i)$. Since I_0 is non-empty, choose element i_m in I_0. Since y

is repeated infinitely many times in the enumeration of D', choose $n > m$ such that $y_n = y$. Then by construction y_n is in $S_{j_n}(B_{j_n})$. Since $j_n > i_m$ and I_0 is upwards closed, j_n is also in I_0. □

A natural way to discharge the hypothesis that D' is countable is just to take a countable elementary structure of the inverse limit D.

Second, in many natural circumstances, we can organize the base frame so that Converse Barcan holds at type e:[39]

Proposition 3.16. *Suppose that I is the natural numbers with its ordering. Then there is a choice of maps $S_i : B_i \to D'$ such that both $h'_i \circ S_i = id_{B_i}$ as well as $S_i(B_i) \subseteq S_j(B_j)$ whenever $j \geq i$. Further, under the additional hypothesis that D' is countably infinite and each B_i is finite and $\sup_i |B_i| = \infty$, these can be defined so that $D' = \bigcup_{j \in I} S_i(B_i)$.*

Proof. The construction is by recursion on n, where we handle $i \leq n$ by stage n of the construction. For $n = 0$, we just choose $S_0 : B_0 \to D'$ such that both $h'_0 \circ S_0 = id_{B_0}$. Suppose that the construction has been completed up to stage n. First note that the map $g_n = h'_{n+1} \circ S_n : B_n \to B_{n+1}$ is injective. For, suppose that b, b' in B_n are such that $g_n(b) = g_n(b')$. Then one has

$$\begin{aligned} b = h'_n(S_n(b)) &= h_{n+1,n}(h'_{n+1}(S_n(b))) = h_{n+1,n}(g_n(b)) \\ =& h_{n+1,n}(g_n(b')) = h_{n+1,n}(h'_{n+1}(S_n(b'))) = h'_n(S_n(b')) = b' \end{aligned}$$

Let $B'_{n+1} \subseteq B_{n+1}$ be the range of the injection $g_n : B_n \to B_{n+1}$. Note that if x is in B'_{n+1} and $b = g_n^{-1}(x)$ then $g_n(b) = x$ and $h'_{n+1}(S_n(b)) = x$ and hence $h'_{n+1}(S_n(g_n^{-1}(x))) = h'_{n+1}(S_n(b)) = x$. Define $S_{n+1} : B_{n+1} \to D$ by $S_{n+1}(x) = S_n(g_n^{-1}(x))$ if x is in B'_{n+1}, and if otherwise then

[39] By Grätzer (1979, pp. 133, Lemma 7), if an index set is countably infinite, then its inverse limit is isomorphic to one whose underlying set is the natural numbers with its ordering. (If the signature includes relation symbols, one again has to assume that the maps between the models are embeddings).

simply choose $S_{n+1}(x) = y$ for some point y of D' with $h_{n+1}(y) = x$. Then we have defined $S_{n+1} : B_{n+1} \to D'$ such that $h'_{n+1} \circ S_{n+1} = id_{B_{n+1}}$. Further, by construction we have

$$S_n(B_n) = S_n(g_n^{-1}(B'_{n+1})) = S_{n+1}(B'_{n+1}) \subseteq S_{n+1}(B_{n+1})$$

Now, further suppose that D' is countably infinite and each B_i is finite and $\sup_i |B_i| = \infty$. Hence, there are infinitely many indexes n where B_{n+1} has higher cardinality than each B_i for $i \leq n$. At these places, we can use the second case break in the definition of S_{n+1} to place elements of D' in the range of S_{n+1} if they are not already in the ranges of the S_i for $i \leq n$. □

Finally, to see why Barcan fails at type e, consider the example displayed in Figure 3.8.

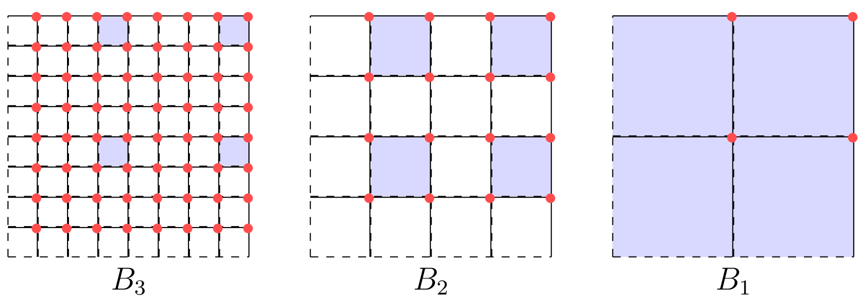

Figure 3.8: Failure of the Barcan formula for type e

We work with divisions of the half-open unit square $(0, 1] \times (0, 1]$ into smaller and smaller squares. In particular, for $i \geq 1$, let B_i consist of $2^i \times 2^i$ squares which are translates of the half-open square $(0, 2^{-i}] \times (0, 2^{-i}]$. The half-open nature of the square is represented by the device, familiar from elementary point-set topology, of using dashed lines on the leftmost and bottom side of the squares, indicating that the points on these lines are not in the squares. For instance, in B_1, the horizontal line running through the middle of B_1 needs to be seen as the top of the southwestern and southeastern squares and so is included in these, but is likewise excluded from the northwestern and northeastern squares. Let P be a unary predicate which we represent in the diagram by coloring in a half-open square with the color blue. In each B_i, exactly four squares have the P-predicate and are drawn blue: the northeast square is blue, the square to the immediate southwest of the center is blue, as is the the northmost square above this one and the eastmost square to the right of this one. For $j \geq i$ let $h_{ji} : B_j \to B_i$ be the map which sends a smaller square in B_j to the unique bigger square in B_i in which it lies. The inverse limit D can be identified with points in the half-open square, and the four blue predicates are the four *points* $(1,1)$, $(\frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 1)$, and $(1, \frac{1}{2})$ (the smaller and smaller four blue squares turn into four points in the limit). The maps $S_i : B_i \to D$ are given simply by selecting a point within each square, and in the diagram we choose the northeastern corner of each square, and mark this in red. This gives us a failure of Barcan, since we have $[\![\forall\ x\ \Box\ Px]\!]_{\mathcal{M},1,\sigma,g} = 1$, since each of the red points in B_1 is always in a blue square, for all $i \geq 1$. But we have $[\![\exists\ x\ \neg Px]\!]_{\mathcal{M},3,\sigma,g} = 1$ since we can choose one of the red dots not in a blue square. However, converse Barcan is maintained here since the red dots in B_1 are still there in B_2 etc. Further, if one replaces D by the countable elementary substructure D' consisting of points with dyadic rational coordinates, then one can further maintain the validity of $\Diamond \mathsf{E}_e(v)$, since all the points with dyadic rational coordinates in the half-open unit square will eventually be covered by red dots.

3.7.4 Inverse limits and perceptual reports

Suppose that φ is a well-formed expression of type t with exactly one free variable v of type e, which we display as $\varphi(v)$. Then define the following well-formed expression of type t, where u is new variable of type e:

$$\exists\!\!\!\exists \varphi(v) \equiv \Sigma\, u\ (u \approxeq v \wedge \varphi(u))$$

Note that v is still free in all of these, and while the notation $\exists\!\!\exists$ is chosen to resemble the existential quantifier, it does not bind the displayed variable v. This definition can be easily generalized to the case where φ has more than one free variable of type e.

Van der Does and van Lambalgen provide a semantics for perceptual reports based off of inverse limits and using this well-formed expression.[40] Their preferred formulation has the same logical form as $\exists\!\!\exists\varphi(v)$ above, but where the non-logical symbols in φ are how the inverse limit itself interprets its first-order signature.[41] By virtue of the results in §3.7.2, we can express how the inverse limit itself interprets the atomics of its first-order signature by using the necessity operator. Hence, when φ is limited to conjunctions of atomics, then the notion that Van der Does and van Lambalgen were interested in is expressible by $\exists\!\!\exists(\Box\varphi)(v)$. Further, using the Vlach operators as in Proposition 3.2, one can express this using only the actualist quantifiers (when the hypothesis that $\Diamond\mathsf{E}_e(v)$ is valid in the model).

For φ being an conjunction of atomics, their semantics for a perceptual demonstrative like "the agent sees this φ" is thus given by $\exists\!\!\exists(\Box\varphi(v))$, where the assignment assigns the variable v the same value x in the outer domain as the demonstrative.[42] Relative to such an assignment, it is true in a model at indexes i, σ iff there is an element y in the outer domain such that the projections $h_i'(x)$ and $h_i'(y)$ are indiscernible as elements of the approximation B_i. In another place in the paper, they consider the variation where it is true in a model at indexes i, σ iff there is an element y in the outer domain such that the projections $h_i'(x)$ and $h_i'(y)$ are identical elements of the approximation.[43]

---

[40]See van der Does and Van Lambalgen (2000).

[41]See van der Does and Van Lambalgen (2000, p. 30). In their notation it is $\mathcal{M}, g \vDash \exists(\varphi \mid \mathcal{B}_i)$, which when φ is an atomic or conjunction of atomics has the same truth-conditions as does our $[\![\exists\!\!\exists((\Box\varphi)(v))]\!]_{\mathcal{M},i,\sigma,g} = 1$. Their equivalence relation $R_i(f,g)$ can be expressed, in our terms, as follows: for all free variables v of type e, one has that the projections of $f(v)$ and $g(v)$ down to B_i are indiscernible there.

[42]See van der Does and Van Lambalgen (2000, p. 49, equations (50)-(51)).

[43]See van der Does and Van Lambalgen (2000, p. 29). In their notation it is $\mathcal{M}, g \vDash \exists_i\varphi$, which when φ is an atomic or conjunction of atomics has the same truth-conditions as does our $[\![\exists\!\!\exists((\Box\varphi)(v))]\!]_{\mathcal{M},i,\sigma,g} = 1$. Their equivalence relation $E_i(f,g)$ can be expressed, in our terms, as follows: for all free variables v of type e, one has that the projections of $f(v)$ and $g(v)$ down to B_i are identical.

Hence, we can view the work of van der Does and van Lambalgen through the lens of the conservative extension of Montague offered here, so that no great deviation from traditional semantics is required for the logic of perceptual reports. Further, this allows us to compare their proposals for the semantics of perceptual reports to other well-known proposals. For instance, another proposal for the semantics of perceptual reports is due to Kraut, which builds in certain ways off of the work of Hintikka.[44] On this proposal, "the agent sees this φ" is true relative to a model, index i, and assignment if and only if "$\Sigma\, z\ (x = \smallsmile z \wedge \Box\varphi(\smallsmile z))$ is true at i, where z is a variable of type $\langle s, e\rangle$, and where x is assigned the same value as the demonstrative.[45] From this, one can see that the van der Does and van Lambalgen proposal implies the Kraut proposal: indeed, taking a value y from the outer domain which witnesses the former proposal, one can just set $z(i) = y$ (that is, z is the constant function y). However, at least formally, the Kraut proposal does not imply the van der Does proposal. For instance, take the model illustrated in Figure 3.9, with just two worlds $j > i$.

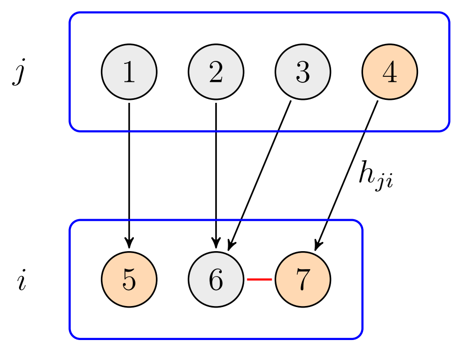

Figure 3.9: Kraut's proposal does not imply van der Does's

There is one unary relation P displayed in orange, and one symmetric binary relation Q displayed in red. There are four elements of the inverse limit, where we list the components

[44]See Kraut (1979), Kraut (1982), and Hintikka (1969b).

[45]See Kraut (1982, p. 288-289). Kraut gives the semantics for "the agent sees a φ" rather than "the agent sees this φ." Hence, in Kraut's own version, there is no "this" or variable x. However, in the earlier paper (Kraut, 1979, p. 210) he suggests such an addition.

with the B_j-element coming first and the B_i-element coming second:

$$D = \{a, b, c, d\}, \quad a = (1, 5), \quad b = (2, 6), \quad c = (3, 6), \quad d = (4, 7)$$

The presence of the red Q has the effect that 5 is distinguishable from both 6 and 7 at world i (since only 7 is orange and yet red-related to a non-orange element, at i).

Consider the function $z : \{i, j\} \to D$ given by $z(i) = a$ and $z(j) = d$. In what follows, set $x = a$. Then at world i it is true that $x = z(i)$ and hence $x = \smile z$. Further at i it is true that $\widetilde{P}(\smile z)$ since this has truth conditions $P^{B_i}(h_i(z(i)))$, and we have $h_i(z(i)) = h_i(a) = 5$, which is orange. Finally at j it is true that $\widetilde{P}(\smile z)$ since this has truth conditions $P^{B_i}(h_j(z(j)))$, and we have $h_j(z(j)) = h_j(d) = 4$, which is also orange. However, we claim that it is not the case that there is an element y of D which is indistinguishable from x at i, and such that $\Box(\widetilde{P}(x) = 1)$ is true at i. For, as mentioned before, the only thing indistinguishable from x at i is itself. And if $\Box(\widetilde{P}(x) = 1)$ was true at i, then $\widetilde{P}(x) = 1$ would be true at j. This has truth conditions $P^{B_j}(h_j(x)) = 1$, which is false since $h_j(x) = 1$ which is not orange. While this is a formal counterexample to the entailment, it might be within Kraut's rights to suggest that it goes against the spirit of his proposal. For, he did not want to consider arbitrary functions $z : I \to D$, but rather those such that "the various values of z taken together represent the typical syndrome of objects, the specified causal effects of which serve to functionally define a certain perceptual state [...]."[46] The idea is then that $z(j)$ plays the same functional role at j as $z(i)$ does at i.

[46]See Kraut (1982, p. 284).

Chapter 4

Reflection Principles and Semantic Entanglement

4.1 Introduction

Horsten and Leigh (2017) provide a philosophical argument in justification of the use of reflection principles to extend a modest base theory of truth to a strong compositional theory. While they show that their method works for a particular initial theory and the usual reflection principles, I consider whether the argument is compatible with a variety of base theories and a new reflection principle, and I show that applying the Horsten-Leigh technique in these cases almost always fails.

Horsten and Leigh single out a weak theory of Tarski biconditionals in support of their thesis, but there are a number of popular and well-studied formal theories of truth. Alongside the choice of base theory, I argue that there is also an alternative choice of reflection principle which is on equal footing with the orthodox one which Horsten and Leigh consider. In parallel with the standard reflection principle, I examine the theories with my new reflection principle to illustrate that this choice has drastic effects on the theories thereby obtained.

Outside of the connections to the Horsten-Leigh project, the study is important because it greatly expands the domain of available formal theories of truth. For what this chapter shows is that reflection principles are semantically entangled: just as the semantic paradoxes forced one to choose between a handful of candidate theories of truth, so too do they necessitate a choice between different reflection principles.

The outline of the paper is as follows. In §4.2, I contextualize the Horsten-Leigh project and provide an informal explanation of both the standard reflection principles and my new reflection principles. Following that, in §4.3 I formulate the basic definitions for formal theories of truth with reflection principles, which allows me to prove a few fundamental model-existence theorems. The next three sections (§4.4, §4.5 and §4.6) examine the extension of three of the most popular formal theories by reflection principles. Then in §4.7 I briefly show that the two types of reflection principles are independent and cannot always be used in conjunction. In §4.8 I provide a summary of how the results impact the Horsten-Leigh project and then consider some possible objections and replies on their behalf. Finally in §4.9, I note further directions for future study, highlighting a possible way out for Horsten and Leigh.

4.2 The Recent Debate and a New Reflection Principle

4.2.1 The Received Positions: Disquotation versus Composition

The study of formal theories of truth has its foundations in the attempt to model, or find adequate definitions for, the natural language concept of truth. Tarski's undefinability theorem placed critical limitations on the project (Tarski, 1936). This shifted the focus toward attempting to "capture the use of the concept" (Horsten and Leigh, 2017, p. 4) rather than providing an explicit definition for truth. The contemporary debate is then centered on clarifying this notion of use and deciding on the desiderata of such a formal theory – for instance, one might want truth to commute with 'and' so that 'p is true and q is true' holds precisely when '(p and q) is true' holds.

To accomplish this, we examine a basic background theory enriched with a predicate expressing truth. In most cases (ours among them), we take our base theory to be a weak version of arithmetic, which runs proxy for the rich scientific theory (or formalization of natural language) in which we want to understand the notion of truth.

In the debate, two principles are in tension: *disquotationalism*, the view that the notion of truth consists in a collection of Tarski biconditionals, where a sentence p holds just in case the sentence 'p is true' does; and the *composition of truth*, our intuition that the truth-predicate commutes with the logical connectives and is preserved by the rules of inference. The first view, under threat of Tarski's undefinability theorem, must be restricted to a smaller class of biconditionals. However, taking a smaller set of biconditionals as one's axioms is sometimes insufficient to derive the truth composition principles.

The relative weakness of disquotationalism then became the target of philosophers in the field. The debate centered on the interactions and justifications of Tarski biconditionals, compositional truth principles and reflection principles. One view was that a theory of truth should be strong enough to derive reflection principles for the base theory.[1] Simple disquotational approaches fail in this regard, so this was taken as motivation for adopting the compositional principles as axioms instead.[2]

[1]Stewart Shapiro is one advocate of this position. He argues that deflationists who "hold that the property of truth is metaphysically thin... must square their views with the fact that truth is not conservative over rich theories" (Shapiro, 1998, pp. 503-504).

[2]This yields a theory like CT, which does prove local reflection for its base theory of Peano Arithmetic. The explicit definition of the system CT can be found in Halbach (2011, Definition 8.35, p. 102) – it and its related subsystems are discussed throughout Part II.

Cieśliński (2010) reverses the argument: we are committed to at least the reflection principles of basic logic (i.e. an empty first-order theory), and adjoining these to a weak base theory brings the theory up to par with the compositional approach. This strategy is novel in that it deliberately gives up the conservativeness of the truth theory over the base theory (which Shapiro (1998) and Ketland (1999) took to be central to a deflationist viewpoint). Horsten and Leigh embrace this strategy and use it to bootstrap a modest collection of Tarski biconditionals to a compositional theory through the iterated application of reflection principles, to which we now turn our attention. Below, we follow Horsten and Leigh (2017) and Cieśliński (2010) in working in a classical setting.

4.2.2 Varieties of Reflection Principles and Their Warrant

Feferman (1991, p. 12-13) provides a canonical gloss of the term 'reflection principle' in the field of arithmetic: "axiom schemata... which express, insofar as is possible without use of the formal notion of truth, that whatever is provable in S is true." In accord with this definition, Horsten and Leigh (2017) identify three types of reflection principles[3]:

Local Reflection Schemas of the form $\mathrm{Bwb}_S(\ulcorner\varphi\urcorner) \to \varphi$ for $\mathcal{L}$-sentences φ and S an $\mathcal{L}$-theory.

