GROUNDEDNESS

JÖNNE SPECK

Its Logic and Metaphysics
2013 – first draft

ACKNOWLEDGMENTS

This work was supported by the *European Research Council* and the University of Oslo.

I am indebted to my supervisor Øystein Linnebo. His constructive criticism has shaped this material significantly.

Chapter 3: I thank Casper Storm Hansen, Sam Roberts and participants at seminars in Bristol and Oslo.

Section 7.1: I thank Antje Rumberg for her patiently sharing with me her extensive knowledge of Bolzano's theory of logic.

Chapter 10: I thank Jon Litland for many inspiring discussions.

Chapter 11: I am very grateful to Albert Visser and Philip Welch for corrections on earlier drafts as well as many inspiring questions.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LISTINGS

ACRONYMS

1 A GENERAL THEORY OF GROUNDEDNESS

Notions of groundedness have figured prominently in the literature on the semantic paradoxes.[1] However, I have in mind a more general conception. It does not only apply to sentences, propositions, or to the truth-values of sentences or propositions. Whenever we are given some things we may ask whether they are grounded; more precisely, whether they are grounded in some designated collection G.

What does it mean for xx to be grounded in yy? Recently, Forster has proposed a generalized iterative conception of set Forster [2008].[2] In this chapter, I further generalize his idea. Thus, a general theory of groundedness emerges that not only subsumes existent accounts, but also systematizes their connections and underlying motivation.

Forster's generalized iterative conception is closely linked to what he calls *recursive datatypes* [2008, p. 99]. Outside of computer science these are better known as *inductive definitions*. Mathematically, the general theory of groundedness will be included in the theory of inductive definitions. Philosophically, there is much more to groundedness. However, developing the philosophical side of the concept I postpone to a later chapter.

Since one intended application of the following is to the universe of sets itself, I will work within *plural* logic.[3] For its primitive of a thing *being among* some things I adopt Burgess' notation '$\propto$'.[4]

For simplicity, I will use the singular locution 'plurality' to refer to some things. I will also use $xx \sqsubseteq yy$ as short for $\forall z(z \propto xx \rightarrow z \propto yy)$.

Finally, I will assume that my plural metalanguage has a plural term forming operator that I denote by the comma sign. Thus, x, y, z is a plural term, as is xx, y.

Forster formulates his iterative conception in terms of *constructor* functions. I will not adopt this terminology. In my study, issues from

1 Herzberger [1970]; Kripke [1975]; Yablo [1982]; McCarthy [1988]; Maudlin [2004]; Leitgeb [2005]

2 Forster uses his generalized iterative conception to argue for the legitimacy of certain non-standard set theories. As I will explain in chapter 5 below, I do not think his argument is conclusive.

3 This choice is for merely practical reasons, and nothing hinges on it. An other framework would do, too, as long as it allows for foundations of set theory. One such alternative framework is Fregean higher-order logic, another may be category theory.

4 Burgess motivates this choice as follows.

> Much as the symbol used in set theory for 'element' is a stylized epsilon '$\in$', the symbol used here for 'is among' is a stylized alpha '$\propto$'. [Burgess, 2004, p. 197]

Using '$\propto$', I deviate from the mainstream that uses '$\prec$' to denote the relation of something being among some others. My reason for deviating is that I will have to use the symbol '$\prec$' for another notion (definition 6).
Further, I acknowledge that there are good reasons to prefer 'is one of' as the informal paraphrasing of the primitive relation Ben-Yami [2009]. However, for terminological clarity I will stick to 'among'. At any rate, doing so I am not committed to any claim about plural logic as a suitable regimentation of natural language plural locutions.

the philosophy of maths will play an important role. In this context, 'constructor' is not a sufficiently neutral word. Therefore, I use the term 'generator' which I hope not to provoke as many philosophical associations. It is certainly intended as a neutral label within a general framework.

The concept of a *generator* is the primitive of my theory of groundedness. For intuition, think of a generator as a recipe by which x is obtained from yy. Formally, a generator Φ is a many-one relation:

$$yy\Phi x$$

Examples are abound. The *formation rules* of a formal language are a generator: the disjunctive formula $\phi \vee \psi$ is generated from the formulae ϕ and ψ. Other examples are the introduction rules of a formal proof system. In propositional logic a theorem $\phi \wedge \psi$ is generated from theorems ϕ and ψ. My interest, however, will be in generators that can be viewed as capturing some of the naïve principles of comprehension or truth, that in the previous chapter we have found to lead to paradox. In particular, the set of the xx is generated from them, and the truth that ϕ is true is generated from the truth that ϕ.

Forster's constructors are functions. I lift this restriction and work with relations. Doing so has the advantage that it dispenses with the need for adding "destructors" to the general framework, functions from something to those things from which it is constructed. Instead, I can take the inverse Φ^{-1} of a generator Φ. For example, if $\phi \vee \psi$ is generated from its disjunct, then they stand in the relation Φ^{-1} to it. I will return to this.

Further, unlike Forster I concentrate on the whole of the ways in which we generate objects from pluralities. Let me give an example.

Example 1. Consider the language of propositional logic based on propositional letters p, q, r, with decorations. From these atomic sentences we *generate* complex sentences using the following formation rules:

$$P_1 \; \frac{\phi}{(\neg\phi)} \qquad \frac{\phi \qquad \psi}{(\phi \vee \psi)} \; P_2$$

$$P_3 \; \frac{\phi \qquad \psi}{(\phi \wedge \psi)} \qquad \frac{\phi \qquad \psi}{(\phi \rightarrow \psi)} \; P_4$$

Presently, what matters are the rules P_1 to P_4 taken together, and it is them together what I will call the generator P. This generalizes Forster's terminology, who would speak of four constructors P_1 to P_4.

In general, if the domain at hand form a set then I will often use set-theoretic resources to speak of them and the generator Φ. This will render the presentation more readable and familiar. For example, as the sentences of propositional logic form a set, I represent the generator P as the union of the following relations.

$$
\begin{aligned}
P_1 &:= \{\langle X, \zeta\rangle : X = \{\phi\}, \zeta = (\neg\phi)\}\\
P_2 &:= \{\langle X, \zeta\rangle : X = \{\phi, \psi\}, \zeta = (\phi \wedge \psi)\}\\
P_3 &:= \{\langle X, \zeta\rangle : X = \{\phi, \psi\}, \zeta = (\phi \vee \psi)\}\\
P_4 &:= \{\langle X, \zeta\rangle : X = \{\phi, \psi\}, \zeta = (\phi \rightarrow \psi)\}
\end{aligned}
$$

However, I emphasize that I do not intend to reduce the notion of a generator to that of a relation, understood set-theoretically, plurally or by some other means. A generator Φ is a way of obtaining an x from some yy, a rule how to move from the yy to x. It is not a collection of pairs $\langle yy, x\rangle$.[5] Rather, Φ is the intension corresponding to such an extensional characterization.

Given a generator Φ, it may be the case that Φ allows for the construction of one and the same object x from distinct pluralities yy, zz. Below I will give reason to focus on cases in which this is ruled out. I will consider generators Φ that are *deterministic*, in the following sense.[6]

Definition 1 (Deterministic generators)**.** We call a generator Φ *deterministic* iff for every x and all pluralities yy, zz

If $yy\Phi x$ and $zz\Phi x$ then the yy are the zz.

Note that the generator **P** from example 1 is deterministic.

Deterministic or not, a generator Φ and some things xx is all that is needed to formulate my general concept of groundedness. I will give two ways of characterizing some y as grounded in the xx through Φ. They are equivalent, but formalize intuitively distinct ideas. Hence, it will prove useful to have available both of these two characterizations of groundedness.

Given a generator Φ, we can ask two prima facie distinct questions. Firstly, we may ask what we may generate from some things. Secondly, we may ask what something is generated from.

The first definition will identify y as grounded in xx if it is arrived at by iterated Φ-generation, starting from xx. Accordingly, I will refer to it as the *upwards* characterization of groundedness. The second definition will call y grounded in xx if tracing down what y is generated

5 Officially, '$\langle xx, y\rangle$' is short for the pairs $\langle x, y\rangle$. There are several ways of understanding pairs $\langle x, y\rangle$ in the present, plural setting. For one, we may assume *super-plurals* in terms of which we can understand a relation holding between one object x and another object y. For another, we may use the machinery of Lewis and Hazen [Lewis, 1991; Hazen, 1997, 2000]. Either way, there is no need to ascent to super-duper-plurals respectively higher-order pairs.

6 To my knowledge, this terminology is due to Peter Aczel [1977, p. 744].

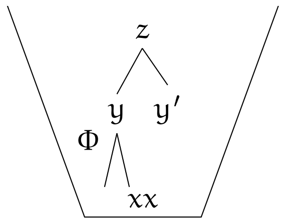

Figure 1: Groundedness

from, we end with them. I will speak of this second characterization of groundedness as *downwards*.

The first author to clearly view the philosophical significance of this distinction was Stephen Yablo 1982. Much of the following may be viewed as a further generalization of Yablo's general, formal theory of groundedness. I will state and explain these connections as they arise.

1.1 UPWARDS: GENERATION

Assume we are given some gg. Then, let us call y grounded in the gg through Φ if y is among the gg, if it stands in the relation Φ to ("is generated from") some $zz \sqsubseteq gg$, or if y is generated from some zz each of which is already grounded in the zz. This idea is visualized well by drawing, as in figure 1, a funnel whose base represents the xx, and every point in its area represents something grounded in them.

To render precise this idea in the present, very general context, I need to clarify what it means to *iterate* generation. Given a well-ordering, we can define the *stages* of an iterated generation through Φ along the well-ordering. Fortunately, for some things ww to well-order some other things yy can be expressed in our present, plural setting [Shapiro, 1991, p. 106]. Let the ww be a well-ordering of the yy. In cases where yy form a set I will, for simplicity, work with its order-type, the ordinal α. Similarly, for readability I will write '$v <_{ww} w$' to say that v precedes w in the ordering ww, and use $w+1$ for the ww-successor of w.

We wish to formalize the iteration of Φ, starting from the gg. Let 0_{ww} be the ww-least thing. Then, we encode the first stage of our iteration of Φ as pairs $\langle 0_w w, g\rangle$ where g is among gg. Given some w that is among the ww, the $w+1$st stage of our iteration of Φ is encoded by pairs $\langle w, x\rangle$, where x is among the wth stage, or there are some xx among these and $xx\Phi x$. Generalizing standard notation from the theory of inductive definitions, I will denote the wth stage by '$I^w_\Phi(gg)$'. For w limit among the ww, let I^w_Φ be all the $I^{w'}_\Phi$, for w' ww-earlier than w, taken together.

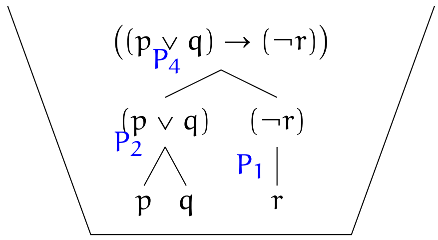

Figure 2: The Grounding Cone of Propositional Logic

Definition 2 (Groundedness)**.** Let Φ be a generator and let gg be some things.

Some y is grounded in gg through Φ ('Φ-grounded in gg', in symbols: $gg <_\Phi y$) iff there is some well-ordering ww and some $w \propto ww$ and y is among the $I^w_\Phi(gg)$.

A mundane example will show just how common groundedness is.

Example 2. The sentences of propositional logic are grounded in propositional letters $p_0, p_1, \ldots$ through the generator P from example 1: they are P-grounded in the $p_0, p_1, \ldots$. See figure 2.

If we have two generators Φ and Φ', they can be combined, giving rise to a more inclusive notion of Φ-Φ'-groundedness. This is done as follows. Let the yy be some things and let Φ and Φ' be generators on them. Now we may obtain things from the yy either through Φ or through Φ': but this is a new way of generating things, a combined generator Φ-Φ'. Think of Φ as the rule to infer y if the xx are so and so, and of Φ' as the rule to infer y if the xx are such and such. The combined generator Φ-Φ' then is the rule which allows us to infer y from the xx if they are so and so *or* such and such.

Example 3. The sentences of propositional *modal* logic are grounded in the propositional letters through the combination of P with the generator M, given by:

$$M_1 \, \frac{\phi}{\Diamond\phi} \qquad \frac{\phi}{\Box\phi} \, M_2$$

We can define, relative to Φ, an operator Γ_Φ that takes some things xx and outputs exactly those yy each of which is Φ-generated from some zz among the xx. Formally,

Definition 3.

$$y \propto \Gamma_\Phi(xx) \Leftrightarrow \exists zz \sqsubseteq xx(zz\Phi y)$$

This operator Γ_Φ allows for the following, useful re-characterization of Φ-groundedness. Some x is Φ-grounded in the yy, we may say, if starting from the yy, and iterating this operator Γ_Φ, we eventually find x in its output. This is equivalent to saying that x is Φ-grounded in the yy iff x is in the least collection containing the yy and closed under Γ_Φ, in other words the least fixed point of Γ_Φ that contains the yy.

1.2 DOWNWARDS: PRIORITY

Given a generator Φ, we define a relation $\prec_\Phi$. $\prec_\Phi$ is meant to express the *priority* of one object over another relative to the generator Φ

Definition 4 (Immediate Partial Priority)**.** Let Φ be a generator. We say that x is immediately, partially prior to y relative to Φ iff there x is among some things from which y is generated through Φ.

For example, the propositional letters p, q are each prior to their disjunction $(p \vee q)$ relative to the generator **P**.

$$p \prec_P (p \vee q)$$
$$q \prec_P (p \vee q)$$

In general, however, $\prec_\Phi$ is not ensured to be irreflexive: some generators Φ allow y to be generated from some xx such that y itself is among the xx.

Of course, there is a corresponding concept of *complete* priority. It is, however, best developed via the following concept of a grounded object's *priority tree*.

Definition 5 (Priority Tree)**.** Let Φ be a generator, and let x and gg be some things. Let $\mathcal{T}$ be a (finite) tree. $\mathcal{T}$ is a *priority* tree of x through Φ and with respect to the gg (a 'Φ, gg-tree' of x) iff (1) x is the root of $\mathcal{T}$, (2) every node z of $\mathcal{T}$ can be generated from the nodes immediately below z by Φ and (3) z is a $\mathcal{T}$-leaf iff it is among the gg.

Example 4. The complex sentences of propositional logic are grounded in the atomic sentences, through the generator P (1). Figure 3 shows a $P, \{p, q\}$-priority tree of $p \wedge (q \to \neg p) \to \neg q$.

Why have I introduced this machinery? The reason is that it allows for a neat re-description of groundedness.

Proposition 1 (Aczel 1977, Yablo 1982)**.** *Let Φ be some generator and let gg be some things. Then x is Φ-grounded in gg just in case x has a Φ, gg-priority tree.*

Proof. x is Φ-grounded just in case for some well-ordering ww x is among $I^w_\Phi(gg)$, for some $w \propto ww$.

The proposition therefore follows directly from lemma 1 below. □

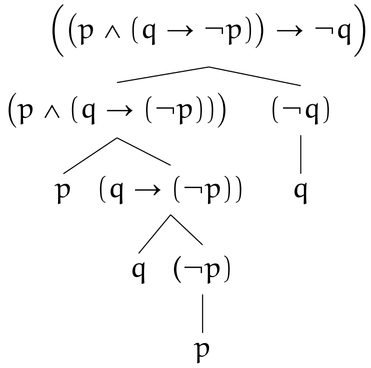

Figure 3: A priority tree

Lemma 1. *Let Φ be some generator and* gg *some things. Let w be any point in the well-ordering ww. Then $x \propto I^{w}_{\Phi}(gg)$ just in case x has a Φ, gg-tree of* length $\leqslant w$.

Proof. The claim is proved by an induction on the well-ordering ww.

(Induction base) ($\sqsubseteq$) Assume $x \propto I^{0}_{\Phi}(gg) = gg$. Then $\langle x \rangle$ is itself a priority tree as required.

($\sqsupseteq$) Let $\mathcal{T}$ be a Φ, gg-priority tree of x of length 0. Hence it must be $\langle x \rangle$ and $x \propto gg$, hence $x \propto I^{0}_{\Phi}$.

(Induction step)

($\sqsubseteq$) Assume $x \propto I^{w}_{\Phi} = \Gamma_{\Phi}(\bigcup_{v <_{ww} w} I^{v}_{\Phi})$. By our induction hypothesis, however, $z \propto \bigcup_{v <_{ww} w} I^{v}_{\Phi}$ just in case z has a Φ, gg-tree $\mathcal{T}^{z}$ of length $< w$. Now we take all these trees and put x on top. We obtain a Φ, gg-tree of length w, as desired.

($\sqsupseteq$) Assume x has a Φ, gg-tree $\mathcal{T}$ of length $\leqslant_{ww} w$. Let z be among the $\mathcal{T}$-nodes zz immediately below x. Let $\mathcal{T}^{z}$ be the sequences $\langle z, \overrightarrow{u_n} \rangle$ for $n \geqslant 0$ and $\langle x, z, \overrightarrow{u_n} \rangle \propto T$. We note that each $\mathcal{T}^{z}$ is a Φ, gg-tree of z, of length less than α. By our induction hypothesis, therefore, $z \propto I^{v}_{\Phi}$, $v <_{ww} w$. Since by assumption, Φ allows us to infer x from the zz, we have that $x \propto \Gamma_{\Phi}(\bigcup_{v<w} I^{v}_{\Phi}) = I^{w}_{\Phi}$, as desired. $\square$

Reflecting on the structure of priority trees leads naturally to the concept of *mediate priority*, or *dependence*. Let us abstract slightly and speak of y having a Φ-priority tree simpliciter if for *some* gg, y has a Φ, gg-priority tree.

Definition 6 (Dependence)**.** Let us say that x *depends* on z through Φ (in symbols: '$z \prec_{\Phi} x$') iff z is a node in a Φ-priority tree of x.

Lemma 2. *Given Φ, $\prec_{\Phi}$ is well-founded on the nodes of every Φ-priority tree.*

The notion of a priority tree becomes even more useful if we assume Φ to be deterministic in the sense of definition 1. The reason is that for *deterministic* generators Φ and a fixed plurality gg, every object x has a *unique* (up to isomorphism) Φ, gg-tree.

Lemma 3. *Let* Φ *be a* deterministic *generator and let* x *be some object. For every* gg, x *has exactly one* Φ, gg*-tree (up to isormophism).*

Proof. Let $\mathcal{T}$ and $\mathcal{T}'$ be Φ, gg-priority trees of x. I show that $\mathcal{T} = \mathcal{T}'$ by induction on the (finite) height of $\mathcal{T}$.

If $\mathcal{T}$ is the single node x, then by clause (3) of the definition of priority trees, x must be among the gg. Hence, for $\mathcal{T}'$ to be a Φ, gg-tree of x it must firstly have x as its root, and secondly have x as a leaf. $\mathcal{T}'$ must therefore be the single node x, and thus identical to $\mathcal{T}$.

Now let $\mathcal{T}$ be of height $n+1$, and let $\mathcal{T} \restriction n$ be $\mathcal{T}$'s largest subtree of height n ($\mathcal{T} \restriction n$ is $\mathcal{T}$ without its leaves). Since $\mathcal{T}$ is a Φ, gg-tree, we know that every leaf of $\mathcal{T} \restriction n$ is generated from the leaves of $\mathcal{T}$ by Φ. The same holds for $\mathcal{T}' \restriction n$ and $\mathcal{T}'$.