Uniform Reflection Schemas of the form $\forall x(\mathrm{Bwb}_S(\ulcorner\varphi(\underline{x})\urcorner) \to \varphi(x))$ for $\mathcal{L}$-formulas φ with a parameter x and S an $\mathcal{L}$-theory.[4]

Global Reflection Axioms of the form $\forall\varphi(\mathrm{Bwb}_S(\ulcorner\varphi\urcorner) \to T(\ulcorner\varphi\urcorner))$, where φ ranges over $\mathcal{L}$-sentences and S is an $\mathcal{L}$-theory.

[3] There is evidently some tension between Feferman's explication of reflection and the formulation of global reflection, which makes explicit use of a formal notion of truth. However, depending on the desired metalogical properties, one is limited on how much can be expressed without such use.

[4] These can be extended to finite tuples of variables in an arithmetic setting with no gain in proof-theoretic strength. Here, the underline notation indicates the function substituting the object-language numeral which corresponds to the value of the term x.

Here, 'Bwb(x)' (abbreviating the German *beweisbar*, 'provable') is a predicate that captures the notion of provability from a set of axioms of arithmetic.[5] Recall that for sufficiently expressive languages, we can enumerate the formulas of the language and interpret predication of the numeric codes as predicating the expressions of the language itself. More concretely, to say 'Bwb($\ulcorner\varphi\urcorner$)' or '$\varphi$ is provable' is to state that there is a number which encodes a proof of φ, where we work with Hilbert-style proofs (sequences of formulas following from basic rules of inference). So a reflection principle licenses the inference from 'Bwb($\ulcorner\varphi\urcorner$)' to φ simpliciter (or in the case of global reflection, to $T(\ulcorner\varphi\urcorner)$). And what's important here is that such inferences don't follow in the standard axiomatizations of arithmetic, e.g. Peano arithmetic.

That last remark is one of the main reasons reflection principles are studied: they increase the proof-theoretic strength of axiomatic systems. In particular, Gödel's second incompleteness theorem informs us that systems like Peano arithmetic can't prove their own consistency (if consistent). However, it is commonly held (and endorsed by Horsten and Leigh) that if one accepts a theory as true, then one is committed to accepting that the theory is sound even if such statements are independent of the theory.[6] One expression of the soundness of a theory is its consistency statement Con(T), which is the formalization of the claim that there is no proof of an absurdity from the axiom system T. What reflection principles do is permit this inference in the object language.

Now, while discussing reflection in general, we saw above three types of reflection principles in practice. This categorization finds support in standard references like Halbach (2011).[7] Local reflection fits our informal examination of reflection principles as a schema. Uniform reflection extends local reflection from sentences (i.e. closed formulas) to formulas with free parameters. This allows a shift in scope. For example, suppose we were studying the formula

[5] When we want to track the particular system S we are working with, we subscript the predicate as $\mathrm{Bwb}_S(x)$.

[6] See Dean (2015) for discussion of the *implicit commitment thesis*.

[7] See section §22.1 (pp. 322-326) for a discussion of types of standard reflection principles.

'x is odd or x is even'. The only way to employ our local reflection principle would be to quantify over the expression internally: if it's provable that *for every* x x is odd or x is even, then *for every* x x is odd or x is even. However, uniform reflection lets us quantify on the outside: *for every* x, if it's provable that x is odd or x is even, then x is odd or x is even. It seems plausible that a proof about all numbers (the antecedent of local reflection) might differ from a series of proofs each about a given number (the antecedent of uniform reflection).

Global reflection principles express the same basic idea of reflection but are implemented in a much different manner. This difference is tangible in their formulation: local and uniform reflection come in the form of schemas comprising infinite sets of sentences, whereas global reflection principles take the form of singular axioms. For an individual working with axiomatic theories, this change has obvious appeal. In this formulation, each instance of the target sentence has been Gödel-coded and put into the scope of a predicate. Consequently, numeric quantification now in essence ranges over coded sentences.

The manner by which this is accomplished is by introducing a truth predicate '$T(x)$', where '$T(\ulcorner\varphi\urcorner)$' can be read as '$\varphi$ is true', which is properly speaking a distinct assertion from 'φ' simpliciter. The connection between the two assertions is an important theme in the grander debate to which this project belongs, but for our present purposes we needn't subscribe to any particular treatment of such expressions; we need only take for granted that there is some type of connection which legitimizes this formulation as a reflection principle of the same character. Now, the addition of another piece of metamathematical machinery might give pause to an individual working in pure arithmetic, but since our end goal is discussion of reflection principles *in formal theories of truth*, we can take global reflection to be on roughly equal footing as the other types.

In fact, Kreisel and Lévy (1968) argue that the global reflection principle is the primary formulation, with the local and uniform principles following (in justificatory grounding) from global reflection. Shapiro (1991, p. 117) summarizes Kreisel and Lévy's position by stating that "a given mathematician believes or accepts the individual instances of the first-order *scheme* only because she (already) believes or accepts the second-order *axiom*". Shapiro illustrates the point by comparison of the induction axiom of second-order arithmetic with the induction schema of first-order arithmetic. Thus far, we're working in just first-order logic, but a similar comparison can be made: global reflection can be formulated as a single axiom (by virtue of object-language quantification over formulas), and we obtain the local and uniform reflection schemes by instantiation of this prior axiom. So global reflection is more than a technical convenience; it provides the most direct link between reflection and truth.

The above formulations constitute the orthodox presentation of reflection principles in arithmetic. However, there is another type of reflection principle which seems to follow from the same considerations that motivated the standard ones. Recall that the general idea of such reflection principles was to infer from provability to truth. Accordingly, such principles are taken to "express the soundness" of a system (Dean, 2015, p. 3). Alternatively, one could consider converse principles, by means of which one infers from truth to some form of unrefutability. Our global reflection principle was of the form $\mathrm{Bwb}(\ulcorner\varphi\urcorner) \to T(\ulcorner\varphi\urcorner)$, looking within the scope of the quantifier. Contraposing that we obtain $\neg T(\ulcorner\varphi\urcorner) \to \neg\mathrm{Bwb}(\ulcorner\varphi\urcorner)$. Changing variables and substituting $\neg\psi$ for φ we have $\neg T(\ulcorner\neg\psi\urcorner) \to \neg\mathrm{Bwb}(\ulcorner\neg\psi\urcorner)$. If our metalinguistic notion of truth (i.e. the 'real' concept) is classical, then this is equivalent to $T(\ulcorner\psi\urcorner) \to \neg\mathrm{Bwb}(\ulcorner\neg\psi\urcorner)$.[8]

[8]As currently formulated, the use of the T predicate is required in the statement of a distinct dual reflection principle. A naive attempt to formulate a uniform version of dual reflection in the same schematic form results in an axiom equivalent to the standard uniform reflection principle. But we can see this an instance of a general phenomenon: schemas which are equivalent in one setting may turn out to be inequivalent in a broader setting. For instance, consider the simple schemas $I(A) \equiv A \leftrightarrow A$ and $J(A) \equiv A \leftrightarrow \neg\neg A$. It is well-known the two are equivalent that in the negative fragment of minimal logic (see Troelstra and Schwichtenberg (2000, p. 48 ff.)). But when we move to the more expansive setting of full minimal logic (or intuitionistic logic), they are provably inequivalent.

We can read this last statement as saying 'if ψ is true, then there is no proof to the contrary'. If the informal idea behind orthodox reflection was 'what follows from accepted axioms is true', then a supporter should also accept the formal rendition of 'what is true doesn't violate the axioms'. This new dual reflection principle has not been studied in the literature, so we will examine it, in conjunction with the standard one, to better understand which properties the two types share. Broadly, we will find that the dual reflection principle is not in general interderivable with the orthodox one but demonstrates similar proof-theoretic strength in terms of deriving consistency statements.

In what sense can this dual reflection principle be said to stand on equal footing with the orthodox formulation? In the context of this chapter, there are two roads to supporting this idea. If one accepts the popular account of reflection principles defended by Shapiro and Ketland, where the justification for reflection principles derives from the compositional truth principles, then any justification of the orthodox principle will apply to dual reflection for the two will be interderivable.[9] This is made precise later in Proposition 4.13, but the idea underlies the informal derivation given above.

If we wish to engage Horsten and Leigh (2017), however, this approach is not open to us, for the authors reject the notion that compositional truth principles are basic. On their view, reflection principles need not be justified; instead such principles are endowed with a distinctive epistemic status known as entitlement.[10] In accepting a theory, we are entitled to an implicit trust in that theory. Reflection principles make this explicit and hence "express our trust in theories" (Horsten and Leigh, 2017, p. 17). If one is also justified in their acceptance of a theory and possesses some further conceptual machinery (regarding, e.g., provability), then one can be entitled to the reflection principle as well. The question is thus: if one is justified in accepting a given theory and understands the requisite formal

[9]This is illustrated in the standard theorem that the compositional system CT proves global reflection (Halbach, 2011, p. 325).

[10]Horsten and Leigh draw this distinction from the locus classicus Burge (2003).

notions, are they entitled to the dual reflection principle as well? The answer seems to be yes. Being justified in accepting a theory encompasses that theory being true and hence consistent, and then an understanding of the provability predicate and the logical constants is sufficient to comprehend the dual reflection principle. So an entitlement to orthodox reflection is sufficient to be entitled to dual reflection.

The above considerations show that the support given to orthodox reflection can be extended to dual reflection. Beyond justification, one might wonder what reasons there are to study these dual reflection principles. The basic inferences licensed by the dual reflection principle are drastically different. For instance, orthodox reflection admits an inference to truth, which is an expression of soundness; dual reflection expresses the consistency of a statement with respect to a set of axioms – which at face value is a notion weaker than truth.[11] Consider also a non-standard interpretation of truth as arithmetic provability *à la* Solovay (1976). On this view, the standard reflection principle reduces to a tautology, but the dual reflection principle becomes an explicit consistency statement for the system. These are not trivial differences. But their distinction from traditional ones is precisely why dual reflection principles are provocative.

One potential criticism of dual reflection principles is that they are, on their own, relatively inert. That is, adjoining dual reflection to a theory of arithmetic which does not already feature a truth theory will not have any substantial effect. It's only in conjunction with an existing set of truth axioms when dual reflection impacts the resulting theory. However, this is a weakness which is shared with standard global reflection principles. In essence, they are unhelpful in building a theory of truth. Instead, it is their interaction with other truth principles which is of interest. And as discussed above, there's a sense in which the dual reflection principle is at work in the case of uniform reflection, since it's equivalent to the standard one in such a context. If one were motivated by their use of standard reflection

[11]For reasons of scope, we do not pursue an analysis of proof-theoretic strength in this chapter. It is an open and interesting technical question to see if the two principles come apart on this matter.

to want to employ global reflection to render explicit certain commitments in their theory, then one ought to be so motivated to do the same with *dual* global reflection. But as we'll see, this sometimes reveals certain defects latent in the base theory.

4.3 Formal Development and Basic Meta-theorems

In what follows, we provide a number of fundamental definitions which will allow us to formulate the axiomatic theories and reflection principles to be examined. Additionally, we'll prove a few general propositions which will assist in demonstrating the consistency of those theories.

Definition 4.1. *The language of truth $L[T]$ is obtained by adding a privileged monadic predicate symbol T to the signature L of first-order PA.*

We will often examine extensions of a particular $L[T]$-theory which we will refer to as $PA[T]$.

Definition 4.2. *$PA[T]$ is the $L[T]$-theory obtained through extending PA by allowing instances of the truth predicate T in the induction schema.*

We now provide some terminology based on the well-known Kripke construction. In what follows, we will use $\mathbb{N}$ to denote the standard model of first-order Peano arithmetic in the signature L of pure arithmetic. We will frequently consider extensions of this model to the language of truth $L[T]$, which we denote $\mathbb{N}[\mathbb{T}]$, where the parameter $\mathbb{T}$ is a collection of Gödel codes of formulas which provide the extension of the truth predicate. In our notation, we distinguish $\mathbb{T}$, a set of numbers in the metalanguage, with T, the unary predicate for truth in the object language.

The theorems in this section make reference to a special class of such models.

Definition 4.3. *A jump model* $\mathbb{N}[\mathbb{T}]$ *is a structure with*

$\mathbb{T} = \{\langle\varphi\rangle \,|\, \mathbb{N}[\mathbb{T}'] \vDash \varphi\}$ *for some* $\mathbb{T}' \subset \omega$.

Here, $\langle\varphi\rangle$ denotes the operation which outputs the Gödel number of the formula φ. Again, $\mathbb{T}$ is a set of numbers.

Definition 4.4. *An* $L[T]$*-theory* S *is sound for jump models if* S *is true on all jump models, i.e.* $\mathbb{N}[\mathbb{T}] \vDash \varphi$ *for all* $\varphi \in S$ *when* $\mathbb{N}[\mathbb{T}]$ *is a jump model.*

The class of theories which are sound for jump models forms an ideal: it is closed under unions and closed under subset.

Definition 4.5. TA *(True Arithmetic) is the set of formulas* φ *of PA with* $\mathbb{N} \vDash \varphi$*. Let* $\langle TA\rangle$ *denote the set of Gödel numbers of the formulas of* TA.

Observe that since the standard model $\mathbb{N}$ satisfies TA, it follows that TA is sound for jump models.

We will begin by introducing a first-order global reflection principle defined for $L[T]$-theories.

Definition 4.6. *If* S *is an* $L[T]$*-theory, let G-Refl*^1_S *be the inference rule expressing that from* φ *relative to* S *derive* $\forall\psi(\mathrm{Bwb}_{S+\varphi}(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner))$*. Here* $\mathrm{Bwb}_{S+\varphi}$ *denotes ordinary provability in the theory* $S \cup \{\varphi\}$.

Note that the term 'inference rule' is being used in a more expansive sense than usual, since unlike the typical rules of inference one encounters, this one is specified relative to a base theory or set of hypotheses.

Here is now the formal implementation of our new reflection principle which we introduced informally in the previous section. Recall that the intuitive idea was that for any system we accept, truths should not conflict with what the system derives. The formulation provided below has not been studied in axiomatic theories of truth. A dim analogue in a purely arithmetical setting can be found in Hájek and Pudlák (1998, Corollary 4.34, p. 108); another arithmetical version connecting partial truth and consistency can be found in Kossak and Schmerl (2006, Mostowski's Reflection Principle, p. 19).

Definition 4.7. *If S is an $L[T]$-theory, let G-$\bar{R}$efl^1_S be the inference rule expressing that from φ relative to S derive $\forall\psi(T(\ulcorner\psi\urcorner) \to \neg\mathrm{Bwb}_{S+\varphi}(\ulcorner\neg\psi\urcorner))$.*

The $L[T]$-theories we study will be obtained by adding G-Refl^1_S to a base theory and taking the closure of that theory in the expanded deductive system.

Definition 4.8. *A grefl-proof of φ_l in an $L[T]$-theory S is a finite sequence $\pi = \langle\varphi_1, \ldots, \varphi_l\rangle$ of $L[T]$ formulas such that for each φ_i, we have one of the following:*

1. *φ_i is a tautology or $\varphi_i \in S$.*
2. *There are $j, k < i$ with $\varphi_k \equiv \varphi_j \to \varphi_i$.*
3. *There is $j < i$ with $\varphi_i \equiv \forall\psi(\mathrm{Bwb}_{S+\varphi_j}(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner))$, i.e. φ_i was obtained via G-Refl^1_S.*

For an $L[T]$-theory S, we define its grefl-deductive-closure SR, where a formula φ is in SR if and only if there is a grefl-proof of φ from S.

Again, in this definition, $\mathrm{Bwb}_{S+\varphi_j}$ denotes ordinary provability in the theory $S \cup \{\varphi_j\}$, which should be distinguished from provability in the expanded deductive system which we have just defined.

Likewise we will also examine theories obtained via G-$\bar{R}$efl^1_S.

Definition 4.9. *A g$\bar{r}$efl-proof of φ_l in an $L[T]$-theory S is a finite sequence $\pi = \langle\varphi_1, \ldots, \varphi_l\rangle$ of $L[T]$ formulas such that for each φ_i, we have one of the following:*

1. *φ_i is a tautology or $\varphi_i \in S$.*

2. *There are $j, k < i$ with $\varphi_k \equiv \varphi_j \to \varphi_i$.*

3. *There is $j < i$ with $\varphi_i \equiv \forall \psi (T(\ulcorner \psi \urcorner) \to \neg \mathrm{Bwb}_{S+\varphi_j}(\ulcorner \neg \psi \urcorner))$, i.e. φ_i was obtained via $G\text{-}\bar{R}efl^1_S$.*

As before, for an $L[T]$-theory S, we define its g$\bar{\mathrm{r}}$efl-deductive-closure by $\mathsf{S\bar{R}}$. Throughout this chapter, the overline notation $\bar{\mathrm{r}}/\bar{\mathrm{R}}/\bar{\mathsf{R}}$ will be used to distinguish usages of the new dual reflection principle with the standard ones.

To avoid confusing our two notions of provability, viz. classical and as extended by our reflection principles, let us now introduce an explicit predicate for this new provability notion.

Definition 4.10. *Let $\mathrm{Bwb}^{\mathsf{R}}_S(\ulcorner \psi \urcorner)$ be the arithmetized version of Definition 4.8.*

In Definition 4.8, we work with only finite sequences of formulas and other arithmetizable notions, so the above definition is coherent. Hence, for an $L[T]$-theory S, we have that $\varphi \in \mathsf{SR}$ if and only if $\mathbb{N} \vDash \mathrm{Bwb}^{\mathsf{R}}_S(\ulcorner \varphi \urcorner)$. We could introduce similar notation for our dual reflection principle, we refrain from doing so since such a notion turns out not to be operative in our proofs.

We can now prove a fundamental theorem for establishing the consistency of such theories.

Theorem 4.11. *Let S be a recursively enumerable $L[T]$-theory extending $PA[T]$ which is sound for jump models and let $\mathbb{T}_0 \subset \omega$ with*

$$\mathbb{T}_{n+1} = \{\langle \varphi \rangle \mid \varphi \in L[T], \mathbb{N}[\mathbb{T}_n] \vDash \varphi\} \tag{4.1}$$

Then if $l \geq 1$ with $\pi = \langle \varphi_1, \ldots, \varphi_l \rangle$ an grefl-proof of φ_l in S, then

$$0 \leq n,\ 1 \leq i \leq l \quad \Rightarrow \quad \mathbb{N}[\mathbb{T}_{l+n}] \vDash \varphi_i \tag{4.2}$$

Likewise, if $l \geq 1$ with $\pi = \langle \varphi_1, \dots, \varphi_l \rangle$ an $\mathrm{g\bar{r}efl}$-proof of φ_l in S, then

$$0 \leq n,\ 1 \leq i \leq l \quad \Rightarrow \quad \mathbb{N}[\mathbb{T}_{l+n}] \vDash \varphi_i \tag{4.3}$$

Proof. We first establish the result for grefl-proofs, then note the differences for g$\bar{r}$efl-proofs.