Now assume for contradiction that $\mathcal{T} \neq \mathcal{T}'$. By our induction hypothesis we know that $\mathcal{T} \restriction n = \mathcal{T}' \restriction n$. So, $\mathcal{T}$ and $\mathcal{T}'$ must differ on their leaves. Hence, some leaf of the subtree $\mathcal{T} \restriction n = \mathcal{T}' \restriction n$ must be generated from some *zz* distinct from those gg that it is generated from in $\mathcal{T}'$. This, however, contradicts our assumption that Φ is *deterministic*. □

The uniqueness of priority trees for deterministic generators ensures that the corresponding relation $\lhd_\Phi$ of *complete priority* (definition **??**) is non-monotone in the sense that if $xx \lhd_\Phi y$ then there is no $zz \sqsupseteq xx$ such that $zz \lhd_\Phi y$.

1.2.1 *Priority Games*

Definition 7 (Priority Games)**.** Let Φ be any generator, let x be any object and let gg be some objects. The $\Phi, gg-$*priority game* of x $G(\Phi, gg, x)$ is played by two players 1 and 2, according to the following rules. Player 1 starts by playing x. 2 responds by playing the objects yy such that $yy \lhd_\Phi x$. In response, 1 plays any of these yy, and so on. A player wins if her opponent cannot make a move. If 2 plays objects among gg, 1 wins. If a run continues indefinitely, 1 loses.

Proposition 2 (Aczel 1977, 1.5.1)**.** *For every* Φ *and every* x*, player 1 has a winning strategy in* $G(\Phi, x)$ *if and only if* x *has a* Φ, gg*-priority tree.*

1.3 CANTORIAN NUMBERS

In his *Grundlagen*, Cantor presents the ordinal numbers, his *extended number sequence*, as those obtained by two *principles of generation* [Cantor, 1932, pp. 195f]. Firstly, given a number α we generate its successor $\alpha + 1$. Secondly, [Ewald, 1996, pp. 907f]

> [...] if any definite succession of defined integers is put forward of which no greatest exists, a new number is created [...], which is thought of as the *limit* of those numbers; that is, it is defined as the next number greater than all of them.

Some comments are in order. Firstly, I following Jané (2010) and I understand Cantor as identifying the relevant ordering of numbers with the order in which they are generated [p. 197]. Thus, it is important not to think of his principles as presupposing the ordering of the numbers. It is not so as if, say, by the first principle we pick from the ordering

Similarly, it is strictly speaking not the case that second principle allows us to generate, for any definite sequence of numbers, their least upper bound – rather, it allows us to generate a number, and doing so to extend the ordering by a least upper bound.

Secondly, a sequence of numbers

Finally,

How exactly to spell out Cantor's talk of 'definite' collections is subject to scholarly debate. For a recent contribution, see Jané [2010].

In the following, I understand these principles of Cantor as giving a generator in the sense of chapter 1, which allows me to apply the general concept of groundedness to the ordinal numbers. The ordinal numbers are grounded in the number nought through Cantor's generators of successor and limit.

Definition 8 (Cantor's Ordinal Generator)**.** Let an ordinal x be generated from some ordinals xx ($xx\mathbf{O}x$), iff x is their least upper bound, or xx are exactly one ordinal and x is its successor.

1.4 THE ITERATIVE CONCEPTION OF SETS

In the previous chapter I have developed Forster's generalized iterative conception into a general theory of groundedness. Now I turn to several cases of groundedness that have particular philosophical significance. The starting point of Forster's generalization is the *iterative conception of set*. Therefore, it is appropriate for me to firstly present how the standard sets exemplify the general theory of the previous chapter.

The iterative conception of set has a venerable history.[7] Possibly the first and arguably the most influential formulation is found in Gödel [Gödel, 1947, p. 180]. He characterizes the iterative conception of set as that view

> [. . .] according to which a set is anything obtainable from the integers (or some other well-defined objects) by [transfinitely] iterated application of the operation "set of" [. . .].

Definition 9 (The Set generator)**.** Let $xx\,\mathbf{S}\,y$ iff xx are the elements of y.

We have that every pure set is **S**-grounded in nothing.

S gives rise to an operator $\Gamma_{\mathbf{S}}$.

$$y \propto \Gamma_{\mathbf{S}}(xx) \Leftrightarrow \exists zz \sqsubseteq xx(y = \{zz\})$$

In other words, $\Gamma_{\mathbf{S}}$ takes some things xx and outputs all sets formed from subpluralities of the xx. If we assume that the xx form a set, then the set of the things $\Gamma_{\mathbf{S}}(xx)$ is the *power-set* of their set. $\Gamma_{\mathbf{S}}$ is a generalized power-set operation.

By the general proposition 1, x is **S**-grounded just in case it has an **S**-priority tree.

This **S**-priority tree of x gives its elements, their elements, and so on. As an example, figure 4 shows the **S**-priority tree of $\{1, \{2, \{\{2\}, \{1\}\}\}\}$ (using numerals to denote the von Neumann ordinals).

The notion of **S**-groundedness is not strange to set theory. Quite the contrary, it is long and well known, even if under a different label. A set x is **S**-grounded if and only if it *well-founded*.

This observation allows us to connect the concepts of chapter 1 with standard terminology from basic set theory.

The **S**-priority tree of x is isomorphic with its *transitive closure*, ordered by set elementhood $\in$.

Thus, my general concept of groundedness includes the more traditional iterative conception of set. In order to show that it is capable of more, I will in the next section apply it to Kripke's theory of grounded truth.

7 For a recent, opinionated overview see [Ferreiros, 1999, p. 441 - 456].

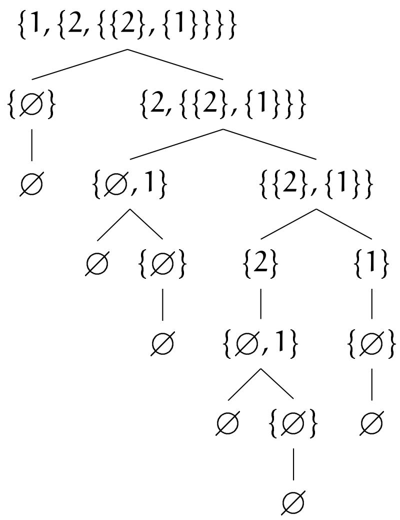

Figure 4: An **S**-priority tree

2 GROUNDED TRUTH

Consider the language $\mathcal{L}_{at}$ of first-order arithmetic extended by a predicate symbol 'T'. For simplicity, let us assume $\neg$, $\vee$ and $\forall$ to be all the primitives and the other connectives and quantifiers to be defined. Let us fix, once and for all, some reasonable method of associating every sentence ϕ with its *Gödel code* $\ulcorner\phi\urcorner$. If X is a set of $\mathcal{L}_{at}$-sentences, I will use '$\ulcorner X\urcorner$' to denote the set of codes $\ulcorner\phi\urcorner$, for $\phi \in X$.

In his seminal [1975], Kripke showed how to expand the standard model of arithmetic $\mathfrak{N}$ by interpretations $\ulcorner X\urcorner$ of 'T' of particular philosophical interest.

The core of his construction is the *Kripke jump* from truth in a model $\mathfrak{M}(\ulcorner X\urcorner)$ to a new interpretation $\ulcorner Y\urcorner$ of 'T', and thus into a new model $\mathfrak{M}(\ulcorner Y\urcorner)$. $\ulcorner X\urcorner$ is a Kripke truth predicate if it is a *fixed point* of such a jump.

My interest is in a certain kind of Kripke truth predicates, those known as predicates of *grounded* truth. The notion of grounded truth is due to Hans Herzberger 1970, but in his 1975 paper, Kripke provides it with new and original content. He does so by telling a story of how an idealized speaker comes to know the concept of truth [Kripke, 1975, pp. 701nn]. This story has been retold many times since Visser [2004]; McGee [1991]; Maudlin [2004]. Nonetheless, I will present it once more, because it renders vivid the close kinship of Kripke's semantic groundedness with the groundedness of sets and other notions of groundedness discussed later; and this aspect of the famous story has not been sufficiently recognized yet.

2.1 LEARNING THE TRUTH

Consider Alice. She speaks a peculiar fragment of English, that is English except for the word 'true'. Further, let us, as usual in philosophy, idealize and assume Alice to have unlimited cognitive capacities and to know for every proposition expressible in her language whether or not it is true.

Now let us present to Alice English sentences that contain 'true'. She does not understand them, because she does not know the meaning of this word. So let us help her. Let us tell Alice that she is to call a proposition 'true' just in case that she can assert it, and 'not true' whenever she is entitled to deny it. Now she already knows that, for instance, snow is white. Recognizing that she can assert this proposition, Alice applies the rule she has just been given and infers that she can also say that 'snow is white' is true. Similarly, she proceeds with everything else she already knows. Due to her omniscience and remarkable cognitive capacities, this means that for every proposition *p expressible in English minus 'true'*, she has now come to understand every sentence in which 'true' applies to a term for *p*.

Now, however, Alice again applies the rules that we have given her, now to these newly understood sentences. Step by step, she therefore

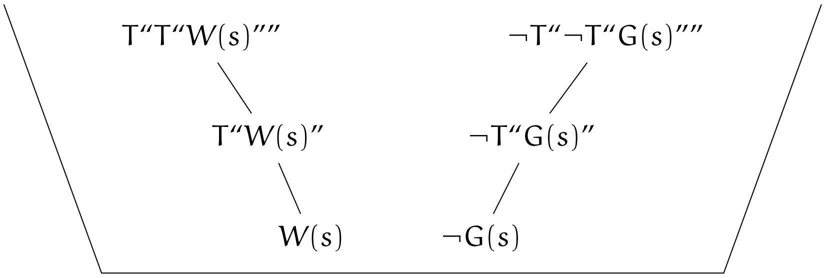

Figure 5: How Alice learns Truth

learns to apply 'true' to more and more sentences, also to those that contain the truth predicate themselves. Since Alice is an idealized subject, it is appropriate to assume that she iterates this step through all natural numbers, and reaches a first limit stage ω. Here, she takes stock of what she has learnt, and understands that she can apply her new word 'true' to every true propositions which she can express so far. Alice keeps going.

Let us focus on how the extension of 'true' increases during this process (see figure 5). At the first stage, 'true' applies to nothing at all. Then, she understands that it applies to every true sentences without 'true'. At the third stage of her learning process, she comes to know that 'true' in addition applies to every true sentence in which 'true' is applied to a sentences that does not contain 'true' itself.

At limit stages, the extension of 'true' is the union of all previous stages.

Now, Kripke suggests that '(...) the "grounded" sentences can be characterized as those which eventually get a truth value in this process', that is, at some stage enter the extension of 'true' [Kripke, 1975, p. 701].

2.2 kripke's construction

Starting out from our base theory, usually the set of truths in $\mathfrak{M}$, the Kripke jump is iterated and more and more sentences containing 'T' enter its interpretation. Kripke calls a sentence "grounded" if it or its negation enters the interpretation at some stage of this construction. The least fixed point extending the base theory collects all and only the grounded sentences.

For this set to be consistent, however, "truth in $\mathfrak{N}(\ulcorner X \urcorner)$" must not mean classical satisfaction. If it did, then for any sentence ϕ containing 'T', the jump of our base theory would contain $\neg T\ulcorner\phi\urcorner$, also if later on $T\ulcorner\phi\urcorner$ comes out as grounded. Let us focus, as usually, on first order arithmetic as our base theory. While $0 = 0$ is a sentence of arithmetic, $T\ulcorner 0 = 0\urcorner$ is not. In this setting, therefore, $\neg T\ulcorner T\ulcorner 0 = 0\urcorner\urcorner$ would be found at the first stage. Likewise, however, this first stage would contain the sentence $T\ulcorner 0 = 0 \urcorner$, since $0 = 0$ is among our base theory. Jumping ahead just once, we would obtain $T\ulcorner T\ulcorner 0 = 0\urcorner\urcorner$, the very sentence whose negation we have just found grounded.

The reason is that classical satisfaction lets $\neg T\ulcorner\phi\urcorner$ come out true whenever ϕ is not in the interpretation of 'T': $\mathfrak{N}(\ulcorner X\urcorner) \vDash T\ulcorner\phi\urcorner$ iff $\phi \notin X$. Hence, Kripke's construction must not be carried out on the basis of classical satisfaction, in particular not on the basis of the classical treatment of *negation*. What is needed is an evaluation scheme that lets a negated sentence come out true, not in the absence of information, but if the available information suffices for it. More precisely, an evaluation scheme m is needed such that $\mathfrak{M}(X) \vDash_m \neg\phi$ only if $\mathfrak{M}(Y) \vDash_m \neg\phi$ holds for every interpretation Y of 'T' extending X. Using a technical term, we need a *monotone* satisfaction [Blamey, 2002]. The crucial feature of a monotone evaluation schema m is that the fact that some sentence code $\ulcorner\phi\urcorner$ is *not* in the extension of 'T' no longer suffices for $T\ulcorner\phi\urcorner$ to come out as true. We no longer have that $\neg T\ulcorner\phi\urcorner$ is true in $\mathfrak{N}(\ulcorner X\urcorner)$ if ϕ is not among the sentences X.

Various monotone schemes m have been used for Kripke's construction. Thus, we have a Kripkean truth predicate based on *Strong* and *Weak* Kleene logic, and constructions that use supervaluational schemes. Note, however, that the need for monotone satisfaction does not imply that our theory of grounded theory cannot be classical. All we have found is that we cannot use classical logic for the Kripke jump, if our goal is a consistent truth predicate. The Kripke jump must be formulated using non-classical logic. But, what we do with our truth predicate thus obtained is a different matter. In particular, we may well reason classically with it. Technically, this means we can take a Kripke truth predicate $\ulcorner X\urcorner$ and work within a classical model $\mathfrak{M}(\ulcorner X\urcorner)$. This approach goes back to Kripke [1975, p. 715] and has been discussed as "closing off" the non-classical model. I will discuss its advantages and disadvantages further below.

Given a monotonic evaluation scheme m, the Kripke jump is standardly formalized by an operator $\mathcal{J}_m$ on sets $\ulcorner X\urcorner$ of (codes of) $\mathcal{L}_{at}$-sentences. For example, $\mathcal{J}_{sk}$ is the standard Kripke jump based on the *Strong Kleene* scheme $\vDash_{sk}$ (definition 11).

$$\ulcorner\phi\urcorner \in \mathcal{J}_m(\ulcorner X\urcorner) \Leftrightarrow \mathfrak{N}(\ulcorner X\urcorner) \vDash_m \phi \tag{1}$$

Kripke called a sentence *grounded* if its code is found in the least fixed point of $\mathcal{J}_m$. My goal is to show that this particular concept of semantic groundedness is a special case of the general concept from section 1. For this, I need to provide a *generator* on the $\mathcal{L}_{at}$-sentences. They form a set *Sent*, which allows me to proceed in the usual set-theoretic setting. In particular, I can represent a generator Φ by a set of pairs $\langle X, \phi\rangle$, where $X \subseteq$ *Sent* and $\phi \in$*Sent*.

Like the generalized power-set operation of section 1.4, the operator $\mathcal{J}_m$ is an example for operators Γ_Φ from chapter **??** (p. 6). However,

$$\neg\big(T\ulcorner\neg\forall x\overline{4}+\dot{x}=\overline{4+\dot{x}}\urcorner \vee \neg T\ulcorner T\ulcorner 0\neq 1\urcorner \vee T\ulcorner 3\neq 1+1\urcorner\urcorner\big)$$

$$\neg\forall x\overline{4}+\dot{x}=\overline{4+\dot{x}} \qquad T\ulcorner 0\neq 1\urcorner \vee T\ulcorner 3\neq 1+1\urcorner$$

$$0\neq 1 \qquad 3\neq 1+1$$

Figure 6: Kripke Groundedness, Coarsely: An Exemplary Priority Tree

what generator Φ does it correspond to? It is a generator that allows us to infer $T\ulcorner\phi\urcorner$ from a set of sentences X if $\mathfrak{N}(\ulcorner X\urcorner)\models_m \phi$. Given X, we generate all sentences $T\ulcorner\phi\urcorner$ such that ϕ is in the Kripke jump of X. Accordingly, I will speak of the *jump* generators and refer to them by **JM**. For example, the Strong Kleene schema $\models_{sk}$ gives rise to the jump generator **JSK**. In general, **JM** is given by the following rule.

$$\frac{X}{T\ulcorner\phi\urcorner} \text{ if } \mathfrak{N}(\ulcorner X\urcorner)\models_m \phi \tag{2}$$

Note that a generator **JM** corresponds to Yablo's notion of *jump-entailment*, or *sufficiency*, for a Kripke jump $\mathcal{J}_m$ [Yablo, 1982, p. 121].[1]

Of course, a generator **JM** allows us also to draw priority trees, as in section 1.2, and thus gives rise to a corresponding notion of *dependence*. Figure 6 provides one example, based on the Strong Kleene jump generator **JSK**. Note that its root, a negated disjunction of the form $\neg(T\ulcorner\phi\urcorner \vee T\ulcorner\psi\urcorner)$ depends not on what is negated, nor on either disjunct, but directly on the sentences of which truth is predicated.

Lemma 4. *Whichever monotone evaluation scheme* m *we consider, the corresponding generator **JM** is not deterministic:*

Proof. If X **JM** ϕ then $\phi = T\ulcorner\psi\urcorner$ and $\psi\in X$, hence for every Y extending X, Y**JM**ϕ. □

Thus, Kripke's concept of semantic groundedness appears to be as simple an instance of the general concept as is the cumulative hierarchy of sets (§**??** above). A sentence is true in Kripke's least fixed point models if and only if it is **JM**-grounded in the empty plurality.

I will refer to this as the *coarse* notion of semantic groundedness. My reason for this label is the following. If we look more closely at Kripke's fixed point construction, we find a richer structure of interacting groundedness than the coarse notion suggests. Given some xx, their set is obtained directly. Given some sentences X, however, its Kripke jump $\mathcal{J}_m(X)$ is better viewed as being obtained in two steps (see figure 7). Firstly, we ascribe truth to all and only the sentence in X, thus moving from complex sentences $\phi\in X$ to atomic sentences

1 Yablo again ascribes this notion of sufficiency to Herzberger [Yablo, 1982, fn. 7].

$T\ulcorner\phi\urcorner$, and infer $\neg T\ulcorner\phi\urcorner$ if $\neg\phi \in X$. Secondly, we close this collection of literals under logic. More precisely, we close the set of literals $\{T\ulcorner\phi\urcorner : \phi \in X\} \cup \{\neg T\ulcorner\psi\urcorner : \neg\psi \in X\}$ under the consequence relation $\models_{\mathfrak{m}}$ that corresponds to the monotone evaluation scheme $\mathfrak{m}$. Doing so, we obtain precisely the $\mathfrak{m}$-complete theory of the model $\mathfrak{N}(\ulcorner X\urcorner)$, in other words the Kripke jump $\mathcal{J}_{\mathfrak{m}}(X)$.

Thus, taking the Kripke jump of a given set X involves two steps: firstly, we ascribe truth, secondly, we close under logic. This fact is missed if we understand Kripke's concept of semantic groundedness in terms of generators **JM**, that allow us to move from the sentences X directly to the complete theory of $\mathfrak{N}(\ulcorner X\urcorner)$. Therefore, generators **JM** provide a merely coarse notion of Kripkean semantic groundedness.

Fortunately, a *finer* understanding of it is available. In the next section, I will outline a general method of replacing a single generator **JM** by two generators **T** and **M** that capture the two distinct steps behind the Kripke jump.

2.3 SEPARATING TRUTH FROM LOGIC

The first step, moving from the set X to the set of literals $\{T\ulcorner\phi\urcorner : \phi \in X\} \cup \{\neg T\ulcorner\psi\urcorner : \neg\psi \in X\}$, corresponds to the generation of sentences $T\ulcorner\phi\urcorner$ from ϕ and $\neg T\ulcorner\phi\urcorner$ from $\neg\phi$. This *truth generator* **T** is common to every variant of Kripke's construction, whichever monotone evaluation scheme $\mathfrak{m}$ we choose. **T** is the core of Kripke's construction.