We proceed by induction on l. For the base case $l = 1$. For a grefl-proof of length 1, it must be that $\varphi_i = \varphi_l$ is a tautology or $\varphi_l \in S$. If φ_l is a tautology, then $\mathbb{N}[\mathbb{T}_{l+n}] \vDash \varphi_l$ automatically. If $\varphi_l \in S$, then $\mathbb{N}[\mathbb{T}_{l+n}] \vDash \varphi_l$ by hypothesis.

For the inductive step, suppose the proposition holds for l. Let $\pi = \langle \varphi_1, \dots, \varphi_l, \varphi_{l+1} \rangle$ be a grefl-proof of φ_{l+1} in S. There are three cases to consider, corresponding to how φ_{l+1} follows in π. Note that the induction hypothesis will take care of things when φ_i for $i \leq l$, so it suffices to examine just φ_{l+1}.

In the first case, φ_{l+1} is a tautology or $\varphi_{l+1} \in S$. This proceeds just as in the base case.

In the second case, φ_{l+1} is obtained by modus ponens, i.e. there are $j, k < l+1$ such that $\varphi_k \equiv \varphi_j \to \varphi_{l+1}$. Applying the induction hypothesis to these indexes, we obtain $\mathbb{N}[\mathbb{T}_{j+m_j}] \vDash \varphi_j$ and $\mathbb{N}[\mathbb{T}_{k+m_k}] \vDash \varphi_k$, where $m_j = l + 1 - j + n$ and $m_k = l + 1 - k + n$. It follows that $\mathbb{N}[\mathbb{T}_{l+1+n}] \vDash \varphi_j \wedge \varphi_k$, so given the definition of φ_k we have $\mathbb{N}[\mathbb{T}_{l+1+n}] \vDash \varphi_{l+1}$.

In the third case, φ_{l+1} is obtained by reflection introduction, i.e. there is $j < l+1$ such that $\varphi_{l+1} \equiv \forall \psi (\mathrm{Bwb}_{S+\varphi_j}(\ulcorner \psi \urcorner) \to T(\ulcorner \psi \urcorner))$. Let ψ be given such that $\mathbb{N}[\mathbb{T}_{l+1+n}] \vDash \mathrm{Bwb}_{S+\varphi_j}(\ulcorner \psi \urcorner)$. A witness to $\mathrm{Bwb}_{S+\varphi_j}(\ulcorner \psi \urcorner)$ encodes a proof in $S + \varphi_j$ of ψ, so we can conclude that $S + \varphi_j \vdash \psi$. By our original hypothesis, we know that $\mathbb{N}[\mathbb{T}_{l+n}] \vDash S$, and from the induction hypothesis we know that $\mathbb{N}[\mathbb{T}_{l+n}] \vDash \varphi_j$, so it must be that $\mathbb{N}[\mathbb{T}_{l+n}] \vDash \psi$, which happens just in case $\mathbb{N}[\mathbb{T}_{l+1+n}] \vDash T(\ulcorner \psi \urcorner)$. Therefore $\mathbb{N}[\mathbb{T}_{l+1+n}] \vDash \forall \psi (\mathrm{Bwb}_{S+\varphi_j}(\ulcorner \psi \urcorner) \to T(\ulcorner \psi \urcorner))$, as desired.

This completes the proof for grefl-proofs.

For g$\bar{r}$efl-proofs, the base case and first two subcases of the inductive step proceed identically, so we focus on the final subcase of the inductive step. So suppose there is $j < l+1$ such that $\varphi_{l+1} \equiv \forall\psi(T(\ulcorner\psi\urcorner) \to \neg\mathrm{Bwb}_{S+\varphi_j}(\ulcorner\neg\psi\urcorner))$. Let ψ be given. We move to demonstrate the contrapositive, so suppose $\mathbb{N}[\mathbb{T}_{l+n+1}] \vDash \mathrm{Bwb}_{S+\varphi_j}(\ulcorner\neg\psi\urcorner)$. It follows that $S + \varphi_j \vdash \neg\psi$. By assumption $\mathbb{N}[\mathbb{T}_{l+n}] \vDash S$ and by the induction hypothesis $\mathbb{N}[\mathbb{T}_{l+n}] \vDash \varphi_j$, so it must be that $\mathbb{N}[\mathbb{T}_{l+n}] \vDash \neg\psi$, i.e. that $\mathbb{N}[\mathbb{T}_{l+n}] \nvDash \psi$. By equation (4.1), it follows that $\langle\psi\rangle \notin \mathbb{T}_{l+n+1}$, and so $\mathbb{N}[\mathbb{T}_{l+1+n}] \vDash \neg T(\ulcorner\psi\urcorner)$. Hence by universal generalization $\mathbb{N}[\mathbb{T}_{l+1+n}] \vDash \forall\psi(T(\ulcorner\psi\urcorner) \to \neg\mathrm{Bwb}_{S+\varphi_j}(\ulcorner\neg\psi\urcorner))$, as desired. □

We can employ the above proposition to construct models of theories satisfying the hypotheses of Theorem 4.11. More specifically, let U be an ultrafilter extending the Fréchet filter on ω. By Łoś's Theorem (Marker, 2002, p. 64), we have that the ultraproduct $\Pi_U\mathbb{N}[\mathbb{T}_n]$ is a model of S. By Theorem 4.11, we know that every grefl-provable formula is satisfied by all later models in the construction after finitely-many steps. The collection of those models as indexed by $\mathbb{T}_n$ will be cofinite and so the ultraproduct will also satisfy the formula.[12] This same technique works completely analogously for g$\bar{r}$efl-proofs and so may be employed to the same end.

Before moving onto our first system, we note a few helpful results.

Corollary 4.12. *Fix an $L[T]$-theory S and recall that* SR *denotes its grefl-deductive-closure and* $\mathsf{S\bar{R}}$ *its g$\bar{r}$efl-deductive-closure. If S satisfies the hypotheses of Theorem 4.11, then both* $\mathsf{SR} \cup TA$ *and* $\mathsf{S\bar{R}} \cup TA$ *are consistent.*

Proof. As remarked above, TA is sound for jump models, and the union of two theories which are sound for jump models is itself sound for jump models, so in particular $\mathsf{SR} \cup TA$ and $\mathsf{S\bar{R}} \cup TA$ are sound for jump models. □

[12]The model generated by this process will be uncountable (since the Fréchet filter on ω is uncountable) and hence non-standard.

The next proposition isolates a sufficient condition for the two reflection principles to derive identical theories.

Proposition 4.13. *If S proves the commutation of negation, i.e. $S \vdash T(\ulcorner \neg\varphi \urcorner) \leftrightarrow \neg T(\ulcorner \varphi \urcorner)$ for all φ in $L[T]$, then* SR *and* $\mathsf{S\bar{R}}$ *are equivalent.*

Proof. Observe that the associated reflection principles are interderivable:

$$
\begin{aligned}
&\forall\psi(\mathrm{Bwb}_{S+\varphi}(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner)) && \\
&\leftrightarrow \forall\psi(\neg T(\ulcorner\psi\urcorner) \to \neg\mathrm{Bwb}_{S+\varphi}(\ulcorner\psi\urcorner)) && \\
&\leftrightarrow \forall\psi(T(\ulcorner\neg\psi\urcorner) \to \neg\mathrm{Bwb}_{S+\varphi}(\ulcorner\psi\urcorner)), && \text{by hypothesis} \\
&\leftrightarrow \forall\theta(T(\ulcorner\theta\urcorner) \to \neg\mathrm{Bwb}_{S+\varphi}(\ulcorner\neg\theta\urcorner)), && \theta \equiv \neg\psi
\end{aligned}
$$

Since the resulting theories are just the closure under their respective reflection principles, they are therefore equivalent theories. □

We conclude this section by noting some properties of theories obtained from grefl-deductive-closure. These will bear on the consistency and inconsistency results in the following sections.

As a note, we use the terminology External Ascent to denote the property of a theory S where if $S \vdash \varphi$ then $S \vdash T(\ulcorner\varphi\urcorner)$. There is no standard terminology here, but some authors, e.g. Friedman and Sheard (1987, p. 5), prefer the term 'T-Intro'.

Theorem 4.14. *Let* SR *be an $L[T]$-theory obtained by grefl-deductive-closure from an $L[T]$-theory S extending $PA[T]$. Then* SR *admits External Ascent, i.e. if* $\mathsf{SR} \vdash \varphi$ *then* $\mathsf{SR} \vdash T(\ulcorner\varphi\urcorner)$.

Moreover, PA proves that $\forall\psi(\mathrm{Bwb}^{\mathsf{R}}_{\mathsf{S}}(\ulcorner\psi\urcorner) \to \mathrm{Bwb}^{\mathsf{R}}_{\mathsf{S}}(\ulcorner T(\ulcorner\psi\urcorner)\urcorner))$.

Proof. Suppose $\mathsf{SR} \vdash \varphi$. By G-Refl$^1_S$, we have that

$$\mathsf{SR} \vdash \forall\psi(\mathrm{Bwb}_{S+\varphi}(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner)) \tag{4.4}$$

Instantiating (4.4) by φ, we have that

$$\mathsf{SR} \vdash \mathrm{Bwb}_{S+\varphi}(\ulcorner\varphi\urcorner) \to T(\ulcorner\varphi\urcorner) \tag{4.5}$$

$\mathsf{SR} \vdash \mathrm{Bwb}_{S+\varphi}(\ulcorner\varphi\urcorner)$ since $S+\varphi \vdash \varphi$ trivially, so by modus ponens we obtain $\mathsf{SR} \vdash T(\ulcorner\varphi\urcorner)$. This proof can be carried out in PA, so the latter half of the theorem follows immediately. □

We note a quick corollary connecting ω-models and consistency with global reflection.

Corollary 4.15. *Let S be an $L[T]$-theory and let SR be the grefl-deductive-closure of S. If SR has an ω-model, then the theory*
$\mathsf{SR}+ = \mathsf{SR} \cup \{\forall\psi(\mathrm{Bwb}^{\mathsf{R}}_S(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner))\}$ *is consistent.*

Proof. By hypothesis, there is some $\mathbb{T}$ such that $\mathbb{N}[\mathbb{T}] \vDash \mathsf{SR}$, so it remains to show that $\mathbb{N}[\mathbb{T}] \vDash \forall\psi(\mathrm{Bwb}^{\mathsf{R}}_S(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner))$.

Let ψ be an $L[T]$-formula and suppose $\mathbb{N}[\mathbb{T}] \vDash \mathrm{Bwb}^{\mathsf{R}}_S(\ulcorner\psi\urcorner)$. It follows that $\mathsf{SR} \vdash \psi$. By Theorem 4.14, we have $\mathsf{SR} \vdash T(\ulcorner\psi\urcorner)$ and hence $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\psi\urcorner)$. □

We alert the reader that we have found no obviously analogous results similar to Theorem 4.14 and Corollary 4.15 for g$\bar{r}$efl-deductive-closures.

4.4 Friedman-Sheard Theory

Here we follow the original article of Friedman and Sheard (1987) in recalling their axioms governing the composition of the truth predicate[13,14]:

(FS1) $\forall\varphi \in L[\Sigma^0_0] \quad T(\ulcorner\varphi\urcorner) \leftrightarrow \mathrm{Sat}^0_0(\ulcorner\varphi\urcorner)$

(FS2) $\forall\varphi \in L[T] \quad T(\ulcorner\neg\varphi\urcorner) \leftrightarrow \neg T(\ulcorner\varphi\urcorner)$

(FS3) $\forall\varphi,\psi \in L[T] \quad T(\ulcorner\varphi\wedge\psi\urcorner) \leftrightarrow (T(\ulcorner\varphi\urcorner)\wedge T(\ulcorner\psi\urcorner))$

(FS4) $\forall\varphi(x) \in L[T] \quad T(\ulcorner\forall x\varphi(x)\urcorner) \leftrightarrow \forall x T(\ulcorner\varphi(\underline{x})\urcorner)$

Definition 4.16. *The Friedman-Sheard Theory with Reflection (FSR) is the grefl-deductive closure of $PA[T]$ with FS1-FS4.*

Halbach (2011, §14.4, pp. 188-192) examines a similar system obtained by enriching FS with reflection principles. He considers a sequence of theories starting with the base Friedman-Sheard axioms where the next theory is obtained by extending the previous theory by its own global reflection principle. He terms the union of these theories FSR, and it is a subtheory of our FSR.[15] Since our interests are the ω-level theories and not their finite iterates, our method for adding reflection proves to be more elegant. More importantly, in the case of our dual reflection principle, it is not apparent that the Halbach technique can be deployed appropriately.

[13] Friedman and Sheard considered a number of axiomatic theories arising from various combinations of axioms. The theory recounted here has emerged as the popular candidate, receiving extensive study in the literature. See Chapter 14 of Halbach (2011, p. 149) for one such analysis.

[14] The first axiom expresses satisfaction in the ambient model for sentences with only bounded quantifiers.

[15] It is open whether this theory is a proper extension of Halbach's. The results proved in this section serve the dual role of illustrating the novel definitions and constructions of this chapter.

4.4.1 The Consistency of FSR

We begin be establishing the consistency of FSR by providing a method for constructing models of the theory.

Proposition 4.17. *FSR satisfies the conditions of Theorem 4.11, hence FSR is consistent.*

Proof. It suffices to establish that $S = \{FS1 - FS4\} \cup PA[T]$ (the base theory for FSR) is sound for jump models.

Let $\mathbb{T} = \{\langle\varphi\rangle \,|\, \mathbb{N}[\mathbb{T}'] \vDash \varphi\}$ for some $\mathbb{T}' \subset \omega$. If $\varphi \in PA[T]$, then $\mathbb{N}[\mathbb{T}] \vDash \varphi$ since $\mathbb{N}[\mathbb{T}]$ is based on the standard model. If $\varphi \in$ FS1-4, observe that the definition $\mathbb{T}$ tells us that the extension of T in $\mathbb{N}[\mathbb{T}]$ is precisely (the Gödel numbers of) the sentences which $\mathbb{N}[\mathbb{T}']$ satisfies, so $\mathbb{N}[\mathbb{T}]$ will satisfy FS1-FS4:

FS1: $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\varphi\urcorner)$ if and only if $\mathbb{N}[\mathbb{T}'] \vDash \varphi$, which holds because in this axiom φ is an arithmetic formula and $\mathbb{N}$ is the standard model.

FS2: $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\neg\varphi\urcorner)$ if and only if $\mathbb{N}[\mathbb{T}'] \vDash \neg\varphi$, if and only if $\mathbb{N}[\mathbb{T}'] \nvDash \varphi$, if and only if $\mathbb{N}[\mathbb{T}] \nvDash T(\ulcorner\varphi\urcorner)$, if and only if $\mathbb{N}[\mathbb{T}] \vDash \neg T(\ulcorner\varphi\urcorner)$.

FS3: $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\varphi \wedge \psi\urcorner)$ if and only if $\mathbb{N}[\mathbb{T}'] \vDash \varphi \wedge \psi$, if and only if $\mathbb{N}[\mathbb{T}'] \vDash \varphi$ and $\mathbb{N}[\mathbb{T}'] \vDash \psi$, if and only if $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\varphi\urcorner)$ and $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\psi\urcorner)$, if and only if $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\varphi\urcorner) \wedge T(\ulcorner\psi\urcorner)$.

FS4: $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\forall x\varphi(x)\urcorner)$ if and only if $\mathbb{N}[\mathbb{T}'] \vDash \forall x\varphi(x)$, if and only if for all $n \in \omega$ $\mathbb{N}[\mathbb{T}'] \vDash \varphi(\bar{n})$, if and only if for all $n \in \omega$ $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\varphi(\bar{n})\urcorner)$, if and only if $\mathbb{N}[\mathbb{T}] \vDash \forall x T(\ulcorner\varphi(\underline{x})\urcorner)$. □

4.4.2 Global Reflection in FSR

To begin, we note that the proof of McGee (1985, p. 399), showing that FS is ω-inconsistent, extends straightforwardly to FSR. For the sake of brevity, we do not repeat the proof here.

Proposition 4.18. *For all* $\mathbb{T} \subset \omega$, $\mathbb{N}[\mathbb{T}] \nvDash \mathit{FSR}$.

We conclude by establishing that FSR is inconsistent with its own global reflection principle.

Proposition 4.19. $\mathsf{FSR+} = \mathsf{FSR} \cup \{\forall\psi(\mathrm{Bwb}^{\mathsf{R}}_{\mathsf{S}}(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner))\}$*, where* $S = \{FS1 - FS4\} \cup PA[T]$*, is inconsistent.*

Proof. Define $T^n(\ulcorner\varphi\urcorner)$ by $T^1(\ulcorner\varphi\urcorner) \equiv T(\ulcorner\varphi\urcorner)$ and $T^{n+1}(\ulcorner\varphi\urcorner) \equiv$
$T(\ulcorner T^n(\ulcorner\varphi\urcorner)\urcorner)$. By the Diagonal Lemma, there is an $L[T]$-sentence ξ such that

$$\mathsf{FSR} \vdash \xi \leftrightarrow \exists x \geq 1(\neg T^x(\ulcorner\xi\urcorner)) \tag{4.6}$$

By McGee's theorem, we have that $\mathsf{FSR} \vdash \xi$. Moving from the metatheory to PA, we have

$$PA \vdash \mathrm{Bwb}^{\mathsf{R}}_{S}(\ulcorner\xi\urcorner) \tag{4.7}$$

By Theorem 4.14, we have

$$PA \vdash \forall\psi(\mathrm{Bwb}^{\mathsf{R}}_{\mathsf{S}}(\ulcorner\psi\urcorner) \to \mathrm{Bwb}^{\mathsf{R}}_{\mathsf{S}}(\ulcorner T(\ulcorner\psi\urcorner)\urcorner)) \tag{4.8}$$

Reasoning inside PA, we can apply induction on (4.7) and (4.8) to obtain

$$PA \vdash \forall x \geq 1(\mathrm{Bwb}^{\mathsf{R}}_{S}(\ulcorner T^x(\ulcorner\xi\urcorner)\urcorner)) \tag{4.9}$$

Since $FSR+$ extends PA, we also have

$$\mathsf{FSR+} \vdash \forall x \geq 1(\mathrm{Bwb}^{\mathsf{R}}_{S}(\ulcorner T^x(\ulcorner\xi\urcorner)\urcorner)) \tag{4.10}$$

But then we apply global reflection to (4.10) to obtain

$$\mathsf{FSR+} \vdash \forall x \geq 1(T^{x+1}(\ulcorner \xi \urcorner)) \tag{4.11}$$

By Theorem 4.14 we know that $\mathsf{FSR} \vdash T(\ulcorner \xi \urcorner)$, and so $\mathsf{FSR+} \vdash T(\ulcorner \xi \urcorner)$ since $\mathsf{FSR+}$ extends FSR. But this combined with (4.11) yields $\mathsf{FSR+} \vdash \neg\xi$, a contradiction. □

When working with his formulation, Halbach derives an analogue of Proposition 4.19, observing that "adding the uniform reflection principle for a system S to an ω-inconsistent system S yields an outright inconsistency, if certain natural conditions are met" (Halbach, 2011, p. 192). In this context, we worked with global reflection principles rather than uniform ones, but the net effect was just that the proof went through the truth predicate.

4.4.3 The Status of FSR-bar

Despite featuring different reflection principles, we find that FSR and $\mathsf{FS\bar{R}}$ are equivalent theories. This is a simple application of Proposition 4.13, since the requisite schema is just FS2. In other words, since the reflection principles are the only differences between the theories, we have $\mathsf{FSR} = \mathsf{FS\bar{R}}$ as any grefl-proof in $\{FS1-4\} \cup PA$ will have a corresponding g$\bar{r}$efl-proof and vice-versa.

4.5 Kripke-Feferman Theory

We now turn our attention to another popular theory of truth. We will examine how the base theory can be consistently extended by reflection, and then consider how supplementing these axioms impacts this consistency.

Recall the Kripke-Feferman axiom schemas from Feferman (1991)[16]:

(KF1) $T(\ulcorner\varphi\urcorner) \leftrightarrow \mathrm{Sat}_0^0(\ulcorner\varphi\urcorner)$

(KF2) $T(\ulcorner\neg\varphi\urcorner) \leftrightarrow \mathrm{Sat}_0^0(\ulcorner\neg\varphi\urcorner)$

(KF3) $T(\ulcorner\neg\neg\varphi\urcorner) \leftrightarrow T(\ulcorner\varphi\urcorner)$

(KF4) $T(\ulcorner\varphi \wedge \psi\urcorner) \leftrightarrow (T(\ulcorner\varphi\urcorner) \wedge T(\ulcorner\psi\urcorner))$

(KF5) $T(\ulcorner\neg(\varphi \wedge \psi)\urcorner) \leftrightarrow (T(\ulcorner\neg\psi\urcorner) \vee T(\ulcorner\neg\varphi\urcorner))$

(KF6) $T(\ulcorner\varphi \vee \psi\urcorner) \leftrightarrow (T(\ulcorner\varphi\urcorner) \vee T(\ulcorner\psi\urcorner))$

(KF7) $T(\ulcorner\neg(\varphi \vee \psi)\urcorner) \leftrightarrow (T(\ulcorner\neg\psi\urcorner) \wedge T(\ulcorner\neg\varphi\urcorner))$

(KF8) $T(\ulcorner\forall x\theta(x)\urcorner) \leftrightarrow \forall x(T(\ulcorner\theta(x)\urcorner))$

(KF9) $T(\ulcorner\neg\forall x\theta(x)\urcorner) \leftrightarrow \exists x(T(\ulcorner\neg\theta(x)\urcorner))$

(KF10) $T(\ulcorner\exists x\theta(x)\urcorner) \leftrightarrow \exists x(T(\ulcorner\theta(x)\urcorner))$

(KF11) $T(\ulcorner\neg\exists x\theta(x)\urcorner) \leftrightarrow \forall x(T(\ulcorner\neg\theta(x)\urcorner))$

(KF12) $T(\ulcorner T(\ulcorner\varphi\urcorner)\urcorner) \leftrightarrow T(\ulcorner\varphi\urcorner)$

(KF13) $T(\ulcorner\neg T(\ulcorner\varphi\urcorner)\urcorner) \leftrightarrow T(\ulcorner\neg\varphi\urcorner)$

Definition 4.20. *The Kripke-Feferman Theory (KF) is the $L[T]$-theory obtained by adding $KF1-13$ to $PA[T]$.*

KFR and $\mathsf{KF\bar{R}}$ will denote the deductive closures of KF under our two types of reflection.

[16]These are given an alternative presentation in Halbach (2011, §15, pp. 195-227).

4.5.1 The Consistency of KFR

In the strong-Kleene logic, there are three truth values, so an $L[T]$-structure $\mathbb{N}[\mathbb{T},\mathbb{F}]$ can be described by two parameters which specify both the extension of the truth predicate $\mathbb{T}$ and its anti-extension $\mathbb{F}$. As is familiar from the Kripke construction, one defines a sequence of structures $\mathbb{N}[\mathbb{T}_\alpha,\mathbb{F}_\alpha]$ by setting the extension and anti-extension to be empty at stage zero and then taking jumps à la Definition 4.3 at successor stages and taking unions at limit stages.

Let $\mathbb{N}[\mathbb{T}_l,\mathbb{F}_l]$ be a fixed point of a strong-Kleene structure satisfying the semantic counterpart of the T-schema, where '$\mathrm{val}_{\mathbb{N}[\mathbb{T}_l,\mathbb{F}_l]}$' gives the truth-value of a formula at stage l of the construction:

$$\mathrm{val}_{\mathbb{N}[\mathbb{T}_l,\mathbb{F}_l]}(T(\ulcorner\varphi\urcorner)) \in \{0,1\} \Rightarrow \mathrm{val}_{\mathbb{N}[\mathbb{T}_l,\mathbb{F}_l]}(T(\ulcorner\varphi\urcorner)) = \mathrm{val}_{\mathbb{N}[\mathbb{T}_l,\mathbb{F}_l]}(\varphi) \tag{4.12}$$

In what follows, we use $\cdot^c$ to denote the relative complement operation in set theory. In particular, we'll apply it to the anti-extension of truth, so that $(\mathbb{F}_l)^c = \omega \setminus \mathbb{F}_l$.

Lemma 4.21. *For any $L[T]$-sentence φ, $\mathbb{N}[(\mathbb{F}_l)^c] \vDash \varphi \to T(\ulcorner\varphi\urcorner)$.*

Proof. We will show that it satisfies the contrapositive, so suppose that $\mathbb{N}[(\mathbb{F}_l)^c] \vDash \neg T(\ulcorner\varphi\urcorner)$. In the Kripke semantics it follows that $\langle\varphi\rangle \in \mathbb{F}_l$. Pick some earlier stage $\alpha < l$ in the Kripke construction such that $\langle\varphi\rangle \in \mathbb{F}_\alpha \subset \mathbb{F}_{\alpha+1}$. In the strong-Kleene structure this means $\mathrm{val}_{\mathbb{N}[\mathbb{T}_\alpha,\mathbb{F}_\alpha]}(\varphi) = 0$. Since the extension and anti-extension of the truth predicate are monotonic, we have $\mathbb{T}_\alpha \subset (\mathbb{F}_l)^c$. It follows that $\mathrm{val}_{\mathbb{N}[(\mathbb{F}_l)^c,\mathbb{F}_l]}(\varphi) = 0$, and so $\mathbb{N}[(\mathbb{F}_l)^c] \vDash \neg\varphi$. $\square$

Proposition 4.22. *The Kripke structure $\mathbb{N}[(\mathbb{F}_l)^c]$ is a model of KFR, so KFR is consistent and in particular ω-consistent.*

Proof. This proof proceeds by induction on proof-length. It is well-known that $\mathbb{N}[(\mathbb{F}_l)^c]$ satisfies KF, so we will focus on the step of the induction for reflection and show that $\mathbb{N}[(\mathbb{F}_l)^c]$ satisfies G-Refl$^1_{\mathsf{KF}}$.

Suppose we are applying G-Refl$^1_{\mathsf{KF}}$ to φ occurring earlier in the proof. By our induction hypothesis we know $\mathbb{N}[(\mathbb{F}_l)^c] \vDash \varphi$. Fix an $L[T]$-sentence ψ and suppose that $\mathbb{N}[(\mathbb{F}_l)^c] \vDash \mathrm{Bwb}_{\mathsf{KF}+\varphi}(\ulcorner\psi\urcorner)$. By the supposition, we know $\mathsf{KF} \cup \{\varphi\} \vdash \psi$. Since $\mathbb{N}[(\mathbb{F}_l)^c] \vDash \mathsf{KF}$ it follows that $\mathbb{N}[(\mathbb{F}_l)^c] \vDash \psi$. By Lemma 4.21 we have that $\mathbb{N}[(\mathbb{F}_l)^c] \vDash T(\ulcorner\psi\urcorner)$ and so $\mathbb{N}[(\mathbb{F}_l)^c] \vDash \forall\psi(\mathrm{Bwb}_{\mathsf{KF}+\varphi}(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner))$ by universal generalization. $\square$

We also note that KFR is consistent with its own global reflection principle.

Proposition 4.23. $\mathsf{KFR+} = \mathsf{KFR} \cup \{\forall\psi(\mathrm{Bwb}^{\mathsf{R}}_{\mathsf{KF}}(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner))\}$ *is consistent.*

Proof. This follows immediately from Proposition 4.22 and Corollary 4.15. $\square$

4.5.2 Classes of models of KF and KFR

The theory KF provides an axiomatization of the strong-Kleene semantics in the Kripke construction. These axioms are in a sense ambivalent about the interpretation of the indeterminate truth-value. One categorization of the models of KF can be to divide them into those satisfying instances of $T(\ulcorner\varphi\urcorner) \to \varphi$ and those satisfying instances of $\varphi \to T(\ulcorner\varphi\urcorner)$ – 'T-out' and 'T-in' in the Friedman and Sheard (1987, p. 5) terminology. The former class corresponds to the models which exclude indeterminate formulas from the T predicate (e.g. $\mathbb{N}[\mathbb{T}_l]$), and the latter to the inclusive ones like $\mathbb{N}[(\mathbb{F}_l)^c]$.

Obviously these properties are incompatible (for threat of contradicting Tarski's undefinability theorem). They correspond to stances on paracompleteness and paraconsistency for the T predicate. We make this explicit in the following two propositions. The first can be found in Halbach (2011, Lemma 15.19, p. 214); the related second result appears to be folklore as well, but we include it for the sake of completeness.[17]

Proposition 4.24. *Over* KF*, the following schemas are equivalent:*

1. $T(\ulcorner\varphi\urcorner) \rightarrow \varphi$
2. $\neg(T(\ulcorner\varphi\urcorner) \wedge T(\ulcorner\neg\varphi\urcorner))$

Proposition 4.25. *Over* KF*, the following schemas are equivalent:*

1. $\varphi \rightarrow T(\ulcorner\varphi\urcorner)$
2. $T(\ulcorner\varphi\urcorner) \vee T(\ulcorner\neg\varphi\urcorner)$

On the other hand, Theorem 4.14 tells us that KFR does not exhibit this indifference.

Proposition 4.26. KFR *is inconsistent with the schema* $T(\ulcorner\varphi\urcorner) \rightarrow \varphi$*.*

Proof. Theorem 4.14 along with the schema yields the Paradox of the Knower, a strengthened form of Tarski's undefinability theorem due to Montague (1963). □

So adopting this approach to reflection requires us to take a stance on the properties of the T predicate: it cannot be consistent. This is somewhat problematic for an advocate of KF: a main attraction of the theory is its ambivalence (and in particular that it is silent about the liar sentence). KFR presents a conflict then by removing one of the appealing features of the theory. And perhaps disturbing is that one's theoretical commitments aren't known from their axioms alone, only becoming explicit after reflection.

[17] The former result is anticipated in Cantini (1989, Lemma 3.2, p. 102). The relations between the schemas are explored in a general setting in Friedman and Sheard (1987).

This is connected to the disparity between the inner and outer logics of KF.[18] The truth and falsity predicates are built up as partial but are featured as total predicates in the end model. The reflective process can be seen as enforcing some degree of parity by ruling out certain completions of the predicates. In the following section, we flesh out this idea by finding similar results to Proposition 4.26.

4.5.3 Incompatible Extensions of KF and the Consistency of KFR-bar

We saw in Lemmas 4.21 and 4.29 that T-in is independent of the KF axioms. In Proposition 4.25, we noted that the T-in schema is equivalent to the desirable property that the truth predicate is complete, i.e. that for every formula, it assigns truth to either it or its negation. This motivates the idea of adding the T-in schema to KF – call this base theory $(\mathsf{KF} + \text{T-in})$. We will examine the closures of this theory under our two reflection principles – call them $(\mathsf{KF} + \text{T-in})\mathsf{R}$ and $(\mathsf{KF} + \text{T-in})\bar{\mathsf{R}}$.

The featured ω-model of KFR also satisfies $(\mathsf{KF} + \text{T-in})\mathsf{R}$.

Proposition 4.27. $\mathbb{N}[(\mathbb{F}_l)^c] \vDash (\mathsf{KF} + \textit{T-in})\mathsf{R}$*, so* $(\mathsf{KF} + \textit{T-in})\mathsf{R}$ *is ω-consistent.*

Proof. This involves induction on length of grefl-proof from $(\mathsf{KF} + \text{T-in})$. But this proceeds exactly as in Proposition 4.22, except that the base case is slightly expanded: all that remains is to show that $\mathbb{N}[(\mathbb{F}_l)^c] \vDash$T-in, which was demonstrated in Lemma 4.21. □

However, it turns out that the same is not true of the dual theory $(\mathsf{KF} + \text{T-in})\bar{\mathsf{R}}$.

Proposition 4.28. $(\mathsf{KF} + \textit{T-in})\bar{\mathsf{R}}$ *is inconsistent.*

[18] See Reinhardt (1986) and Burge (1979) for discussion.

Proof. For reductio, suppose there were a model $\mathcal{M}$ such that $\mathcal{M} \vDash (\mathsf{KF} + \text{T-in})\bar{\mathsf{R}}$.

Let λ denote the usual liar sentence so that $\mathcal{M} \vDash \neg\lambda \leftrightarrow T(\ulcorner\lambda\urcorner)$. By T-in, $\mathcal{M} \vDash \lambda \to T(\ulcorner\lambda\urcorner)$. Combining the two we have $\mathcal{M} \vDash \lambda \to \neg\lambda$, so we have $\mathcal{M} \vDash \neg\lambda$, i.e. $\mathcal{M} \vDash T(\ulcorner\lambda\urcorner)$.

Let φ be a fixed tautology. By G$\bar{R}$efl, $\mathcal{M} \vDash T(\ulcorner\lambda\urcorner) \to \neg\text{Bwb}_{(\mathsf{KF}+\text{T-in})+\varphi}(\ulcorner\neg\lambda\urcorner)$. So $\mathcal{M} \vDash \neg\text{Bwb}_{(\mathsf{KF}+\text{T-in})+\varphi}(\neg\lambda)$. This is a Π^0_1 statement so the standard model satisfies $\mathbb{N} \vDash \neg\text{Bwb}_{(\mathsf{KF}+\text{T-in})+\varphi}(\neg\lambda)$, thus $(\mathsf{KF} + \text{T-in}) + \varphi \nvdash \neg\lambda$.[19] So there is another model $\mathcal{N}$ with $\mathcal{N} \vDash (\mathsf{KF} + \text{T-in}) + \varphi + \neg\neg\lambda$. But $\mathcal{N} \vDash$T-in, so $\mathcal{N} \vDash \neg\lambda$ by the reasoning in the previous paragraph, a contradiction.

□

In parity with 4.22, we demonstrate the consistency of $\mathsf{KF}\bar{\mathsf{R}}$. We accomplish this by consideration of the theory $(\mathsf{KF} + \text{T-out})\bar{\mathsf{R}}$ which extends it.

Lemma 4.29. *For any $L[T]$-sentence φ, $\mathbb{N}[\mathbb{T}_l] \vDash T(\ulcorner\varphi\urcorner) \to \varphi$.*

Proof. Suppose $\mathbb{N}[\mathbb{T}_l] \vDash T(\ulcorner\varphi\urcorner)$. It follows that $\langle\varphi\rangle \in \mathbb{T}_l$. Then there is some earlier stage $\alpha < l$ in the Kripke construction such that $\langle\varphi\rangle \in \mathbb{T}_\alpha$. It follows that $\text{val}_{\mathbb{N}[\mathbb{T}_\alpha,\mathbb{F}_\alpha]}(\varphi) = 1$. Since $\mathbb{T}_\alpha \subset \mathbb{T}_l$ and $\mathbb{F}_\alpha \subset (\mathbb{T}_l)^c$, we have that $\text{val}_{\mathbb{N}[\mathbb{T}_l,(\mathbb{T}_l)^c]}(\varphi) = 1$ and hence $\mathbb{N}[\mathbb{T}_l] \vDash \varphi$. □

Proposition 4.30. *The Kripke structure $\mathbb{N}[\mathbb{T}_l]$ is a model of $(\mathsf{KF}+\textit{T-out})\bar{\mathsf{R}}$, so $(\mathsf{KF}+\textit{T-out})\bar{\mathsf{R}}$ is consistent and in particular ω-consistent.*

Proof. As in Proposition 4.22, the proof proceeds by induction on proof length. It is well-known that $\mathbb{N}[\mathbb{T}_l]$ satisfies KF, and from Lemma 4.29 we know $\mathbb{N}[\mathbb{T}_l] \vDash$ T-out, so it remains to demonstrate that $\mathbb{N}[\mathbb{T}_l]$ satisfies G-$\bar{R}$efl$^1_{(\mathsf{KF}+\text{T-out})}$.

Suppose we are applying G-$\bar{R}$efl$^1_{(\mathsf{KF}+\text{T-out})}$ to φ occurring earlier in the proof. Fix an $L[T]$-sentence ψ. We will show that $\mathbb{N}[\mathbb{T}_l]$ satisfies the contrapositive, so further suppose that

[19] See Kaye (1991, p. 28, Corollary 2.8) for details.

$\mathbb{N}[\mathbb{T}_l] \vDash \mathrm{Bwb}_{(\mathsf{KF}+\text{T-out})+\varphi}(\ulcorner\neg\psi\urcorner)$ for some φ g$\bar{r}$efl-provable from $(\mathsf{KF} + \text{T-out})$. By the latter supposition, we know $(\mathsf{KF} + \text{T-out}) \cup \{\varphi\} \vdash \neg\psi$. Since $\mathbb{N}[\mathbb{T}_l] \vDash (\mathsf{KF} + \text{T-out})$ it follows that $\mathbb{N}[\mathbb{T}_l] \vDash \neg\psi$. Suppose toward a contradiction that $\mathbb{N}[\mathbb{T}_l] \vDash T(\ulcorner\psi\urcorner)$. By Lemma 4.29 we would have that $\mathbb{N}[\mathbb{T}_l] \vDash \psi$, a contradiction. So it must be that $\mathbb{N}[\mathbb{T}_l] \vDash \neg T(\ulcorner\psi\urcorner)$, as desired. Thus $\mathbb{N}[\mathbb{T}_l] \vDash \forall\psi(\mathrm{Bwb}_{(\mathsf{KF}+\text{T-out})+\varphi}(\ulcorner\neg\psi\urcorner) \to \neg T(\ulcorner\psi\urcorner))$ by universal generalization. □

Since $\mathsf{KF}\bar{\mathsf{R}}$ is a sub-theory of $(\mathsf{KF} + \text{T-out})\bar{\mathsf{R}}$, it follows that $\mathsf{KF}\bar{\mathsf{R}}$ is also ω-consistent.

Finally, we note a dual to Proposition 4.28.