Definition 10 (Truth Generator)**.** Let **T** be the generator given by the following rules, for sentences ϕ.

$$\text{T-Intro}\ \frac{\phi}{T\ulcorner\phi\urcorner} \qquad \frac{\neg\phi}{\neg T\ulcorner\phi\urcorner}\ \neg\text{T-Intro}$$

Lemma 5. *T is* deterministic.

By itself, **T** allows us to generate more and more statements of the form "it is true that .." and "it is not true that ...", from some given set of sentences, say the truths of arithmetic. However, we will not arrive at any conjunction, disjunction or quantification of such statements. For this, we need to close the set $\{T\ulcorner\phi\urcorner : \phi \in X\} \cup \{\neg T\ulcorner\psi\urcorner : \neg\psi \in X\}$ under logic.

What, however, does it mean to close a set of literals under logic? This depends on our choice of a monotone evaluation scheme $\mathfrak{m}$. In the next section, I will identify *logic* generators **M** such that $\mathfrak{N}(\ulcorner X\urcorner) \models_{\mathfrak{m}} \phi$ iff ϕ is **M**-grounded in the literals $T\ulcorner\zeta\urcorner$, $\zeta \in X$ and $\neg T\ulcorner\zeta\urcorner$, $\neg\zeta \in X$.[2] For example, ϕ is **WK**-grounded in them just in case ϕ holds in the *Weak Kleene* model expanding the standard numbers $\mathfrak{N}$ by an interpretation $\ulcorner X\urcorner$ for 'T'. On this analysis, it is the first step in going from

2 As usual, I call a sentence ϕ a "literal" if it is atomic or the negation of an atomic sentence.

$$\begin{tikzcd}
\zeta_0, \zeta_1, \ldots, \neg\xi_0, \neg\xi_1, \ldots \arrow[d, "\mathbf{T}"] \arrow[dr, "\mathcal{J}_m"] & \\
T\ulcorner\zeta_0\urcorner, \ldots, \neg T\ulcorner\xi_0\urcorner, \ldots \arrow[r, "\mathbf{M}"] & \psi_0, \psi_1, \ldots, \neg\chi_0, \neg\chi_1, \ldots
\end{tikzcd}$$

Figure 7: Splitting Kripke's jump into truth-generation **T** and closure under logic .

X to its Kripke jump $\mathcal{J}_m(\ulcorner X \urcorner)$, the ascription of truth to the sentences in X, that becomes the core of semantic groundedness. And indeed, this first step is the same whichever monotone scheme m we choose.

Kripke characterized the notion of grounded truth by the least fixed point of his jump operator $\mathcal{J}_m$ (equation 1 on p. 18 above). This formal concept is easily viewed as a special case of the general theory of groundedness from section 1, through generators **JM** (equation 2 on p. 19).

However, there also is a finer analysis of Kripke's notion of grounded truth. Kripke's move from a set X to its Kripke jump falls into two steps. Firstly, the sentences X are ascribed truth, and if $\neg\xi \in X$ then ξ is inferred not to be true. We obtain a set of literals $T\ulcorner\xi\urcorner$, $\neg T\ulcorner\xi\urcorner$. It is only in a second step that this set of sentences is closed under the logic determined by our chosen evaluation scheme m. Figure 7 presents these two steps and how the Kripke jump $\mathcal{J}_m$ of the coarse reading combines them into one.

The second step corresponds to a generator **M**. It is closely related to the relation of semantic consequence induced by the monotone evaluation scheme m on which the construction is based. In fact, **M** simply is the way of deriving complex sentences from literals according to m. For this reason, I will speak of *logic generators* **M**. Accordingly, **M** varies between the different ways of carrying out of Kripke's construction. There is a logic generator for the Strong Kleene scheme, a different one for the Weak Kleene scheme, and again other generators for the various supervaluational schemes.

In order for Kripke's construction to give consistent truth predicates, m is assumed to be any *monotone* evaluation scheme. This, however, implies that whichever logic generator **M** we choose, it will not be *deterministic* in the sense of definition 1.

In combination, **T** and **M** provide us with the following analysis of Kripke's concept of grounded truth.

Proposition 3. *If m be Weak or Strong Kleene, then $\ulcorner\phi\urcorner$ is in the least fixed point of Kripke's jump operator $\mathcal{J}_m$ if and only if ϕ is T-**M**-grounded in the $\mathcal{L}_a$-literals true in $\mathfrak{N}$.*

*If m is a supervaluational schema, then $\ulcorner\phi\urcorner$ is in the least fixed point of Kripke's jump operator $\mathcal{J}_m$ if and only if ϕ is T-**M**-grounded in the $\mathcal{L}_a$-sentences true in $\mathfrak{N}$.*

Proof. From lemmata 7, 9 and 12 below. □

My conclusion of the foregoing discussion is that within the general framework of section 1 there are two ways of understanding Kripke's concept of semantic groundedness. On the one hand, there is the *coarse* notion. Given some grounded truths X, more are generated by taking the Kripke jump of X. In particular, $T\ulcorner\phi\urcorner$ is generated not from ϕ alone but from *all* sentences that have already entered the interpretation of 'T'.

On the other hand, there is the *fine* notion based on the combination of a uniform truth generator **T** with one of the logic generators **M**. On this reading, $T\ulcorner\phi\urcorner$ is generated from ϕ alone. If this ϕ itself is a complex sentence then it in turn is grounded in some $\mathcal{L}_{at}$-literals through the logic generator **M**. I now turn to present these logic generators that correspond to the non-classical, monotone evaluation schemes $\mathfrak{m}$.

2.4 STRONG KLEENE LOGIC

Recall the Strong Kleene evaluation scheme $\vDash_{sk}$, as defined, for example, in [Halbach, 2011b, 15.10].[3]

Definition 11 (Strong Kleene)**.** Let $\mathcal{L}$ be a first order language with $\neg, \vee, \forall$ as its primitive logical symbols. Let $\mathfrak{M}$ be an $\mathcal{L}$-model that assigns to every $\mathcal{L}$-relations symbol R^n an extension $J^+(R^n)$ *as well as* an anti-extension $J^-(R^n)$. Let β assign a $\mathcal{L}$-variable to an object of $\mathfrak{M}$'s domain M.

We define $\mathfrak{M} \vDash_{sk} \phi[\beta]$ by induction on the positive complexity of ϕ.

$$\begin{aligned}
\mathfrak{M} \vDash_{sk} R^n\overrightarrow{x_n}[\beta] &\Leftrightarrow \beta(\overrightarrow{x_n}) \in J^+(R^n) \\
\mathfrak{M} \vDash_{sk} \neg R^n\overrightarrow{x_n}[\beta] &\Leftrightarrow \beta(\overrightarrow{x_n}) \in J^-(R^n) \\
\mathfrak{M} \vDash_{sk} \neg\neg\phi[\beta] &\Leftrightarrow \mathfrak{M} \vDash_{sk} \phi[\beta] \\
\mathfrak{M} \vDash_{sk} \phi \vee \psi[\beta] &\Leftrightarrow \mathfrak{M} \vDash_{sk} \phi[\beta] \text{ or } \mathfrak{M} \vDash_{sk} \psi[\beta] \\
\mathfrak{M} \vDash_{sk} \neg(\phi \vee \psi)[\beta] &\Leftrightarrow \mathfrak{M} \vDash_{sk} \neg\phi \text{ and } \mathfrak{M} \vDash_{sk} \neg\psi[\beta] \\
\mathfrak{M} \vDash_{sk} \forall x\phi(x)[\beta] &\Leftrightarrow \text{ for all } m \in M\, \mathfrak{M} \vDash_{sk} \phi(x)[\beta(x:m)] \\
\mathfrak{M} \vDash_{sk} \neg\forall x\phi(x)[\beta] &\Leftrightarrow \text{ there is an } m \in M\, \mathfrak{M} \vDash_{sk} \neg\phi(x)[\beta(x:m)]
\end{aligned}$$

In order to give the Strong Kleene Kripke jump $\mathcal{J}_{sk}$, we can recover the anti-extension of 'T' from its extension, in the following manner. Given a set of sentence X, let $\neg X$ denote the set of all sentences ϕ

3 It is a slight strengthening of Kleene's original truth tables due to Albert Visser [2004].

such that $\neg\phi \in X$. If X comprises the sentences true under some interpretation of 'T', then $\neg X$ are the sentences false under it. $\neg X$ allows us to extract the anti-extension from the extension. Consequently, the Strong Kleene jump $\mathcal{J}_{sk}$ need not work on pairs of sets but is well viewed as taking single sets of sentence codes $\ulcorner X \urcorner$ from which is extracted both positive and negative information.

$$\ulcorner\phi\urcorner \in \mathcal{J}_{sk}(\ulcorner X\urcorner) \Leftrightarrow \mathfrak{N}(\ulcorner X\urcorner, \ulcorner\neg X\urcorner) \vDash_{sk} \phi \tag{3}$$

If we wish, as Kripke does, '$\neg Tx$' to hold of everything that is not a sentence, then we need to make one additional assumption. We let the set of sentence codes $\ulcorner\neg X\urcorner$ contain not only all all codes $\ulcorner\phi\urcorner$ such that $\neg\phi\urcorner \in X$ but also all objects of the domain that do not encode a sentence. In what follows, I will tacitly assume that this trick is implemented.

From the Strong Kleene jump (equation 3) we obtain the generator **JSK**, and on its basis the *coarse* notion of Kripkean groundedness, relative to the Strong Kleene evaluation scheme. Figure 6 gives an example of the corresponding priority trees.

My present interest, however, is in the alternative, *fine* understanding of Strong Kleene groundedness. It is groundedness through the combination of the truth generator **T** (p. **??**) with a logic generator **M**. In order to apply this schema to the case of Kripke's Strong Kleene construction, I therefore need to give a Strong Kleene generator **SK**. This is easily done: I only have to turn the clauses of definition 11 into rules.