Proposition 4.31. $(\mathsf{KF} + \textit{T-out})\mathsf{R}$ *is inconsistent.*

Proof. Let λ denote the liar sentence such that $(\mathsf{KF} + \text{T-out})\mathsf{R} \vdash \neg\lambda \leftrightarrow T(\ulcorner\lambda\urcorner)$. By T-out, $(\mathsf{KF}+\text{T-out})\mathsf{R} \vdash T(\ulcorner\lambda\urcorner) \to \lambda$, so $(\mathsf{KF}+\text{T-out})\mathsf{R} \vdash \neg\lambda \to \lambda$, i.e. $(\mathsf{KF}+\text{T-out})\mathsf{R} \vdash \lambda$. By Theorem 4.14, $(\mathsf{KF} + \text{T-out})\mathsf{R} \vdash T(\ulcorner\lambda\urcorner)$ and thus $(\mathsf{KF} + \text{T-out}) \vdash \neg\lambda$. □

The pattern of results for these extensions of the Kripke-Feferman theory reveal the entanglement of reflection with the semantic notion of truth. Here, taking a stance on reflection principles is tantamount to choosing (or excluding) one of the T-rules, which by the above propositions commits one to the para-completeness or para-consistency of the truth predicate.

4.6 Typed Tarski Biconditionals

We can't have an exhaustive disquotation schema without contradicting Tarski's undefinability theorem. But any theory of truth should at least handle (dis-)quotation for the underlying T-free theory, i.e. Peano Arithmetic.

We begin by noting some properties of a minimal theory of Tarski biconditionals and then examine the extension of this theory by reflection principles with a focus on consistency.

Definition 4.32. *The system* TB_0 *is the* $L[T]$*-theory extending* $PA[T]$ *obtained by adding Tarski biconditionals for every sentence of* PA*:*

$$T(\ulcorner\varphi\urcorner) \leftrightarrow \varphi : \quad \varphi \in \mathcal{L}_{PA} \textit{ a closed formula} \tag{4.13}$$

To help distinguish sets of formulas and sets of Gödel numbers, we adopt the following notation: if $\mathbb{S} \subset \omega$ is a set Gödel numbers, let $\mathring{\mathbb{S}}$ denote the set of corresponding formulas, i.e. $\mathbb{S} = \{\langle\varphi\rangle \mid \varphi \in \mathring{\mathbb{S}}\}$.

Lemma 4.33. $\mathbb{N}[\mathbb{T}] \vDash \mathsf{TB}_0$ *iff* $\mathring{\mathbb{T}} \cap \mathrm{Sent}(PA) = TA$*.*

Proof. ($\Rightarrow$) First suppose $\mathbb{N}[\mathbb{T}] \vDash \mathsf{TB}_0$. Let $\varphi \in \mathring{\mathbb{T}} \cap \mathrm{Sent}(PA)$. Then $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\varphi\urcorner)$ since $\ulcorner\varphi\urcorner \in \mathbb{T}$. Then since $\mathbb{N}[\mathbb{T}] \vDash \mathsf{TB}_0$ and φ is a arithmetic sentence, we have $\mathbb{N}[\mathbb{T}] \vDash \varphi$. Since $\varphi \in \mathrm{Sent}(PA)$, we have $\mathbb{N} \vDash \varphi$ hence $\varphi \in TA$. Now to show that $TA \subset \mathring{\mathbb{T}} \cap \mathrm{Sent}(PA)$, let $\varphi \in TA$. It suffices to verify that $\varphi \in \mathring{\mathbb{T}}$. Since φ is a true arithmetic sentence, $\mathbb{N} \vDash \varphi$, and so $\mathbb{N}[\mathbb{T}] \vDash \varphi$. By hypothesis, $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\varphi\urcorner) \leftrightarrow \varphi$, so $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\varphi\urcorner)$, thus $\varphi \in \mathring{\mathbb{T}}$.

($\Leftarrow$) Suppose $\mathring{\mathbb{T}} \cap \mathrm{Sent}(PA) = TA$. Let $\varphi \in \mathrm{Sent}(PA)$ and consider $\psi \equiv T(\ulcorner\varphi\urcorner) \leftrightarrow \varphi$. Suppose $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\varphi\urcorner)$. Then $\varphi \in \mathring{\mathbb{T}}$ and hence $\varphi \in TA$ by hypothesis. It follows that $\mathbb{N} \vDash \varphi$ and so $\mathbb{N}[\mathbb{T}] \vDash \varphi$. Now suppose $\mathbb{N}[\mathbb{T}] \vDash \varphi$. Since $\varphi \in \mathrm{Sent}(PA)$ we have $\mathbb{N} \vDash \varphi$ and so $\varphi \in TA$. It follows that $\langle\varphi\rangle \in \mathbb{T}$ and so $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\varphi\urcorner)$. Thus $\mathbb{N}[\mathbb{T}] \vDash \psi$, an arbitrary Tarski-biconditional, therefore $\mathbb{N}[\mathbb{T}] \vDash \mathsf{TB}_0$. ☐

TB_0 has some deficiencies, making it an unsatisfactory theory of truth. For example, it cannot prove certain compositional truth principles in universal form.[20]

[20] What follows is a quick proof involving machinery developed earlier in this chapter. The standard proof is sketched in, e.g., Horsten and Leigh (2017, p. 6), but the result has been known since Tarski (1936).

Proposition 4.34. TB_0 *does not prove the commutation (composition) of truth with conjunction, i.e.:*

$$\mathsf{TB}_0 \nvdash \forall\varphi\forall\psi(T(\ulcorner\varphi\wedge\psi\urcorner)\leftrightarrow(T(\ulcorner\varphi\urcorner)\wedge T(\ulcorner\psi\urcorner))) \tag{4.14}$$

Proof. Let φ,ψ be $L[T]$-formulas such that $\varphi,\psi\notin \mathrm{Sent}(PA)$. Consider $\mathbb{T} = TA\cup\{\langle\varphi\wedge\psi\rangle\}$. Since $\mathring{\mathbb{T}}\cap\mathrm{Sent}(PA) = TA$, $\mathbb{N}[\mathbb{T}]\vDash\mathsf{TB}_0$ by Lemma 4.33. Hence $\mathbb{N}[\mathbb{T}]\vDash T(\ulcorner\varphi\wedge\psi\urcorner)$ but $\mathbb{N}[\mathbb{T}]\nvDash T(\ulcorner\varphi\urcorner)\wedge T(\ulcorner\psi\urcorner)$. □

However, the system might suffice as a starting point.

Proposition 4.35. TB_0 *is consistent, i.e. it has a model.*

Proof. Let $\mathbb{T} = \langle TA\rangle$, then by Lemma 4.33, $\mathbb{N}[\mathbb{T}]\vDash\mathsf{TB}_0$. □

4.6.1 The Consistency of $\mathsf{TB}_0\mathsf{R}$ and $\mathsf{TB}_0\mathsf{R}$-bar

We now examine the system $\mathsf{TB}_0\mathsf{R}$, obtained taking the grefl-deductive-closure of TB_0.

Proposition 4.36. $\mathsf{TB}_0\mathsf{R}$ *satisfies the conditions of Theorem 4.11, hence* $\mathsf{TB}_0\mathsf{R}$ *is consistent.*

Proof. It suffices to establish that TB_0 is sound for jump models. By *Lemma*4.33, we just need to verify that all jump models extend TA which is immediate since $\mathbb{N}\vDash TA$. □

We can strengthen the previous result by providing an ω-model for $\mathsf{TB}_0\mathsf{R}$.

Proposition 4.37. $\mathsf{TB}_0\mathsf{R}$ *is ω-consistent, i.e. it has an ω-model.*

Proof. Let $\mathbb{T} = \{\langle\varphi\rangle \,|\, \varphi \in \mathsf{TB_0R} \cup TA\}$. We proceed by induction on the length l of an grefl-proof $\pi = \langle\varphi_1, \ldots, \varphi_l\rangle$ from $\mathsf{TB_0}$ that $\mathbb{N}[\mathbb{T}] \vDash \varphi_l$.

For the base case, $l = 1$, it suffices to show that $\mathbb{N}[\mathbb{T}] \vDash \mathsf{TB_0}$ since the length 1 proofs are just instances of axioms. By Lemma 4.33, this reduces to verifying that $\mathring{\mathbb{T}} \cap \mathrm{Sent}(PA) = TA$. The right-to-left inclusion is immediate. Now suppose that $\varphi \in \mathring{\mathbb{T}} \cap \mathrm{Sent}(PA)$. If $\varphi \in TA$, then there's nothing to show. If $\varphi \in \mathsf{TB_0R}$, then by Corollary 4.12 we get that $\varphi \in TA$.

For the inductive step, suppose that the result holds for proofs of length l. We have three cases to consider.

In the first case, φ_{l+1} is a tautology or an axiom of $\mathsf{TB_0}$. This is analogous to the base case.

In the second case, φ_{l+1} is obtained by modus ponens, i.e. there are $j, k \leq l$ such that $\varphi_j \equiv \varphi_k \to \varphi_{l+1}$. Since $\mathbb{N}[\mathbb{T}]$ satisfies φ_k and φ_j by induction hypothesis, it must also satisfy φ_{l+1}.

In the third case, $\varphi_{l+1} \equiv \forall\psi(\mathrm{Bwb}_{\mathsf{TB_0}+\varphi_j}(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner))$ for some $j \leq l$. Suppose that $\mathbb{N}[\mathbb{T}] \vDash \mathrm{Bwb}_{\mathsf{TB_0}+\varphi_j}(\ulcorner\psi\urcorner)$ for some ψ. Since π is an $l+1$ length grefl-proof of φ_{l+1}, it follows that φ_j is grefl-provable in a length $j \leq l$ proof. By our supposition, we have that $\mathsf{TB_0} \cup \{\varphi_j\} \vdash \psi$. Since $\mathsf{TB_0R}$ is deductively-closed and since both $\mathsf{TB_0}$ and φ_j are in $\mathsf{TB_0R}$, we then have $\langle\psi\rangle \in \mathsf{TB_0R} \subset \mathbb{T}$. Thus $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\psi\urcorner)$. □

We note two quick corollaries. First, $\mathsf{TB_0R}$ is consistent with its own global reflection principle.

Corollary 4.38. $\mathsf{TB_0R+} = \mathsf{TB_0R} \cup \{\forall\psi(\mathrm{Bwb}^{\mathsf{R}}_{\mathsf{TB_0}}(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner))\}$ *is consistent.*

Proof. This follows immediately from Proposition 4.37 and Corollary 4.15. □

Shifting our focus to $\mathsf{TB_0\bar{R}}$, we find the following:

Corollary 4.39. $\mathsf{TB_0\bar{R}}$ *is consistent. Moreover, the theory* $\mathsf{TB_0R} + \mathsf{TB_0\bar{R}}$ *is consistent.*

Proof. This follows from the same reasoning in Proposition 4.36. □

4.7 Examples of Joint Inconsistency

As we saw for FS (trivially) and $\mathsf{TB_0}$, the base theories were consistent with either reflection principle and in fact the resulting theories were jointly consistent. This motivates the question of whether this is always the case. We can construct a simple theory establishing the contrary.

Proposition 4.40. *Let* $S = PA + \neg\mathrm{Con}(PA)$*.* SR *and* $S\bar{R}$ *are individually consistent but jointly inconsistent.*

Proof. Let $\mathcal{N}$ be a model of S. Extend $\mathcal{N}$ to a new model $\mathcal{N}'$ by providing an extension for the truth predicate $T = dom(\mathcal{N})$; similarly define $\mathcal{N}''$ by taking $T = \varnothing$. By rote induction, $\mathcal{N}' \vDash SR$ and $\mathcal{N}'' \vDash S\bar{R}$.

Now, suppose toward a contradiction that $SR + S\bar{R}$ were consistent with model $\mathcal{M}$. We know that $PA \vdash 0 \neq 1$, and since S extends PA, we have $S \vdash 0 \neq 1$. By Theorem 4.14, it follows that $SR \vdash T(\ulcorner 0 \neq 1 \urcorner)$ and hence $\mathcal{M} \vDash T(\ulcorner 0 \neq 1 \urcorner)$. Let φ be some fixed tautology. Then by G-$\bar{R}\mathrm{efl}^1_S$ we have $\mathcal{M} \vDash \neg\mathrm{Bwb}_{S+\varphi}(\ulcorner \neg 0 \neq 1 \urcorner)$. Since $\mathrm{Bwb}(n)$ captures classical provability, this happens just in case $\mathcal{M} \vDash \neg\mathrm{Bwb}_{S+\varphi}(\ulcorner 0 = 1 \urcorner)$, i.e. $\mathcal{M} \vDash \mathrm{Con}(S + \varphi)$. Since S extends PA, we then have $\mathcal{M} \vDash \mathrm{Con}(PA)$, a contradiction. □

The theory S employed above is an unnatural theory which no philosopher would adopt. While it does provide an example of a theory for which the two reflection principles are incompatible, if we take the implicit commitment thesis seriously (Dean (2015)), then we have prior reasons for excluding this theory. So some caution should be taken in interpreting the proposition; at the very least, it motivates further investigation to see if there are more-plausible theories of truth for which we find similar results. What we find is that KF is in fact such a theory. The following proof bears similarity to the proof of Halbach and Horsten (2006, p. 688, Lemma 11) that KF is incompatible with the NEC and CONEC rules.

Proposition 4.41. KF *is such that* KFR *and* $\mathsf{KF\bar{R}}$ *are individually consistent but jointly inconsistent.*

Proof. We observed their individual consistency in §4.5, so it remains to demonstrate their joint inconsistency. Let $S = \mathsf{KFR} + \mathsf{KF\bar{R}}$. By diagonalization, we have both

$$PA \vdash \lambda \vee T(\ulcorner \lambda \urcorner) \tag{4.15}$$

$$PA \vdash \neg\lambda \vee \neg T(\ulcorner \lambda \urcorner) \tag{4.16}$$

Applying G-Refl$^1_{\mathsf{KF}}$ to each of (4.15) and (4.16) gives

$$S \vdash T(\ulcorner \lambda \vee T(\ulcorner \lambda \urcorner) \urcorner) \tag{4.17}$$

$$S \vdash T(\ulcorner \neg\lambda \vee \neg T(\ulcorner \lambda \urcorner) \urcorner) \tag{4.18}$$

Applying KF6 to each of (4.17) and (4.18) followed by KF12 and KF13 respectively yields the following after reduction by basic logical laws:

$$S \vdash T(\ulcorner \lambda \urcorner) \tag{4.19}$$

$$S \vdash T(\ulcorner \neg\lambda \urcorner) \tag{4.20}$$

By conjoining (4.19) and (4.20) and applying KF4 we then obtain

$$S \vdash T(\ulcorner \lambda \wedge \neg\lambda \urcorner) \tag{4.21}$$

Applying G-$\bar{R}\mathrm{efl}^1_{\mathsf{KF}}$ to (4.21) gives that

$$S \vdash \neg\mathrm{Bwb}_{KF+\varphi}(\ulcorner\neg(\lambda \wedge \neg\lambda)\urcorner) \tag{4.22}$$

Here, φ is some arbitrary formula (say, some fixed tautology) provable in KF. But this is a contradiction, since $PA \vdash \neg(\lambda \wedge \neg\lambda)$ and KF extends PA, which S witnesses. $\square$

Also notable in the first counterexample is the lack of any explicit truth axioms. This motivates the question: are there natural or minimal truth principles such that for any $L[T]$-theory S deriving them, SR and $S\bar{R}$ are jointly consistent if they are independently consistent? We prove a pair of preliminary results below. These reinforce the centrality of the completeness and consistency truth schemas described in §4.5.2. The ambivalence of KF toward these schemas hence compromised its joint consistency.

Proposition 4.42. *Let S be an $L[T]$-theory such that there is $\mathbb{T} \subset \omega$ with $\mathbb{N}[\mathbb{T}] \vDash SR$. Then if $S \vdash \forall\psi(\neg(T(\ulcorner\psi\urcorner) \wedge T(\ulcorner\neg\psi\urcorner)))$ it follows $\mathbb{N}[\mathbb{T}] \vDash S\bar{R}$. Hence SR and $S\bar{R}$ are jointly consistent.*

Proof. We proceed by induction on g$\bar{r}$efl-proof length from S. The base case and the first two subcases of inductive step follow from the fact that $\mathbb{N}[\mathbb{T}] \vDash SR$. Now suppose $\varphi_{l+1} \equiv \forall\psi(\mathrm{Bwb}_{S+\varphi_j}(\ulcorner\neg\psi\urcorner) \to \neg T(\ulcorner\psi\urcorner))$. Suppose toward a contradiction that this was false on the model. Then there would be ψ such that $S + \varphi_j \vdash \neg\psi$ and $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\psi\urcorner)$. But we have that $SR \vdash T(\ulcorner\neg\psi\urcorner)$ by Theorem 4.14, so $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\psi\urcorner) \wedge T(\ulcorner\neg\psi\urcorner)$, a contradiction since $\mathbb{N}[\mathbb{T}] \vDash S$. $\square$

Proposition 4.43. *Let S be an $L[T]$-theory such that there is $\mathbb{T} \subset \omega$ with $\mathbb{N}[\mathbb{T}] \vDash S\bar{R}$. Then if $S \vdash \forall\psi(T(\ulcorner\psi\urcorner) \vee T(\ulcorner\neg\psi\urcorner))$ it follows $\mathbb{N}[\mathbb{T}] \vDash SR$. Hence SR and $S\bar{R}$ are jointly consistent.*

Proof. We proceed by induction on grefl-proof length from S. The base case and the first two subcases of inductive step follow from the fact that $\mathbb{N}[\mathbb{T}] \vDash SR$. Now suppose $\varphi_{l+1} \equiv$

$\forall\psi(\mathrm{Bwb}_{S+\varphi_j}(\ulcorner\psi\urcorner) \to T(\ulcorner\psi\urcorner))$. Suppose toward a contradiction that this was false on the model. Then there would be ψ such that $S + \varphi_j \vdash \psi$ and $\mathbb{N}[\mathbb{T}] \vDash T(\ulcorner\neg\psi\urcorner)$ (since $\mathbb{N}[\mathbb{T}] \vDash \forall\psi(T(\ulcorner\psi\urcorner) \vee T(\ulcorner\neg\psi\urcorner)))$. Since $\mathbb{N}[\mathbb{T}] \vDash S\bar{R}$, we have from the latter conjunct that $\mathbb{N}[\mathbb{T}] \vDash \neg\mathrm{Bwb}_{S+\varphi_j}(\ulcorner\neg\neg\psi\urcorner)$. It follows that $S + \varphi_j \nvdash \neg\neg\psi$, a contradiction. □

The previous two propositions also hold under the weaker hypothesis that the model $\mathbb{N}[\mathbb{T}]$ satisfies the S-theorem identified, rather than that S itself proves this. In light of this, we note a quick corollary which strengthens Proposition 4.39.

Corollary 4.44. *Let* $\mathbb{T} = \{\langle\varphi\rangle \,|\, \varphi \in \mathsf{TB_0R} \cup TA\}$. *Then* $\mathbb{N}[\mathbb{T}] \vDash \mathsf{TB_0R} + \mathsf{TB_0\bar{R}}$.

Proof. From Proposition 4.36, we know that $\mathbb{N}[\mathbb{T}] \vDash \mathsf{TB_0R}$. Since $\mathsf{TB_0R} + TA$ is consistent, it follows that $\mathbb{N}[\mathbb{T}] \vDash \forall\psi(\neg(T(\ulcorner\psi\urcorner) \wedge T(\ulcorner\neg\psi\urcorner)))$. Hence by Proposition 4.42, $\mathbb{N}[\mathbb{T}] \vDash \mathsf{TB_0\bar{R}}$. □

4.8 Prospects for the Horsten-Leigh Project

The results of sections §4.4-4.7 are summarized in Table 4.1. In particular, we recall the consistency results (abbreviated as 'Cons.').

Theory S	SR Cons.?	$S\bar{R}$ Cons.?	$SR + S\bar{R}$ Cons.?
FS	✓	✓	✓
KF	✓	✓	×
$(\mathsf{KF} + \text{T-in})$	✓	×	×
$(\mathsf{KF} + \text{T-out})$	×	✓	×
$\mathsf{TB_0}$	✓	✓	✓
$PA + \neg\mathrm{Con}(PA)$	✓	✓	×

Table 4.1: The consistency results for each theory.

Horsten and Leigh (2017) apply their method to the system of Tarski biconditionals $\mathsf{TB_0}$ as a validation of their argument for the means by which one is entitled to reflection. Though they only examine the theory under a few iterations of reflection, the results of §4.6 give further support by demonstrating consistency at the ω-th level in $\mathsf{TB_0R}$. In addition to the orthodox reflection principle, we find that our dual reflection principle enjoys the same success and is even consistent with the standard principle. But while this case seems to validate their philosophical argument, we find that other implementations are not so fortuitous.

Our extensions of Kripke-Feferman theory (KF + T-in) and (KF + T-out) revealed that the choice of reflection principle impacts the success of the method. In particular, each of the extensions is compatible with only one of the reflection principles. If Horsten and Leigh wish to support either both reflection principles or one in particular, then they'll be in effect ruling out a class of base theories and ought to provide an argument in support of this exclusion. And beyond complicating the class of permissible theories, if they support precisely one of the reflection principles, the authors need a more narrow philosophical argument which is incompatible with the informal contraposition argument given at the end of §4.2.2. The joint inconsistency result for KF presses the issue by highlighting the difficulty in endorsing both types of reflection.

As we have seen, the choice of base theory can also impact the resulting theory significantly. As noted in the previous paragraph, certain theories are compatible with only certain reflection principles. But as we saw with the Friedman-Sheard theory of truth in §4.4, even if we find no conflict with the type of reflection principle, we can still run into problems. In particular, we recovered the known result that FSR is inconsistent with its own global reflection principle.

A more conservative way to view the results regarding FS and KF is to say that reflection is working as intended: it renders explicit what was implicit in the base theory. This includes defects like ω-inconsistency or incompatible inner and outer logics. But still, further argument for why one ought not to accept an only implicitly defective theory is warranted.

Having posed some criticisms of the Horsten-Leigh project, it's worth spending a moment to consider some possible objections a proponent of the project might have. First, Horsten and Leigh demonstrated that their method worked with TB_0 and in §4.6 I extend their results in further validation of the theory. So it might appear that Horsten and Leigh have succeeded in their goal and so should not be too bothered by the concerns noted above. In order to understand why this is not the case, it's important to separate out two aspects of the Horsten-Leigh paper: the argument for reflection and the role of TB_0. Horsten and Leigh begin by providing a philosophical argument for iterative reflection, then on the basis of the legitimacy of the technique, apply it to TB_0 to illustrate that a simple base theory of truth implicitly yields the desired compositional truth principles. Nothing in the philosophical argument singles out TB_0 as the theory we should adopt, and so my analysis of the other theories to which the argument could apply reveals that the technique is problematic.

A second objection could arise in a comparison of the formal implementations of reflection between this chapter and Horsten and Leigh (2017): I apply reflection to the ω-th level, but Horsten and Leigh examine only a finite number of iterations. I find my approach to be most consonant with Horsten and Leigh's philosophical argument for reflection, for each application of reflection renders explicit new commitments which can then be reflected upon. I believe Horsten and Leigh's focus on a finite stage of the reflective process isn't grounded in a belief that that particular stage is the ultimate one. Rather, it required only finitely-many iterations to derive the compositional principles which are normally cited as deficiencies of TB_0 (Horsten and Leigh, 2017, Theorems 2 and 3, p. 13).

A third worry concerns the theories of truth I examine: TB_0 is not compositional but KF and FS are, and part of Horsten and Leigh's goal was to recover compositionality via reflection. While I think the force of my argument would be greater had I found similar results for another non-compositional base theory, the given evidence is sufficient.[21] After all, at some

[21]This is not to say that I have found positive results for such theories. Rather, there are only a handful of axiomatic theories in the literature, and I have found no better candidates to examine.

point in the reflective process we aim to obtain the compositional rules. If one is to continue the iterative process of reflection, there appears to be no harm in accepting the rules earlier. Further, the authors claim that justification in accepting a theory is sufficient for "adopting a reflection principle for that theory" (p. 15). Without an argument for why one cannot be justified in accepting, e.g., KF, this claim appears to be in general false – it needs further qualification to support its application to TB_0 and exclusion from KF.

Outside of the Horsten and Leigh project, an opponent of global reflection principles might find some vindication. A bird's-eye view of the analysis conducted here witnesses the many pitfalls which may befall a proponent of global reflection. Weaker reflection principles, or ones based on a non-classical notion of provability, would be appealing to an advocate of KF as a means of avoiding the issues introduced here.

4.9 Further Directions

In the previous section, we weighed the evidence against Horsten and Leigh's assertion that "truth is simple". A possible way out for Horsten and Leigh would be to move to higher-order reflection principles. This is motivated by a connection to set theory. Set-theoretic reflection principles are given by the following schema[22]:

$$V \vDash \varphi \Rightarrow \exists \alpha V_\alpha \vDash \varphi \tag{4.23}$$

Here V denotes the von Neumann universe of sets and V_α is the α-th level of the cumulative hierarchy. Contraposing (4.23) and substituting $\neg\psi$ for φ gives

$$\forall \alpha V_\alpha \vDash \psi \Rightarrow V \vDash \psi \tag{4.24}$$

[22] For one such formulation, see Theorem 29 of Jech (2006, p. 89)

We can gloss (4.24) as 'if φ is true on all (standard) models, then φ is true in the universe'. By completeness, truth on all models is equivalent to provability, so we then obtain 'if ψ is provable, then ψ is true in the universe', which is comparable to the typical formulation of an arithmetic reflection principle.

As discussed above, the arithmetic reflection principles are motivated by a connection between provability from true axioms and truth, i.e. that provability preserves truth. Set-theoretic reflection principles are instead canonically viewed as illustrating that the full universe can't be captured by a (first-order) formula, in that there will be some initial segment which suffices in satisfying it. The aim of the above transformation is to present the two usages of 'reflection principle' as having more in common than phonetics.

Despite differences in motivation, perhaps the two contexts share a common moral regarding consistency. Recall that higher-order reflection principles are inconsistent in set theory (Koellner, 2009). This is due to higher-order logic forcing a 'match up' of theory and metatheory (e.g. third-order validity is higher-order and set-theoretic). Maybe in the arithmetical setting lower-order reflection principles result in inconsistency due to first-order logic forcing a 'match up' of theory and metatheory (e.g. provability is first-order and arithmetic).

This motivates a study of higher-order reflection principles in arithmetic. What this would involve is a move to higher-order arithmetic (e.g. second-order Peano arithmetic) and a generalization of the reflection principles to apply to higher-order sentences. The hope for Horsten and Leigh in this setting is that the negative results we found for theories like KF might not apply to their higher-order counterparts. But I leave the study of higher-order reflection principles in arithmetic to future work.

Another area for future research is to broaden or generalize the methods of this chapter. It would be of general interest, and relevant to the Horsten and Leigh project, to apply this analysis to other theories of truth, such as the supervaluationist theory VF (Cantini, 1990). And while we focused on candidate theories of truth, a better understanding of the reflection principles could be obtained by finding their minimal interactions with various truth principles in the spirit of Friedman and Sheard (1987). For instance, only a handful of the Kripke-Feferman axioms are needed to obtain the results of §4.5 and §4.7 and hence the propositions there apply to a more general class of theories.

Bibliography

Bacon, Andrew, John Hawthorne, and Gabriel Uzquiano (2016). Higher-order free logic and the Prior-Kaplan paradox. *Canadian Journal of Philosophy*, 46(4-5): 493–541.

Bale, Alan Clinton (2008). A universal scale of comparison. *Linguistics and Philosophy*, 31(1): 1.

Barwise, Jon and Robin Cooper (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2): 159–219.

Bennett, Jonathan Francis (2003). *A Philosophical Guide to Conditionals*. Oxford University Press.

Bennett, Karen (2005). Two axes of actualism. *The Philosophical Review*, 114(3): 297–326.

Berto, Francesco (2013). *Existence as a Real Property*. Springer.

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

Boolos, George (1995). *The Logic of Provability*. Cambridge university press.

Boolos, George (1998). *Logic, Logic, and Logic*. Harvard University Press, Cambridge, MA.

Burge, Tyler (1979). Semantical paradox. *The Journal of Philosophy*, 76(4): 169–198.

Burge, Tyler (2003). Perceptual entitlement. *Philosophy and Phenomenological Research*, 67(3): 503–548.

Burgess, John P (1997). Quinus ab omni nævo vindicatus. *Canadian Journal of Philosophy*, 27(sup1): 25–65.

Cantini, Andrea (1989). Notes on formal theories of truth. *Mathematical Logic Quarterly*, 35(2): 97–130.

Cantini, Andrea (1990). A theory of formal truth arithmetically equivalent to ID_1. *The Journal of Symbolic Logic*, 55(1): 244–259.

Carnap, Rudolf (1932). The elimination of metaphysics through logical analysis of language. *Logical empiricism at its peak: Schlick, Carnap, and Neurath*, 2: 10.

Chalmers, David J. (2006). Two-dimensional semantics. In Lepore, E. and B. Smith, editors, *Oxford Handbook of the Philosophy of Language*. Oxford University Press.

Chang, C C and H J Keisler (1990). *Model Theory*, volume 73 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, Amsterdam, third edition.

Cieśliński, Cezary (2010). Truth, conservativeness, and provability. *Mind*, 119(474): 409–422.

Cresswell, Max J (1976). The semantics of degree. In *Montague Grammar*, pages 261–292. Elsevier.

Cresswell, Maxwell John (1990). *Entities and Indices*, volume 41 of *Studies in Linguistics and Philosophy*. Springer.

Davies, Martin and Lloyd Humberstone (1980). Two notions of necessity. *Philosophical Studies*, 38(1): 1–31.

Dean, Walter (2015). Arithmetical reflection and the provability of soundness. *Philosophia Mathematica*, 23(1): 31–64.

Dowty, David R., Robert Eugene Wall, and Stanley Peters (1981). *Introduction to Montague Semantics*. Springer.

Elbourne, Paul (2013). *Definite Descriptions*. Oxford University Press, Oxford.

Feferman, Solomon (1991). Reflecting on incompleteness. *J. Symb. Logic*, 56(1): 1–49.

Fine, Kit (1981). Model theory for modal logic—part iii existence and predication. *Journal of Philosophical Logic*, 10(3): 293–307.

Fine, Kit (2005). Prior on the construction of possible worlds and instants. In *Modality and Tense: Philosophical Papers*, pages 152–159. Oxford University Press.

Fitting, Melvin and Richard Mendelsohn (1998). *First-order Modal Logic*. Kluwer Academic, Dordrecht Boston.

Frege, Gottlob (1884). *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl*. w. Koebner.

Frege, Gottlob (1892). Über begriff und gegenstand. *Vierteljahrsschrift für wissenschaftliche Philosophie*.

Friedman, Harvey and Michael Sheard (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33(1): 1–21.

Fritz, Peter (2013). Modal ontology and generalized quantifiers. *J. Philos. Logic*, 42(4): 643–678.

Fritz, Peter and Jeremy Goodman (2016). Higher-order contingentism, part 1: Closure and generation. *Journal of Philosophical Logic*, 45(6): 645–695.

Fritz, Peter and Jeremy Goodman (2017). Counting incompossibles. *Mind*.

Gallin, Daniel (2016). *Intensional and Higher-order Modal Logic*. Elsevier.

Gamut, L. T. F. (1991). *Logic, Language, and Meaning, Volume 2: Intensional Logic and Intensional Grammar*. Chicago: Univ. of Chicago Press.

Goodman, Jeremy (2016a). An argument for necessitism. *Philosophical Perspectives*, 30(1): 160–182.

Goodman, Jeremy (2016b). Williamson on necessitism. *Canadian Journal of Philosophy*, 46(4-5): 613–639.

Grätzer, George (1979). *Universal Algebra*. Springer.

Groenendijk, Jeroen and Martin Stokhof (1991). Dynamic predicate logic. *Linguistics and philosophy*, pages 39–100.

Hájek, Petr and Pavel Pudlák (1998). *Metamathematics of First-order Arithmetic*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin. Second printing.

Halbach, Volker (2011). *Axiomatic Theories of Truth*. Cambridge University Press.

Halbach, Volker and Leon Horsten (2006). Axiomatizing kripke's theory of truth. *J. Symbolic Logic*, 71(2): 677–712.

Halpern, Joseph Y and Yoav Shoham (1991). A propositional modal logic of time intervals. *Journal of the Association for Computing Machinary*, 38(4): 935–962.

Hintikka, Jaakko (1956). Identity, variables, and impredicative definitions. *The Journal of Symbolic Logic*, 21(3): 225–245.

Hintikka, Jaakko (1969a). Existential presuppositions and their elimination. In *Models for Modalities*, pages 23–44. Springer.

Hintikka, Jaakko (1969b). The logic of perception. In Hintikka, Jaakko, editor, *Models for Modalities*. Reidel.

Hodges, Wilfrid (1993). *Model Theory*, volume 42 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

Horsten, Leon and Graham E. Leigh (2017). Truth is simple. *Mind*, 126(501): 195–232.

Hughes, George Edward and Max J Cresswell (1996). *A New Introduction to Modal Logic*. Psychology Press.

Janssen, T.M.V. (1986). Foundations and applications of montague grammar. part 2: Applications to natural language. *CWI Tracts*, 28(2).

Jech, Thomas (2006). *Set Theory*. Springer Monographs in Mathematics. Springer Berlin Heidelberg.

Kanamori, Akihiro (2003). *The Higher Infinite*. Springer, Berlin, second edition.

Kant, Immanuel (1998). *Critique of Pure Reason (Translated and Edited by Paul Guyer & Allen W. Wood)*. Cambridge: Cambridge University Press.

Kaplan, David and Richard Montague (1960). A paradox regained. *Notre Dame journal of formal logic*, 1(3): 79–90.

Kaye, Richard (1991). *Models of Peano Arithmetic*, volume 15 of *Oxford Logic Guides*. Oxford University Press, New York.

Keisler, H Jerome (1964). Ultraproducts and saturated models. *Indag. Math*, 26: 178–186.

Kennedy, Christopher and Louise McNally (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, pages 345–381.

Ketland, Jeffrey (1999). Deflationism and Tarski's paradise. *Mind*, 108(429): 69–94.

Koellner, Peter (2009). On reflection principles. *Annals of Pure and Applied Logic*, 157(2): 206–219.

Koellner, Peter (2010). On the question of absolute undecidability. In Gödel, Kurt, Solomon Feferman, Charles Parsons, and Stephen G. Simpson, editors, *Philosophia Mathematica*, pages 153–188. Association for Symbolic Logic.

Kossak, Roman and Jim Schmerl (2006). *The Structure of Models of Peano Arithmetic*. Oxford Logic Guides. Clarendon Press.

Kratzer, Angelika (1989). An investigation of the lumps of thought. *Linguistics and Philosophy*, 12(5): 607–653.

Kratzer, Angelika (2012). *Modals and Conditionals: New and Revised Perspectives*, volume 36. Oxford University Press.

Kratzer, Angelika (2019). Situations in natural language semantics. In Zalta, Edward N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2019 edition.

Kraut, Robert (1979). Attitudes and their objects. *Journal of Philosophical Logic*, 8(1): 197–217.

Kraut, Robert (1982). Sensory states and sensory objects. *Noûs*.

Kreisel, Georg and Azriel Lévy (1968). Reflection principles and their use for establishing the complexity of axiomatic systems. *Mathematical Logic Quarterly*, 14(7-12): 97–142.

Kripke, Saul A (1972). Naming and necessity. In *Semantics of Natural Language*, pages 253–355. Springer.

Kunen, Kenneth (1980). *Set Theory*, volume 102 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, Amsterdam.

Lang, Serge (2002). *Algebra*, volume 211 of *Graduate Texts in Mathematics*. Springer, New York, third edition.

Luce, Robert Duncan, Patrick Suppes, and David H Krantz (2007). *Foundations of Measurement: Representation, Axiomatization, and Invariance*, volume 3. Courier Corporation.

Maddy, Penelope (1997). *Naturalism in Mathematics*. Oxford University Press.

Marker, David (2002). *Model Theory: An Introduction*, volume 217 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.

Marx, Maarten and Yde Venema (1997). *Multi-Dimensional Modal Logic*. Springer, Dordrecht.

McGee, Vann (1985). How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, 14(4): 399–410.

Montague, Richard (1961). Fraenkel's addition to the axioms of zermelo. In Y. Bar-Hillel, M. O. Rabin, E. I. J. Poznanski and A. Robinson, editors, *Essays on the Foundations of Mathematics*, pages 91–114. The Magnes Press, Jerusalem.

Montague, Richard (1963). Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica*, 16: 153–167.

Montague, Richard (1970). Universal grammar. *Theoria*, 36(3): 373–398.

Montague, Richard (1973). The proper treatment of quantification in ordinary english. In Suppes, Patrick, Julius Moravcsik, and Jaakko Hintikka, editors, *Approaches to Natural Language*, pages 221–242. Dordrecht.

Montague, Richard (1974). *Formal Philosophy*. Yale University Press, New Haven and London. Edited by Richmond H Thomason.

Oliver, Alex and Timothy Smiley (2013). *Plural Logic*. Oxford University Press, Oxford.

Parsons, Terence (1980). *Nonexistent Objects*. Yale University Press, New Haven.

Putnam, Hilary (1973). Meaning and reference. *Journal of Philosophy*, 70(19): 699–711.

Quine, Willard V (1948). On what there is. *The review of metaphysics*, 2(1): 21–38.

Reinhardt, William N. (1986). Some remarks on extending and interpreting theories with a partial predicate for truth. *Journal of Philosophical Logic*, 15(2): 219–251.

Rogers, Brian and Kai F Wehmeier (2012). Tractarian first-order logic: Identity and the n-operator. *The Review of Symbolic Logic*, 5(4): 538–573.

Russell, Bertrand (1903). *The Principles of Mathematics*. Cambridge University Press.

Sakai, Katsuro (2013). *Geometric Aspects of General Topology*. Springer, Tokyo.

Schindler, Ralf (2014). *Set Theory: Exploring Independence and Truth*. Springer.

Shapiro, Stewart (1991). *Foundations Without Foundationalism: A Case for Second-Order Logic*. Oxford University Press.

Shapiro, Stewart (1998). Proof and truth: Through thick and thin. *Journal of Philosophy*, 95(10): 493–521.