Definition 12 (Strong Kleene Truth generator)**.** Let **SK** be the generator given by the following rules.

$$\mathrm{SK}_{\vee L}\ \frac{\phi}{\phi \vee \psi} \qquad \frac{\neg\phi \qquad \neg\psi}{\neg(\phi \vee \psi)}\ \mathrm{SK}_{\neg\vee}$$

$$\mathrm{SK}_{\vee R}\ \frac{\psi}{\phi \vee \psi} \qquad \frac{\phi}{\neg\neg\phi}\ \mathrm{SK}_{\neg\neg}$$

$$\mathrm{SK}_{\forall}\ \frac{\psi(\overline{0}) \qquad \psi(\overline{1}) \qquad \dots}{\forall x\big(\psi(x)\big)} \qquad \mathrm{SK}_{\neg\forall}\ \frac{\neg\psi(\overline{n})}{\neg\forall x\big(\psi(x)\big)}\ \text{for some } n$$

Recall that A is the set of $\mathcal{L}_a$-literals true in $\mathfrak{N}$.

Lemma 6. $\mathfrak{N}(\ulcorner X\urcorner, \ulcorner\neg X\urcorner) \vDash_{sk} \phi$ *if and only if* ϕ *is* ***SK****-grounded in* $A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in X\}$

Proof. Naturally, the lemma is proved by an induction on the positive complexity of ϕ. At the base, let ϕ be a literal. If it does not contain 'T' then we have that ϕ holds in the model $\mathfrak{N}(\ulcorner X\urcorner, \ulcorner\neg X\urcorner)$ iff it is among the A, hence **SK**-grounded in the A. So assume that ϕ is of the form $T\ulcorner\psi\urcorner$ or $\neg T\ulcorner\psi\urcorner$. We observe that

$$\mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} T\ulcorner\psi\urcorner \Leftrightarrow \psi \in X \Leftrightarrow T\ulcorner\psi\urcorner \in \{T\ulcorner\zeta\urcorner : \zeta \in X\}$$
$$\mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \neg T\ulcorner\psi\urcorner \Leftrightarrow \psi \in \neg X \Leftrightarrow \neg T\ulcorner\psi\urcorner \in \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in X\}$$

Either way, ϕ is **SK**-grounded in $A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in X\}$.

At the induction step, let ϕ be the disjunction $\psi \vee \chi$ and assume that the lemma holds for both ψ and χ.

$$\begin{aligned}\mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \psi \vee \chi &\Leftrightarrow \mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \psi \text{ or } \mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \chi \\ &\Leftrightarrow \psi \text{ or } \chi \textbf{ SK}\text{-grounded in } A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in X\} \\ &\overset{SK_\vee}{\Leftrightarrow} \psi \vee \chi \textbf{ SK}\text{-grounded in } A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in X\}\end{aligned}$$

Now let ϕ be the negated disjunction $\neg(\psi \vee \chi)$.

$$\begin{aligned}\mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \neg(\psi \vee \chi) &\Leftrightarrow \mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \neg\psi \text{ and } \mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \neg\chi \\ &\overset{\text{I.H.}}{\Leftrightarrow} \neg\psi \text{ and } \neg\chi \textbf{ SK}\text{-grounded in } A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in \\ &\overset{SK_{\neg\vee}}{\Leftrightarrow} \neg(\psi \vee \chi) \textbf{ SK}\text{-grounded in } A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in\end{aligned}$$

Finally, let ϕ be a quantified sentence $\forall x\psi(x)$.

$$\begin{aligned}\mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \forall x\psi(x) &\Leftrightarrow \text{ for every } n \in \omega, \mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \psi(\overline{n}) \\ &\overset{\text{I.H.}}{\Leftrightarrow} \text{ for every } n \in \omega, \psi(\overline{n}) \textbf{ SK}\text{-grounded in } A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg \\ &\overset{SK_{10}}{\Leftrightarrow} \forall x\psi(x) \textbf{ SK}\text{-grounded in } A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in X\}\end{aligned}$$

$$\begin{aligned}\mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \neg\forall x\psi(x) &\Leftrightarrow \text{ for some } n \in \omega, \mathfrak{N}(\ulcorner X\urcorner, \ulcorner \neg X\urcorner) \vDash_{sk} \neg\psi(\overline{n}) \\ &\overset{\text{I.H.}}{\Leftrightarrow} \text{ for some } n \in \omega, \psi(\overline{n}) \textbf{ SK}\text{-grounded in } A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg \\ &\overset{SK_{10}}{\Leftrightarrow} \neg\forall x\psi(x) \textbf{ SK}\text{-grounded in } A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in X\end{aligned}$$

□

Lemma 7. *As before, let A be the set of $\mathcal{L}_a$-literals true in $\mathfrak{N}$ and let ϕ be any sentence of the extended language $\mathcal{L}_{at}$. We have that ϕ is true in the least fixed point of Kripke's Strong Kleene jump just in case ϕ is **T-SK**-grounded in the A.*

Proof. Volker Halbach shows that a set of sentence codes X is a $\mathcal{J}_{sk}$-fixed point if and only if it contains the (codes of the) A and is closed under rules corresponding to T-Intro, $\neg$T-Intro, $SK_{\neg\neg}$ to $SK_{\neg\forall}$ [Halbach, 2011b, 15.14]. In particular, the least $\mathcal{J}_{sk}$-fixed point is the least such set. □

Lemma 7 justifies the *fine* understanding of Kripke's semantic groundedness based on Strong Kleene logic. It allows us to view the grounded

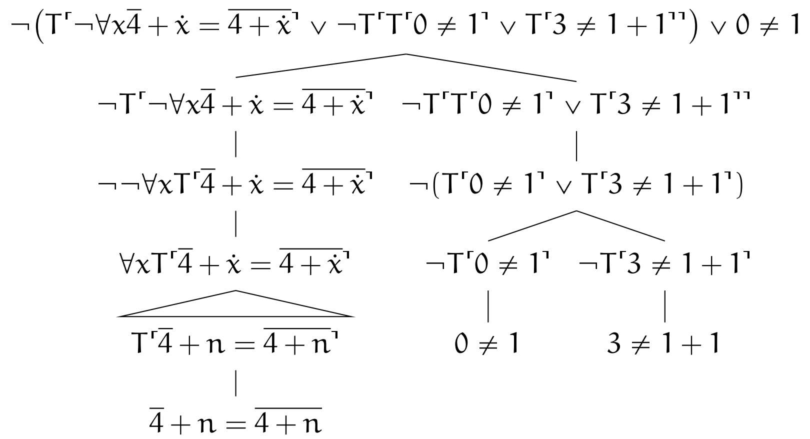

Figure 8: Kripke Groundedness, Finely: An Exemplary **T**-**SK** Priority Tree

sentences as generated from the arithmetical truths, by the combined application of the general truth generator **T** and the Strong Kleene logic generator **SK**. To see the advantages of this fine analysis of Strong Kleene groundedness, compare the **T**-**SK**-priority tree in figure 8 with its coarse analogue 6. Note in particular how the former tracks immediate, logical dependencies such as that of $\neg(T\ulcorner 0 \neq 1\urcorner \vee T\ulcorner 3 \neq 1+1\urcorner)$ on $\neg T\ulcorner 0 \neq 1\urcorner$.

2.5 WEAK KLEENE LOGIC

A variant of Kripke's theory that has gained some attention only recently is based on a *Weak* Kleene evaluation scheme [Feferman, 2008; Fujimoto, 2010]. As on the Strong Kleene scheme considered in the previous section, a relation symbol is assigned both an extension and an anti-extension, and a literal of the form $\neg T\ulcorner\psi\urcorner$ is true not in the absence of ψ from the extension of 'T' but only if ψ is present in its anti-extension. The schemes differ in how complex sentences are treated. On the Weak Kleene approach, a complex sentence is true only if every constituent clause has a definite truth value. For example, the disjunction $\phi \vee \psi$ is true only if both disjuncts are true, ϕ is true and ψ is false or ϕ is false but ψ is true. Accordingly, $\mathfrak{M} \vDash_{wk} \phi[\beta]$ is defined similarly to the Strong Kleene scheme (definition 11), only that the clauses for negated disjunction $\neg(\phi \vee \psi)$ and negated universal quantification $\neg\forall x\phi(x)$ are extended by further conditions.

Definition 13 (Weak Kleene)**.** Let $\mathcal{L}$ and $\mathfrak{M}$ be as in definition 11.

$$
\begin{aligned}
\mathfrak{M} \vDash_{wk} R^n\overrightarrow{x_n}[\beta] &\Leftrightarrow \beta(\overrightarrow{x_n}) \in J^+(R^n)\\
\mathfrak{M} \vDash_{wk} \neg R^n\overrightarrow{x_n}[\beta] &\Leftrightarrow \beta(\overrightarrow{x_n}) \in J^-(R^n)\\
\mathfrak{M} \vDash_{wk} \neg\neg\phi[\beta] &\Leftrightarrow \mathfrak{M} \vDash_{wk} \phi[\beta]\\
\mathfrak{M} \vDash_{wk} \phi \vee \psi[\beta] &\Leftrightarrow \big(\mathfrak{M} \vDash_{wk} \phi[\beta] \text{ and } \mathfrak{M} \vDash_{wk} \psi[\beta]\big) \text{ or } \big(\mathfrak{M} \vDash_{wk} \phi[\beta] \text{ and } \mathfrak{M} \vDash_{wk}\\
&\quad \text{or } \big(\mathfrak{M} \vDash_{wk} \neg\phi[\beta] \text{ and } \mathfrak{M} \vDash_{wk} \psi[\beta]\big)\\
\mathfrak{M} \vDash_{wk} \neg(\phi \vee \psi)[\beta] &\Leftrightarrow \mathfrak{M} \vDash_{wk} \neg\phi \text{ and } \mathfrak{M} \vDash_{wk} \neg\psi[\beta]\\
\mathfrak{M} \vDash_{wk} \forall x\phi(x)[\beta] &\Leftrightarrow \text{ for all } m \in M\, \mathfrak{M} \vDash_{wk} \phi(x)[\beta(x:m)]\\
\mathfrak{M} \vDash_{wk} \neg\forall x\phi(x)[\beta] &\Leftrightarrow \text{ there is an } m \in M\, \mathfrak{M} \vDash_{wk} \neg\phi(x)[\beta(x:m)]\\
&\quad \text{and for all } m \in M\, \mathfrak{M} \vDash_{wk} \phi(x) \vee \neg\phi(x)[\beta(x:m)]
\end{aligned}
$$

Definition 14 (Weak Kleene Truth generator)**.** Let **WK** be the generator given by the following rules.

$$
\mathrm{WK}_{\neg}\ \frac{\phi}{\neg\neg\phi} \qquad \frac{\neg\phi \qquad \neg\psi}{\neg(\phi \vee \psi)}\ \mathrm{WK}_{\neg\vee}
$$

$$
\mathrm{WK}_{\neg\vee A}\ \frac{\phi \qquad \psi}{\phi \vee \psi} \qquad \mathrm{WK}_{\neg\vee B}\ \frac{\phi \qquad \neg\psi}{\phi \vee \psi} \qquad \mathrm{WK}_{\neg\vee C}\ \frac{\neg\phi \qquad \psi}{\phi \vee \psi}
$$

$$
\mathrm{WK}_{\forall}\ \frac{\psi(\overline{0}) \qquad \psi(\overline{1}) \qquad \dots}{\forall x\big(\psi(x)\big)} \qquad \frac{\forall x\big(\psi(x) \vee \neg\psi(x)\big) \qquad \neg\psi(\overline{n})}{\neg\forall x\big(\psi(x)\big)}\ \text{for some } n \qquad \mathrm{WK}_{\neg\forall}
$$

Lemma 8. $\mathfrak{N}(\ulcorner X \urcorner, \ulcorner \neg X \urcorner) \vDash_{wk} \phi$ *if and only if* ϕ *is* ***WK****-grounded in* $A \cup \{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in X\}$

Proof. The claim is proved just analogously to how we proved lemma 6. □

Lemma 9. *Let* A *be as before, and* ϕ *be any sentence of the extended language* $\mathcal{L}_{at}$*. We have that* ϕ *is true in the least fixed point of Kripke's* Weak *Kleene jump just in case* ϕ *is* ***T****-****WK****-grounded in the* A.

2.6 SUPERVALUATION

I now turn to Kripkean theories of truth based on supervaluational logic $\vDash_{sv}$. As is well known, it differs from the Strong and Weak Kleene approach quite significantly. It is not compositional. A disjunction may be true without either disjunct being so. To give a notorious example, $T\ulcorner\lambda \vee \neg\lambda\urcorner$ is in Kripke's least supervaluational fixed points without, of course, neither the liar nor its negation being so.

This specific character of the supervaluational variants of Kripke's theory justifies a more detailed exposition. I will first develop the standard, *coarse* presentation of Kripke's theory, although in a more general format than usual. After that, I will develop a *fine* presentation in terms of the truth generator **T** and specifically supervaluational logic generators **SV**. Since supervaluational logic is not compositional, however, these logic generators will be of a different form than the generators **SK** and **WK** of the previous section.

I begin by generalizing the customary, coarse reading of supervaluational Kripke theories. Recall that this is the reading based on Kripke's jump operator $\mathcal{J}_{\mathrm{m}}$, which turns truth in a model into a new model. The idea behind a supervaluational Kripke jump $\mathcal{J}_{\mathrm{sv}}$ is the following. Given a set of sentence codes X, we consider arbitrary extensions Y of X, each of which induces a classical model $\mathfrak{N}(\ulcorner Y \urcorner)$, with Y interpreting 'T'. Then we use plain classical semantics to determine which sentences come out true in all of these models and add exactly these sentence (codes) to the interpretation X from which we started.

$$\ulcorner \phi \urcorner \in \mathcal{J}_{\mathrm{sv}}(\ulcorner X \urcorner) \Leftrightarrow \forall Y(X \subseteq Y \Rightarrow \mathfrak{N}(\ulcorner Y \urcorner) \models \phi) \tag{4}$$

Of course, the more extensions Y we consider, the less agreement will there be between the models $\mathfrak{N}(\ulcorner Y \urcorner)$, and the less sentences will be in the interpretation of the truth predicate in the resulting new model. Usually, therefore, an *admissibility* condition is imposed on the range of extensions Y considered. Which interpretation Y is considered admissible depends in parts on the set X. Therefore, I will focus on the relation of some set Y *admissibly extending* a set X, in symbols: $X \overset{+}{\subseteq} Y$. For example, Burgess [1986] considers a supervaluational Kripke jump that requires not only Y to extend X, but also it not to contain any sentence ϕ such that $\neg\phi \in X$. Let $\overline{X}$ denote the complement of X, and recall that $\neg X$ denote the set $\{\phi : \neg\phi \in X\}$.
Then, we can define Burgess condition on Y being an admissible extension of X as follows.

$$X \overset{+}{\subseteq} Y :\leftrightarrow X \subseteq Y \subseteq \overline{\neg X}$$

Burgess restricts quantification to sets Y "sandwiched between" the given set X and those sentences ψ whose negation does not occur in X. This choice of an admissibility condition gives rise to following Kripke jump.

$$\ulcorner \phi \urcorner \in \mathcal{J}_{\mathrm{bs}}(\ulcorner X \urcorner) \Leftrightarrow \forall Y\big(X \subseteq Y \subseteq \overline{\neg X} \Rightarrow \mathfrak{N}(\ulcorner Y \urcorner) \models \phi\big) \tag{5}$$

In the literature, various admissibility conditions $\overset{+}{\subseteq}$ have been made use of. To give just one other example, Cantini [1990] works with a Kripkean theory based on the stronger admissibility condition of Y being a *consistent* extension of X.

$$\ulcorner \phi \urcorner \in \mathcal{J}_{\mathrm{cs}}(\ulcorner X \urcorner) \Leftrightarrow \forall Y\big(X \subseteq Y \,\&\, \mathit{Con}(\ulcorner Y \urcorner) \Rightarrow \mathfrak{N}(\ulcorner Y \urcorner) \models \phi\big) \tag{6}$$

Other, stronger admissibility conditions are conceivable, too. In the following, I will reason schematically, for an arbitrary admissibility condition $\Subset$. In particular, I will not assume it to be definable but treat an admissibility condition as a set of candidate interpretations of 'T'.

Thus, $\mathcal{J}_{bs}$ and $\mathcal{J}_{cs}$ are instances of the following general schema.

$$\ulcorner\phi\urcorner \in \mathcal{J}_{as}(\ulcorner X\urcorner) \Leftrightarrow \forall Y(X \Subset Y \Rightarrow \mathfrak{N}(\ulcorner Y\urcorner) \models \phi) \tag{7}$$

Given an admissibility condition $\Subset$ the corresponding Kripke jump $\mathcal{J}_{as}$ determines a way of generating sentences of the form $T\ulcorner\phi\urcorner$ from a given set of sentences X.

$$\textbf{JAS}\ \frac{X}{T\ulcorner\phi\urcorner}\ \text{if } \phi \in \mathcal{J}_{as}(\ulcorner X\urcorner) \tag{8}$$

We find that the least $\mathcal{J}_{as}$-fixed point comprises exactly the sentences grounded in nothing through this generator. This is not very surprising, as equation 8 does hardly more than rewriting the step from one stage of Kripke's construction to the next. Still, we have thus given a *coarse* reading of Kripke's supervaluational concept of grounded truth.

However, it is not satisfactory yet. The reason is this. Whichever admissibility condition is imposed, we have restricted our perspective to classical extension *of the standard model* $\mathfrak{N}$. Consequently, all arithmetical truths are gotten for free, in the following precise sense. If a supervaluational jump is applied to the *empty* set, then agreement is required between *all* admissible interpretations of 'T'. However, the interpretation of the base language $\mathcal{L}_a$ of first order arithmetic is fixed. Hence, even if no information is available yet all $\mathcal{L}_a$-sentences true in $\mathfrak{N}$ are obtained.

$$\mathcal{J}_{as}(\emptyset) = \{\ulcorner\phi\urcorner : \mathfrak{N} \models \phi\}$$

As a result, it appears as if supervaluational Kripke truth over first-order arithmetic does not need first-order arithmetic as basis. Truth seems to be *grounded* in the empty set. But this is an illusion, due to the fact that we have built truth-in-$\mathfrak{N}$ into the jump operator $\mathcal{J}_{as}$. Really, Kripkean truth is grounded in true arithmetic.

A theory of groundedness should not leave this important fact implicit. Therefore, I generalize the supervaluational approach captured in equation 7 by quantifying not only over admissible interpretation of 'T', but also over arbitrary $\mathcal{L}_a$-models $\mathfrak{M}$.

Definition 15. Given an admissibility condition $\Subset$ let the *generalized* supervaluational jump $g\mathcal{J}_{as}$ be defined as follows.

$$\ulcorner\phi\urcorner \in g\mathcal{J}'_{as}(\ulcorner X\urcorner) :\Leftrightarrow \forall\mathfrak{M}\forall Y\big(X \Subset Y \Rightarrow \mathfrak{M}(\ulcorner Y\urcorner) \models \phi\big)$$

Lemma 10. *Let* N *be the set of* $\mathcal{L}_a$*-sentences true in* $\mathfrak{N}$*:* $\mathrm{N} \coloneqq \{\phi : \phi \in \mathcal{L}_a$ *and* $\mathfrak{N} \models \phi\}$*. Let* $\Subset$ *be an admissibility condition,* $\mathfrak{J}_{as}$ *the corresponding supervaluational Kripke jump, and* $g\mathfrak{J}_{as}$ *its generalized variant as defined in* 15.