Solovay, Robert M. (1976). Provability interpretations of modal logic. *Israel Journal of Mathematics*, 25(3): 287–304.

Stalnaker, Robert (2012). *Mere Possibilities: Metaphysical Foundations of Modal Semantics*. Princeton University Press.

Tarski, Alfred (1936). *The Concept of Truth in Formalized Languages*. Oxford University Press.

Troelstra, Anne and Helmut Schwichtenberg (2000). *Basic Proof Theory*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press.

Updike, Eric T (2019). Is objectual identity really dispensable? *Erkenntnis*, pages 1–22.

van der Does, Jaap M and Michiel Van Lambalgen (2000). A logic of vision. *Linguistics and Philosophy*, 23(1): 1–92.

von Stechow, Arnim (1984). Comparing semantic theories of comparison. *Journal of semantics*, 3(1-2): 1–77.

Wehmeier, Kai F (2004). Wittgensteinian predicate logic. *Notre Dame Journal of Formal Logic*, 45(1): 1–11.

Wehmeier, Kai F (2008). Wittgensteinian tableaux, identity, and co-denotation. *Erkenntnis*, 69(3): 363–376.

Wehmeier, Kai F (2012). How to live without identity—and why. *Australasian Journal of Philosophy*, 90(4): 761–777.

Wehmeier, Kai F (2018). The proper treatment of variables in predicate logic. *Linguistics and Philosophy*, 41(2): 209–249.

Williamson, Timothy (2010). Necessitism, contingentism, and plural quantification. *Mind*, 119(475): 657–748.

Williamson, Timothy (2013). *Modal Logic as Metaphysics*. Oxford University Press.

Williamson, Timothy (2016). Reply to Goodman. *Canadian Journal of Philosophy*, 46(4-5): 640–653.

Wittgenstein, Ludwig (1971). *Tractatus Logico-philosophicus*. Routledge & Kegan Paul. Translated by D. F. Pears and B. F. McGuinness.

Zalta, Edward (2012). *Abstract Objects: An Introduction to Axiomatic Metaphysics*, volume 160. Springer Science & Business Media.

Appendix A

Replies to Updike

In this appendix, we address Updike's criticisms regarding the viability of an existence predicate for the W-logician.

To avoid skewing the debate by taking a hard stance the ontological status of an existence predicate, Updike (2019) entertains the solution of taking the meaning of the existence predicate to be (partially) described by a semantic constraint, i.e. the condition provided in the satisfaction clause for our models. Even specified this way, Updike claims the existence predicate is still problematic for the W-logician, enumerating a number of concerns. We first examine an earlier claim of Updike's, and then consider his problems in turn.

In motivating the existence predicate, Updike considers the consequences of a locally-universal predicate (by which he means formula of one free variable), like our $\top(x)$. Updike's desideratum is the W-logical translation of metaphysical theses like the necessity of being. Below is Updike's QML$^{=}$-rendition.

$$\forall x \, \Box \, \exists y \; y = x \tag{$\mathrm{NB_I}$}$$

He observes that one possible translation would use a universal predicate to obtain a formula like $\forall x \, \Box \, \top(x)$. This requires an additional background assumption like the *being constraint* (that predication implies existence), contrary to our more general apparatus in Definition 1.21. However, crucial in Updike's assumption is that the being constraint applies to not just atomic predicates, but any formula with free variables. Schematically, his gloss of the being constraint below takes the universal $\forall P$ to range over any formula $\varphi(x)$:

$$\forall P \, \Box \, \forall x \, \Box \, (Px \to \exists y \; y = x) \tag{BC}$$

With this assumption in place, Updike argues that any strategy of translation like $\forall x \, \Box \, \top(x)$ will be unsatisfactory because it will always be a theorem (derivable from the being constraint) and hence inadequate for metaphysical debate.

But this assumption in (BC) is contentious. In axiomatizing necessitism for formal comparison with contingentism, Williamson (2010) explicitly restricts the being constraint to non-logical atomic predicates (and uses a predicate rather than the quantificational definition of existence) (p. 688). In his book, Williamson (2013) renders (BC) like Updike, but anticipates the objection, noting that the problem "assumes that $[Fx \to Fx]$ instantiates the being constraint... which it [does] if and only if $Fx \to Fx$ is a substitution instance of Fx" (p. 156). In fairness, Williamson continues to then criticize the objection, arguing that the unrestricted quantification in the being constraint is appropriate. However, opponents of his view who don't reject the being constraint entirely advocate for this restriction. For example, Stalnaker (2012) distinguishes between "properties" and "singularly propositional functions", with the being constraint crucially only applying to the former (139ff.).

Thus Updike's principal objection hinges on a particular view about properties and predication. Given this background, with Williamson arguing for the non-translatability of necessitism and contingentism, and the role the existence predicate typically plays in mediating the views (of possibilism and actualism), it is perhaps unsurprising that fixing a Williamsonian view trivializes certain metaphysical theses. With this issue in mind, we now turn our attention to Updike's various problems.

First, Updike claims that a move to second-order quantified modal W-logic will reveal more trivializations of metaphysics (and modal claims generally) like the above (p. 10). This criticism is targeted at the defense of excluding existence from the range of quantifier over properties. As before, if we adopt the restricted view of such quantifications, then the strategy of a defined universal like $\top(x)$ is still open. Our preferred resolution is to admit existence as a predicate, denying Updike's secondary point and accept that certain metaphysical claims can be resolved on the basis of logic alone (or at least go beyond the formal debate at hand).

It should also be noted that the discipline of higher-order modal metaphysics is still in its infancy. Bacon et al. (2016) show that the particular restrictions on quantification into sentence position one adopts have immediate implications for the necessitism-contingentism debate, and the higher-order Barcan formulas (also consequences of the choice of restriction) operate like comprehension schemas and hence can rig the debate. Williamson (2013) attacks contingentism on the basis that thorough contingentism (first- and higher-order) is untenable and a mixed view of first-order contingentism and higher-order necessitism is potentially philosophically suspect. Fritz and Goodman (2016) develop a coherent system for higher-order contingentism but note problems for the being constraint.

Second, Updike claims that the being constraint expressed in terms of an existence predicate still derives the necessity of being (p. 11). Our objection is as before: the claim only goes through if one allows $\top(x)$ as an instantiation of the predicate in the being constraint. Atomic predicates like $\mathcal{E}$ won't do the work either, only deriving a rote tautology.

Third, Updike claims that "some instances of the being constraint will fail" (p. 11). The argument again relies on unrestricted instantiation in the being constraint. Hence this observation motivates the restriction. Regarding Updike's commentary on comprehension, we again highlight Bacon et al.'s results entangling comprehension and modal ontology.

Fourth, Updike claims that "W-logician must take a stand on the actualist versus possibilist debate" (p. 13). While Updike's initial statement that an existence predicate co-extensive with what actually exists entails a slide into possibilism seems confused, our discussion of the restrictions on the being constraint indicate that use of an existence predicate might narrow one's options in modal metaphysics, even if it doesn't directly commit one to a particular side.

Fifth, Updike claims that the W-logician must employ actualist quantification and that the converse Barcan formula and the necessity of being follow from possibilist quantification. This latter observation is correct, well-known in $QML^{=}$ (see Proposition 4.9.10 in p. 113 of Fitting and Mendelsohn (1998)). But this is just a routine consequence of the choice of background constant domain semantics. Similarly, we do employ actualist quantification for the variable domain semantics, which is the only setting requiring the introduction of an existence predicate.

In sum, Updike's criticisms are by no means fatal for the W-logician. That is, for the contexts in which the W-actualist may want to admit an existence predicate, they face no technical obstacles, nor do they bear a philosophical burden that their standard counterpart avoids.

Appendix B

Addenda to Chapter Two

In this appendix, we present the technical results which were omitted from the main text of the chapter so as to not distract from the flow of the narrative.

B.1 Another Williamson Debate

Here we examine a more complicated example of Williamson's methodology. Consider a Williamson debate over the set-theoretic universe given by a base theory $T = ZFC$, and

$$T_0 = T + 0^\sharp \text{ exists} \tag{B.1}$$

$$T_1 = T + V = L \tag{B.2}$$

In this setting we take the neutral sentences (i.e. the extension of C) to be those whose quantifiers and predicates are restricted to constructible sets. In other words, neutral sentences are restricted to L. The set $0^\sharp$ is not constructible and hence incompatible with the axiom of constructibility (Jech, 2006, p. 339). Recall that $0^\sharp$ is a Δ^1_3-definable subset of ω (Schindler, 2014, p. 202). The set $0^\sharp$ is of interest to set theorists because it is implied by many large cardinal axioms; see Kanamori (2003, esp. p. 472) for a selection of these results.

We claim that Player 0 wins the Williamson debate, i.e. there is a Williamson-map for Player 1 but there is none for Player 0. For Player 1, we just relativize to L to obtain a Williamson-map, much like we did in the example of Infinity.

For Player 0, we show that there is no Williamson-map $(\cdot)^0$ such that for all $\varphi(x)$:

$$T_0 \vdash \forall x \in L(\varphi(x) \leftrightarrow \varphi^0(x)) \tag{B.3}$$

The argument proceeds by reductio. So let $\varphi(x) \equiv x \in 0^\sharp$ and suppose that the proposition failed, where we use the Δ^1_3 definition of membership in $0^\sharp$ (which in particular has no parameters). Then $T_0 \vdash \forall x \in L(\varphi(x) \leftrightarrow \varphi^0(x))$. In particular, $T_0 \vdash \forall x \in \omega(\varphi(x) \leftrightarrow \varphi^0(x))$. By the reflection theorem (Kunen, 1980, Theorem 7.5, p. 137), we have that $0^\sharp$ is first-order definable in L_β for some $\beta > \omega$, and hence $0^\sharp \in L_{\beta+1} \subset L$, a contradiction.

We have that Player 0 wins the Williamson debate, giving us evidence in support of T_0, which denies the axiom of constructibility. Note that we've had to allow parameters in the Williamson-map to get our result, as displayed in the quantifier in (B.3). These parameters are restricted to the concrete objects, however, so this should be unobjectionable.

B.2 Proofs

In this section we provide proofs for the results of the main text.

B.2.1 Toward Williamson's Original Result

Here we work toward a novel proof of Williamson's Theorem 2.2. We begin by reviewing the classic result of the non-first-orderizability of the Geach-Kaplan sentence.

Theorem 2.1. *There is a sentence of second-order logic which is non-first-orderizable. Indeed, the Geach-Kaplan sentence (GK) is non-first-orderizable.*

Proof. The proof proceeds by reductio. So suppose toward a contradiction that (GK) had a first-order equivalent. Note that (GK) contains the non-logical relation Axy ('admires'). Since definition of equivalence quantifies over all models and hence all possible interpretations of Axy, we look in particular at the class of models with zero and successor where Axy is interpreted arithmetically as $x = 0 \vee x = S(y)$. Substituting by this formula yields the arithmetized Geach-Kaplan sentence:

$$\exists X \forall x \forall y (\exists z X z \wedge ((X x \wedge (x = 0 \vee x = S(y))) \rightarrow (X y \wedge x \neq y))) \tag{GK'}$$

First, observe that the standard model of first-order Peano arithmetic is such that $\mathbb{N} \nvDash$ GK', i.e. that the arithmetized Geach-Kaplan is false on the standard model.[1] To see this, we must show that there is no witness to the existential. For suppose not, that there were such a witness X, which we can think of as a collection of numbers. We can separate the argument into two cases corresponding to whether zero belongs to the set X. After the second-order existential quantifier $\exists X$, there is a pair of universal quantifiers over numbers $\forall x \forall y$, so we need to verify that the subformula holds for all values of x and y.

[1]Throughout this paper, we will refer to models of true arithmetic – while the definition only requires first-order expressions, usually we'll talk of second-order models, as understood by the presence of second-order formulas in the scope of the turnstile. The second-order models are obtained by a natural extension in the standard semantics.

In the first case, with $0 \in X$, consider the choice of x and y with $x = y = 0$. The antecedent of the conditional is clearly satisfied, but the second conjunct of the consequent is made false by our choice of x and y, since it requires that they be nonidentical. In the second case, with $0 \notin X$, X must be nonempty for the first conjunct to be true, so let x be the smallest number in X, and let $y = x - 1$. Again the antecedent is satisfied since $x = S(y)$, but we can't have Xy for that would contradict our choice of x – x was supposed to be the smallest number in X but y is the predecessor of x.

So we have that the standard model is indeed not a model of the arithmetized Geach-Kaplan (GK'). Now, let us consider in turn an arbitrary non-standard model of true arithmetic, which exists by the Compactness Theorem of first-order logic. Let us show that this model is by contrast a model of the arithmetized Geach-Kaplan (GK'). To see this, we must provide a witness to the existential. Our witness X, will consist of all the non-standard elements in our non-standard model, that is, all of the elements which are not finitely reachable from zero via the successor operation. In order words, X is just the domain of our model with the standard part removed. In terms of the usual visualization of non-standard models of arithmetic as composed of $\mathbb{Z}$-chains, our witness X consists of all of the so-called $\mathbb{Z}$-chains.

The first conjunct of (GK') is satisfied since X is non-empty by construction. We must show the conditional holds, so assume the antecedent, that x is in X with x being either 0 or the successor of y for some choice of y. By our choice of X, $0 \notin X$, since 0 is in the standard part, so for the antecedent to be true the it must be that the other disjunct is satisfied, hence $x = S(y)$. The non-standard part of the model is infinite downward, meaning that no finite iterations of the predecessor operation (the opposite of successor) will lead back into the standard part. In other words, the $\mathbb{Z}$-chains are closed under predecessor. This means that $y \in X$ since y is the predecessor of a non-standard number x and hence non-standard itself (and X consisted of all the non-standard elements). Hence the consequent holds. Thus any non-standard model of arithmetic will be a model of the arithmetized Geach-Kaplan (GK').

So the truth-value of (GK') varies across second-order models of true arithmetic: it's false on the standard model but true on non-standard models.

Recall that we are assuming toward a contradiction that there were some first-order sentence φ equivalent to (GK'). By the definition of equivalence, which states that the two formulas must have the same truth-values on all models, we have that $\mathcal{N} \vDash$ GK' iff $\mathcal{N} \vDash \varphi$ for any $\mathcal{N}$. By our two previous results, it must be that $\mathbb{N} \nvDash \varphi$ (since $\mathbb{N} \nvDash$ GK') and $\mathcal{M} \vDash \varphi$ (since $\mathcal{M} \vDash$ GK'). But since we're working with models of true arithmetic, it must be that $\mathbb{N}$ and $\mathcal{M}$ agree on first-order sentences, in particular φ, yielding a contradiction. Thus (GK') and hence (GK) are non-first-orderizable. □

We have demonstrated that the move to second-order logic results in a genuine increase in expressive power. As we will see, Williamson exploits this phenomenon to achieve his result in the context of formal metaphysics. Now we provide a direct proof of Theorem 2.2.

Theorem 2.2. *There is a sentence φ of second-order quantified modal logic such that there is no neutral ψ with Aux[Nec] $\vDash \varphi \leftrightarrow \psi$. Indeed, (MGK) is such a sentence (Theorem 2.15 in Williamson (2010, p. 743)).*

Proof. We begin by defining a functor which takes a model $\mathcal{M}$ of PA^1 to an arithmetized model $\underset{\rangle}{\mathcal{M}}$ of Aux[Nec]. Let the set of worlds W for $\underset{\rangle}{\mathcal{M}}$ be $|\mathcal{M}|$, the domain of the original model. When we have a specific number x in mind, we will refer to it's corresponding world as w_x to avoid confusing the domains with the worlds themselves. Let the domain $d(w)$ at world w also be $|\mathcal{M}|$ for every world $w \in W$. This is because we want $\underset{\rangle}{\mathcal{M}}$ to model Aux[Nec], so the domain must be invariant across worlds, as forced by the first conjunct of the axiom. Since (MGK) contains two instances of non-logical vocabulary (the predicates R and C), we will specify their interpretation. Let the interpretation $i(w)(R) = \{\langle w, w\rangle, \langle w, w+1\rangle, \langle w+1, w\rangle, \langle w+1, w+1\rangle\}$, where we appeal to the successor function in $\mathcal{M}$ to define $w + 1$. Similarly, we let $i(w)(C) = \{w, w+1\}$ be the interpretation for C.

A first observation: on these models, every neutral formula is equivalent to a first-order formula of PA[1], i.e. for any second-order modal formula φ there is first-order non-modal arithmetic formula ψ such that $\mathcal{M} \vDash \varphi$ iff $\mathcal{M} \vDash \psi$. This is because neutral formulas are restricted to C (via the $(\cdot)^{\mathrm{Con}}$ map) and hence refer to at most two objects at a world, which can be captured by finite conjunctions in the language of arithmetic.

So, to show that (MGK) is not neutral, let $\mathbb{N}$ be the standard model of PA[1], and let $\mathcal{M}$ be any non-standard model. Our strategy will be to derive a contradiction from supposing the neutrality of MGK. To start, we will show that $\mathbb{N} \vDash \neg\mathrm{MGK}$ but $\mathcal{M} \vDash \mathrm{MGK}$. Our focus is on validity, i.e. truth at all worlds in a frame, so we do not make reference to a specific world. This follows a similar argument to establishing the non-first-orderizability of GK.

Let us consider the extension of the standard model $\mathbb{N}$ first. We show that there can be no such X witnessing the existential in (MGK). For suppose there were. Then there is x such that Xx since X must be non-empty by the first conjunct. From the interpretation of R it follows that $\Diamond R(x, x-1)$ and $\Diamond R(x, x+1)$, so we then have $X(x-1)$ and $X(x+1)$. This process can be repeated indefinitely, continually expanding membership in X. Since for any n we can show that Xn, it follows that $X = |\mathbb{N}|$, falsifying the second conjunct, which requires that X not include the entire domain. So $\mathbb{N} \vDash \neg\mathrm{MGK}$.

Now let's look at $\mathcal{M}$. As when we work with GK, let's take $X = |\mathcal{M}| \setminus |\mathbb{N}|$, the domain minus the standard part. Clearly the first two conjuncts are satisfied: X and its complement are non-empty. Now suppose $\Diamond Rxy$. This happens only when $|x - y| \leq 1$, i.e. when the two indexes differ by at most 1. So if x lies on a $\mathbb{Z}$-chain, then so must y, hence $Xx \rightarrow Xy$. Thus $\mathcal{M} \vDash \mathrm{MGK}$.

Now suppose toward a contradiction that (MGK) were neutral. Further, assume that $\mathcal{M}$ is a model of true arithmetic. Since $\mathcal{M} \vDash \mathrm{MGK}$, we know from the first observation that there is a sentence ψ of non-modal first-order arithmetic such that $\mathcal{M} \vDash \psi$. Since we're working with models of true arithmetic, they satisfy the same first-order formulas, so it follows that $\mathbb{N} \vDash \psi$. But it must then be that $\mathbb{N} \vDash$ MGK, a contradiction. Thus (MGK) is not neutral.

$\square$

B.2.2 Neutralizing MGK

Here we show that the intensional resources render neutral (MGK) by providing a neutral equivalent.

Proposition B.1. *Aux[Nec]* $\vDash MGK \leftrightarrow INT$.

Proof. Fix a model $\mathcal{M}$ of Aux[Nec]. First we show the left to right direction. So suppose $\mathcal{M} \vDash MGK$. Let X be a witness to (MGK). Define a function Δ on worlds by $w \mapsto X \cap int(C)(w)$. To show that Δ is a witness to (INT), we verify each conjunct of (INT).

First conjunct: Pick some $a \in X$. By Aux[Nec], there is w such that $a \in int(C)(w)$. It follows that $a \in \Delta(w)$, i.e $a \in \check{\Delta}$ at w.

Second conjunct: Now pick some $b \notin X$. Since $b \notin X$, $b \notin \check{\Delta}$ at w for any world w.

Third conjunct: Let worlds w_1 and w_2 and constants c and d be given such that $c \in int(C)(w_1)$ and $d \in int(C)(w_2)$, and suppose that $\Diamond Rcd$ and $\Diamond c \in \check{\Delta}$. By the latter supposition, there is some w_3 with $c \in \Delta(w_3) \subset X$. By (MGK) with the former supposition it follows that $d \in X$. We were given that $d \in int(C)(w_2)$, so we have that $d \in \Delta(w_2)$ and hence $\Diamond d \in \check{\Delta}$.

Now we show the right to left direction. So suppose $\mathcal{M} \vDash INT$ and let Δ be a witness. Define a subset of the domain $X = \bigcup_{w \in W} \Delta(w)$. To show that X is a witness to (MGK), we verify each conjunct of (MGK).

First conjunct: By (INT), there is a world w and object a such that $a \in \check{\Delta}$ at w and so X is non-empty.

Second conjunct: Similarly, there is a b such that for any world w we have that $b \notin \check{\Delta}$ at w and thus $b \notin X$.

Third conjunct: Now let c and d be given such that $\Diamond Rcd$ and Xc. Let w_1 be the world witnessing $\Diamond Rcd$. By Aux[Nec], it follows that $c, d \in int(C)(w_1)$. From the definition of X, there is some w_2 with $c \in \Delta(w_2)$. By (INT) we then have that $\Diamond d \in \check{\Delta}$, so $d \in X$. $\square$

B.2.3 Extending the Nec map

In this subsection, we provide an extension the $(\cdot)^{\text{Nec}}$ map to the new resources and show that the analogue to Theorem 2.2 fails in the expanded setting.

First, we extend the definition from Williamson (2010, p. 731) of the possibilification of a formula $(\cdot)^{\text{Nec}}$, maintaining the original clauses[2]:

$$\begin{aligned} (Xx)^{\text{Nec}} &= (Xx \wedge Cx) \\ (\exists X A)^{\text{Nec}} &= \exists \Delta ((A[X \triangleleft \Delta])^{\text{Nec}}) \\ (\exists \Delta A)^{\text{Nec}} &= \exists \Delta (A)^{\text{Nec}} \end{aligned} \tag{Nec+}$$

Here, $A[X \triangleleft \Delta]$ indicates the result of substituting $\Diamond x \in \check{\Delta}$ for each occurrence of Xx in A.

Next we recall from Williamson (2010, pp.731-732) the $(\cdot)^{\text{Nec}}$ operation on models $M = \langle W, D, dom, int \rangle$: $M^{\text{Nec}} = \langle W, D, dom^N, int^C \rangle$, where $dom^N(w) = \bigcup_{w' \in W} dom(w') \cap int(C)(w')$ and $int^C(F)(w) = int(F)(w) \cap int(C)(w)^n$.

[2] I.e. those for the non-logical predicates, identity, propositional connectives, diamond, Vlach operators and first-order quantifier.

We proceed by formulating two initial results. The first is an analogue of Proposition 1.13 from Williamson (2010, p. 732). The motivation for the (α) and (β) conditions should hopefully be clear given these constructions in the proof of Proposition B.1.

Theorem B.2. *Fix a model M. For all formulas $A(X_1, \dots, X_n)$ with all higher-order variables displayed, for all assignments a with*

$$a(X_i) \subset \bigcup_{w \in W} dom(w) \cap int(C)(w) \tag{$*$}$$

and

$$\begin{aligned} & a(\Delta_i)(w) = a(X_i) \cap int(C)(w) && (\alpha) \\ or \quad & a(X_i) = \bigcup_{w \in W} a(\Delta_i)(w) && (\beta) \end{aligned}$$

we have

$$\begin{aligned} & M, w, s, \bar{a} \vDash (A[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n])^{Nec} \\ & \textit{iff } M^{Nec}, w, s, \bar{a} \vDash A(X_1, \dots, X_n) \end{aligned}$$

Proof. This proceeds by induction on the complexity of A.

The base case has two subcases. In the first case[3], suppose $A(X_1, \dots, X_n) \equiv X_i x$. Let a sequence of worlds s, world w and variable assignment a satisfying the above conditions be given. We differentiate the cases where a satisfies (α) and those where a satisfies (β) and prove each direction separately.

[3]The argument for free second-order variables of higher arities is exactly parallel, and for the sake of readability we restrict to $X_i x$. Incorporating higher arities involves a generalization of the intensional operators to tuples of concrete objects, which can be implemented straightforwardly.

$(\alpha), \Rightarrow$: Suppose $M, w, s, a \vDash (A[X_1 \triangleleft \Delta_1, \ldots, X_n \triangleleft \Delta_n])^{\text{Nec}}$. Then $M, w, s, a \vDash \Diamond(x \in \breve{\Delta}_i \wedge Cx)$. It follows that there is w' with $M, w', s, a \vDash x \in \breve{\Delta} \wedge Cx$. Then $a(x) \in a(X_i) \cap int(C)(w')$ by (α). Then $a(x) \in a(X_i) \subset dom^N(w)$ by $(*)$, so $M^{\text{Nec}}, w, s, a \vDash X_i x$.

$(\alpha), \Leftarrow$: Suppose $M^{\text{Nec}}, s, w, a \vDash X_i x$. Then $a(x) \in a(X_i) \subset dom^N(w) = \bigcup_w dom(w) \cap int(C)(w)$. So choose w' witnessing this with $a(x) \in dom(w') \cap int(C)(w')$. Then $M, w', s, a \vDash x \in \breve{\Delta}_i \wedge Cx$ by (α). Then $M, w, s, a \vDash \Diamond(x \in \breve{\Delta}_i \wedge Cx)$. So $M, w, s, a \vDash (A[X_1 \triangleleft \Delta_1, \ldots, X_n \triangleleft \Delta_n])^{\text{Nec}}$.

$(\beta), \Rightarrow$: Suppose $M, w, s, a \vDash (A[X_1 \triangleleft \Delta_1, \ldots, X_n \triangleleft \Delta_n])^{\text{Nec}}$. Then $M, w, s, a \vDash \Diamond(x \in \breve{\Delta}_i \wedge Cx)$. It follows there is w' with $M, w', s, a \vDash x \in \breve{\Delta} \wedge Cx$. Then $a(x) \in a(\Delta_i)(w') \subset a(X_i) \subset dom^N(w)$ by (β) and then by $(*)$, so $M, w, s, a \vDash Xx$.

$(\beta), \Leftarrow$: Suppose $M^{\text{Nec}}, s, w, a \vDash X_i x$. We have $a(x) \in a(X_i)$ so there is w' with $a(x) \in a(\Delta_i)(w')$ by (β). Then $M, w', s, a \vDash x \in \breve{\Delta}_i \wedge Cx$ because $\Delta_i(w') \subset int(C)(w')$. Then $M, w, s, a \vDash \Diamond(x \in \breve{\Delta}_i \wedge Cx)$. So $M, w, s, a \vDash (A[X_1 \triangleleft \Delta_1, \ldots, X_n \triangleleft \Delta_n])^{\text{Nec}}$.

The case where $A(X_1, \ldots, X_n) \equiv Fx$ a predicate is unchanged from Williamson (2010, Proposition 1.13, p. 732).

For the inductive step, the cases for the propositional connectives follow straightforwardly from the induction hypothesis. There remain three subcases to verify.

First, suppose $A(X_1, \dots, X_n) \equiv \exists x A_0(X_1, \dots, X_n)$. Then we have the following chain of equivalences:

$$
\begin{aligned}
& M, w, s, a \vDash (\exists x A_0[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n])^{\text{Nec}} \\
\text{iff } & M, w, s, a \vDash \uparrow \Diamond \exists x (Cx \wedge \downarrow (A_0[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n])^{\text{Nec}}) \\
\text{iff } & M, w, s\frown w, a \vDash \Diamond \exists x (Cx \wedge \downarrow (A_0[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n])^{\text{Nec}}) \\
\text{iff } & \exists w' \in W: \ M, w', s\frown w, a \vDash \exists x (Cx \wedge \downarrow (A_0[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n])^{\text{Nec}}) \\
\text{iff } & \exists w' \in W, \exists x\text{-variant } b \text{ of } a: \ M, w', s\frown w, b \vDash Cx \wedge \downarrow (A_0[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n])^{\text{Nec}} \\
\text{iff } & \exists w' \in W, \exists x\text{-variant } b \text{ of } a: \ M, w', s\frown w, b \vDash Cx \\
& \qquad \text{and } M, w', s\frown w, b \vDash \downarrow (A_0[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n])^{\text{Nec}} \\
\text{iff } & \exists w' \in W, \exists x\text{-variant } b \text{ of } a: \ M, w', s\frown w, b \vDash Cx \\
& \qquad \text{and } M, w, s, b \vDash (A_0[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n])^{\text{Nec}} \\
\text{iff } & \exists w' \in W, \exists x\text{-variant } b \text{ of } a: \ M, w', s\frown w, b \vDash Cx \\
& \qquad \text{and } M^{\text{Nec}}, w, s, b \vDash A_0 \text{ (by I.H.)} \\
\text{iff } & \exists w' \in W: \ M, w', s\frown w, a \vDash \exists x Cx \\
& \qquad \text{and } M^{\text{Nec}}, w, s, a \vDash \exists x A_0
\end{aligned}
$$

The first conjunct of the last line is vacuous (for it to be false would require the domain to be empty), so we've concluded this step up the induction.

Second, suppose $A(X_1, \dots, X_n) \equiv \exists X A_0(X_1, \dots, X_n, X)$. Following the definition of the $(\cdot)^{\text{Nec}}$ map $M, w, s, a \vDash (A[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n])^{\text{Nec}}$ iff $M, w, s, a \vDash \exists \Delta (A_0[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n, X \triangleleft \Delta])^{\text{Nec}}$. We prove each direction separately.

For the left-to-right direction, suppose $M, w, s, a \vDash \exists \Delta (A_0[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n, X \triangleleft \Delta])^{\text{Nec}}$. Choose a Δ-variant b such that $M, w, s, b \vDash (A_0[X_1 \triangleleft \Delta_1, \dots, X_n \triangleleft \Delta_n, X \triangleleft \Delta])^{\text{Nec}}$. Note that X does not appear in formula $(A_0[X_1 \triangleleft \Delta_1, \ldots, X_n \triangleleft \Delta_n, X \triangleleft \Delta])^{\text{Nec}}$ since all instances of it are replaced. Then we can assume without loss of generality that $b(X) = \bigcup_{w \in W} b(\Delta(w))$. Then by the induction hypothesis (note that b satisfies $(*)$ and (β)), we have that $M^{\text{Nec}}, w, s, b \vDash A_0(X_1, \ldots, X_n, X)$. Then of course $M^{\text{Nec}}, w, s, a \vDash \exists X A_0(X_1, \ldots, X_n, X)$.

For the right-to-left direction, suppose $M^{\text{Nec}}, w, s, a \vDash \exists X A_0(X_1, \ldots, X_n, X)$. Choose an X-variant b such that $M^{\text{Nec}}, w, s, b \vDash A_0(X_1, \ldots, X_n, X)$. Since Δ does not appear in A_0, we can assume without loss of generality that $b(\Delta)(w) = b(X) \cap int(C)(w)$. We find that b satisfies $(*)$ and (α), so by the induction hypothesis we have $M, w, s, b \vDash (A_0[X_1 \triangleleft \Delta_1, \ldots, X_n \triangleleft \Delta_n, X \triangleleft \Delta])^{\text{Nec}}$. Then it follows that $M, w, s, a \vDash \exists\Delta((A_0[X_1 \triangleleft \Delta_1, \ldots, X_n \triangleleft \Delta_n, X \triangleleft \Delta])^{\text{Nec}})$, which is what we wanted to show.

Third, suppose $A \equiv \exists \Delta A_0(X_1, \ldots, X_n, \Delta)$. Following the definition of the map $M, w, s, a \vDash (A(X_1, \ldots, X_n))^{\text{Nec}}$ iff $M, w, s, a \vDash \exists\Delta((A_0[X_1 \triangleleft \Delta_1, \ldots, X_n \triangleleft \Delta_n])^{\text{Nec}})$, iff there is a Δ-variant b with $M, w, s, b \vDash (A_0[X_1 \triangleleft \Delta_1, \ldots, X_n \triangleleft \Delta_n](\Delta))^{\text{Nec}}$, iff there is a Δ-variant b with $M^{\text{Nec}}, w, s, b \vDash A_0(X_1, \ldots, X_n, \Delta)$ via induction hypothesis, iff $M^{\text{Nec}}, w, s, a \vDash \exists\Delta(A_0(X_1, \ldots, X_n))$. In this, we are appealing to the fact that M and M^{Nec} agree on the range of the quantifiers over intensions Δ, Γ, etc. $\square$

The following result was proven in Proposition 1.20 of Williamson (2010, p. 734).

Lemma B.3. *If $M, w, s, a \vDash$ Aux[Nec], then $M^{Nec} = M$.*

We now establish that the analogue Theorem 2.2 does not hold in the expanded setting with the intensional resources.

Corollary B.4. *For any sentence A, Aux[Nec] $\vDash (A)^{Nec} \leftrightarrow A$.*

Proof. Let a model M of Aux[Nec] and sentence A be given. Then $M, w, s, a \vDash (A)^{\text{Nec}}$ iff $M^{\text{Nec}}, w, s, a \vDash A$ by Theorem B.2, iff $M, w, s, a \vDash A$ by Lemma B.3. Thus $M, w, s, a \vDash (A)^{\text{Nec}} \leftrightarrow A$. $\square$

B.2.4 Fritz-neutrality

Here we provide the proof which evidences the claim that the intensional resources are neutral in Fritz's model-theoretic sense.

Proposition B.5. *Let $M, M' \in \mathsf{P}$ be two models which concrete-coincide. Then for all formulas A, for all worlds w, for all variable assignments g and g' which agree on the free variables of A, for all (possibly empty) subsets of second-order variables $\{X_1, \dots, X_n\}$ among the free variables of A, and for all $\{\Delta_1, \dots, \Delta_n\}$ not among the free variables of $(A)^{Nec}$ such that $g(\Delta_i) = g'(\Delta_i)$ for $i \in \{1, \dots, n\}$, we have*

$$M, w, g \vDash A[X_1 \lhd \Delta_1, \dots, X_n \lhd \Delta_n]^{Nec} \textit{ iff } M', w, g' \vDash A[X_1 \lhd \Delta_1, \dots, X_n \lhd \Delta_n]^{Nec}$$

We are interested in the case where A is a sentence, for the proposition yields that $(A)^{\mathrm{Nec}}$ is Fritz-neutral. However, the proof requires that we work with formulas generally.

Proof. The proof proceeds by induction on the complexity of A. Fritz (2013, Proposition 13, p. 666) showed that his criterion was equivalent to Williamson's in the first order case, so we consider just the second-order extensions.

For the base case, suppose $A \equiv Xx$. There are two subcases corresponding to whether X is among the set of free variables mentioned in the statement of the proposition: we must show (i) $M, w, g \vDash (A)^{\mathrm{Nec}}$ iff $M', w, g' \vDash (A)^{\mathrm{Nec}}$ and (ii) $M, w, g \vDash (A[X \lhd \Delta_1])^{\mathrm{Nec}}$ iff $M', w, g' \vDash (A[X \lhd \Delta_1])^{\mathrm{Nec}}$.

For (i), $M, w, g \vDash (A)^{\mathrm{Nec}}$ iff $M, w, g \vDash Xx \wedge Cx$, iff $M', w, g' \vDash Xx \wedge Cx$ (since g and g' agree on the free variables X and x, and M and M' concrete-coincide), iff $M', w, g' \vDash (A)^{\mathrm{Nec}}$.

For (ii), $M, w, g \vDash (A[X \lhd \Delta_1])^{\mathrm{Nec}}$ iff $M, w, g \vDash (\Diamond x \in \breve{\Delta}_1)^{\mathrm{Nec}}$, iff $M, w, g \vDash \Diamond(x \in \breve{\Delta}_1 \wedge Cx)$, iff there is $w' \in W$ such that $M, w', g \vDash x \in \breve{\Delta}_1 \wedge Cx$, iff there is $w' \in W$ such that $M', w', g' \vDash$

$x \in \widetilde{\Delta_1} \wedge Cx$ (since g and g' agree on the free variables Δ_1 and x, and M and M' concrete-coincide), iff $M', w, g' \vDash \Diamond(x \in \widetilde{\Delta_1} \wedge Cx)$, iff $M', w, g' \vDash (A[X \triangleleft \Delta_1])^{\mathrm{Nec}}$.

For the inductive step, we have two cases to consider.

First, suppose that $A \equiv \exists X A_0$. Then $M, w, g \vDash (\exists X A_0)^{\mathrm{Nec}}$ iff $M, w, g \vDash \exists\Delta((A_0[X \triangleleft \Delta])^{\mathrm{Nec}})$, iff there is a Δ-variant h of g such that $M, w, h \vDash (A_0[X \triangleleft \Delta])^{\mathrm{Nec}}$, iff there is a Δ-variant h' of g' such that $M', w, h' \vDash (A_0[X \triangleleft \Delta])^{\mathrm{Nec}}$ (by induction hypothesis, for if g and g' agree on the free variables of A, then the Δ-variant h' of g' with $h'(\Delta) = h(\Delta)$ agrees with h on the free variables of $(A_0[X \triangleleft \Delta])^{\mathrm{Nec}}$), iff $M', w, g' \vDash \exists\Delta((A_0[X \triangleleft \Delta])^{\mathrm{Nec}})$, iff $M', w, g' \vDash (\exists X A_0)^{\mathrm{Nec}}$.

Second, suppose that $A \equiv \exists\Delta A_0$. Then $M, w, g \vDash (\exists\Delta A_0)^{\mathrm{Nec}}$ iff $M, w, g \vDash \exists\Delta((A_0)^{\mathrm{Nec}})$, iff there is a Δ-variant h of g such that $M, w, h \vDash (A_0)^{\mathrm{Nec}}$, iff there is a Δ-variant h' of g' such that $M', w, h' \vDash (A_0)^{\mathrm{Nec}}$ (by induction hypothesis using the same reasoning above), iff $M', w, g' \vDash \exists\Delta((A_0)^{\mathrm{Nec}})$, iff $M', w, g' \vDash (\exists\Delta A_0)^{\mathrm{Nec}}$. □