For every set of sentence-codes X*,*

$$\phi \in \mathfrak{J}_{as}(\ulcorner X \urcorner) \Leftrightarrow \phi \in g\mathfrak{J}_{as}(\ulcorner X \urcorner) \text{ or } \phi \in \mathrm{N}$$

Proof. I show that for every sentence ϕ of the language $\mathcal{L}_{at}$ we have that

$$\forall Y(X \Subset Y \Rightarrow \mathfrak{N}(\ulcorner Y \urcorner) \models \phi)$$
$$\Leftrightarrow$$
$$\forall \mathfrak{M} \forall Y\big(X \Subset Y \Rightarrow \mathfrak{M}(\ulcorner Y \urcorner) \models \phi\big) \vee \mathfrak{N} \models \phi$$

The claim is trivial for arithmetical ϕ. For ϕ that contain 'T' I reason by induction on its syntactic complexity. Note that the right-to-left direction is dealt with uniformly, and easily, by letting $\mathfrak{M} = \mathfrak{N}$. A little care is needed for the left-to-right direction. At the induction base, ϕ is assumed to be of the form $T\ulcorner\psi\urcorner$. In this case, however, if $\mathfrak{N}(\ulcorner Y \urcorner) \models \phi$ then indeed for every model $\mathfrak{M}$, $\mathfrak{M}(\ulcorner Y \urcorner) \models \phi$, as desired. For the induction step, assume that $\phi = \neg\chi$, and that the claim holds for χ. We assume that $\mathfrak{N}(\ulcorner Y \urcorner) \models \neg\chi$ holds for every admissible superset Y of X, let $\mathfrak{M}$ be any $\mathcal{L}_a$ model and Y an admissible extension of X. For contradiction, assume that $\mathfrak{M}(\ulcorner Y \urcorner) \not\models \neg\chi$, hence $\mathfrak{M}(\ulcorner Y \urcorner) \models \chi$. But then, by the induction hypothesis, $\mathfrak{N}(\ulcorner Y \urcorner) \models \chi$, contradiction.

The other cases follow from the induction hypothesis directly. For example, if $\phi = \chi \vee \psi$ then we know that either $\mathfrak{N}(\ulcorner Y \urcorner) \models \chi$ or $\mathfrak{N}(\ulcorner Y \urcorner) \models \psi$; in both cases, however, the induction hypothesis ensures that $\mathfrak{M}(\ulcorner Y \urcorner) \models \chi \vee \psi$. □

The generalized supervaluational Kripke jump (equation 15) corresponds to a generalized *coarse* Kripke generator **gJAS**, such that every sentence in the least $g\mathfrak{J}_{as}$-fixed point is **gJAS**-grounded in the truths of arithmetic.

$$\textbf{gJAS}\ \frac{X}{T\ulcorner\phi\urcorner} \text{ if } \ulcorner\phi\urcorner \in g\mathfrak{J}_{as}(\ulcorner X \urcorner) \tag{9}$$

I now turn to the *fine* understanding of Kripke's semantic groundedness. According to it, the sentences of Kripke's least fixed point based on the monotone evaluation scheme m are grounded in the base theory through the combination of the truth generator **T** with a logic generator **M**. This **M** is the generator through which a sentence ϕ is grounded in the literals $T\ulcorner\zeta\urcorner$, $\neg T\ulcorner\xi\urcorner$ (plus the base language literals true in the base model), if and only if ϕ is m-true under the interpretation of 'T' by exactly those $\zeta, \neg\xi$. In the previous two sections, I gave such a generator **SK** for Strong Kleene logic as well as a

Weak Kleene generator **WK**. Now, my goal is to identify a generator for supervaluation.

As noted above, supervaluational satisfaction schemes are not compositional. Therefore, closing under supervaluational logic

Consequently, the supervaluational generators **AS** will not be given by neat rules such as $SK_\vee$. In order to generate a sentence ϕ from some other sentences X, we need to consider a range of admissible models of the language.[4] These admissible models, however, are all classical.

...

What I called the Strong-Kleene generator in fact allows us to generate, given all literals true in a *classical* model its complete classical theory.

Fact 1 (McGee 1990, example 5.5)**.** *Let* A *be the* $\mathcal{L}_a$*-literals true in* $\mathfrak{N}$*. That is, let* A *be the set* $\{s = t : \mathfrak{N} \models s = t\} \cup \{s \neq t : \mathfrak{N} \models s \neq t\}$.

For every complex $\mathcal{L}_{at}$*-sentence* ϕ, $\mathfrak{N}(\ulcorner X \urcorner) \models \phi$ *iff* ϕ *is* ***SK****-grounded in* $A \cup \{T\ulcorner\xi\urcorner : \xi \in X\} \cup \{\neg T\ulcorner\xi\urcorner : \xi \notin X\}$.

Thus, taking the (classically) complete theory of a model $\mathfrak{N}(\ulcorner X \urcorner)$ is to take the $\mathcal{L}_{at}$-sentences **SK**-grounded in the literals true in $\mathfrak{N}(\ulcorner X \urcorner)$.

...

Now, let us say that given an admissibility condition $\Subset$ and sets of sentences X, Z, the set $\{T\ulcorner\zeta\urcorner : \zeta \in Z\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in Z\}$ *admissibly extends* $\{T\ulcorner\xi\urcorner : \xi \in X\} \cup \{\neg T\ulcorner\xi\urcorner : \neg\xi \in X\}$ *truth-wise* if $X \Subset Z$. Then, the supervaluational generator **As** is well viewed as function of **C**-groundedness in admissible extension.

Definition 16 (Supervaluational generators)**.** Given an admissibility condition $\Subset$ and a set of $\mathcal{L}_{at}$-sentences X, let us say that ϕ is **AS**-generated from the X if and only if

- X is a set of literals $T\ulcorner\xi\urcorner, \neg T\ulcorner\xi\urcorner$ and
- ϕ is **C**-grounded in every set of such literals that admissibly extends X truth-wise.

For example, **BS** is the generator corresponding to Burgess' Kripke jump (equation 5). We have that

Lemma 11. *Let* $\Subset$ *be an admissibility condition, and* X *a set of* $\mathcal{L}_{at}$*-sentences.*

We have that $\phi \in \mathfrak{g}\mathfrak{J}_{as}(X)$ *if and only if* ϕ *is* ***AS****-grounded in the* $\{T\ulcorner\zeta\urcorner : \zeta \in X\} \cup \{\neg T\ulcorner\zeta\urcorner : \neg\zeta \in X\}$

Proof. Immediate from the definitions 16 and 15. □

4 Closely related to this is the fact that supervaluational consequence is not computably enumerable. Further below, when I ask for ways of axiomatizing grounded theories, I will discuss these computational aspects of Kripke's constructions.

Recall that 'N' denotes the set of base language -sentences ϕ such that $\mathfrak{N} \vDash \phi$. Together with lemma **??**emma 11 implies that Kripkean groundedness given by a supervaluational fixed point construction based on $\Subset$ is groundedness in N through the corresponding generator **AS**.

Lemma 12. *Let $\Subset$ be an admissibility condition, and ϕ any $\mathcal{L}_{at}$-sentence. We have that $\ulcorner\phi\urcorner$ is in the least fixed point of the Kripke jump based o$\Subset$ n just in case ϕ is **T**-**AS**-grounded in* N.

Proof. From lemma 10 we know that ϕ is in the least fixed point of $\mathcal{J}_{as}$ if and only if ϕ is in the least fixed point of $g\mathcal{J}_{as}$ that extends N. Now, by lemma 12 this set is is the smallest set extending N that is closed under **AS** and the rules T-Intro and $\neg$T-Intro, i.e. the truth generator **T**. Therefore, ϕ is in the least fixed point of $\mathcal{J}_{as}$ if and only if ϕ is **T**-**AS** grounded in N. □

2.7 LEITGEB GROUNDEDNESS

Recently, Leitgeb has developed a variant of Kripke's theory based on a concept of one sentence ϕ *semantically depending* on a set of sentences X. ϕ depends on X if there is

> [...] no difference in the truth value of φ without a corresponding difference in the extension of the truth predicate as far as the members of [X] are concerned [...] [Leitgeb, 2005, p. 160]

Leitgeb continues to point out that

> [...] the notion of dependence which we aim at is a kind of supervenience: the truth value of φ supervenes on which members of are to be found in the extension of ['T'] (ibid.)

To emphasize this aspect of Leitgeb's concept, as well as to avoid confusion with the more general concept of dependence from section 1 (definition 6 on p. 8), I will henceforth speak of Leitgeb's concept as semantical *supervenience*.

Leitgeb works within the usual setting of formal truth theory. To the language of arithmetic $\mathcal{L}$ is added a monadic predicate 'T', to be read 'is true'. The extension of this predicate is a set of sentences of the extended language $\mathcal{L}_t$. Whichever set is taken for this, a different model of $\mathcal{L}_t$ is obtained. The idea behind Leitgeb's notion of groundedness is that the truth value of sentences containing 'T' *supervenes* on this choice.

Definition 17. The sentence ϕ supervenes, relative to arithmetic, on the set of sentences X if variation in truth value requires variation of the interpretation of 'T' with respect to X: for all sets Y, Z, $\mathfrak{N}(Y) \vDash \phi \not\Leftrightarrow \mathfrak{N}(Z) \vDash \phi$ only if $\mathfrak{N}(Y \cap X) \vDash \phi \not\Leftrightarrow \mathfrak{N}(Z \cap X) \vDash \phi$.

Thus, a sentence ϕ is said to supervene on a set of sentences X, if it matters to the interpretation of ϕ at most whether or not the X are in the extension of 'T'. For example, 'T⌜0 < 1⌝' supervenes on {'0 < 1'}. The sentence can be true in one model and false in another only if the extension of 'T' as interpreted in the first differs from the extension of 'T' in the latter with respect to {'0 < 1'}.

Leitgeb defines a D^{-1} that maps a set of sentences to the set of just those sentences which supervene on the first.

Definition 18 (Leitgeb's Supervenience Jump)**.**

$$\phi \in D^{-1}(\ulcorner X \urcorner) \Leftrightarrow \phi \text{ supervenes on } X$$

D^{-1} is monotone on the sets of $\mathcal{L}_{at}$-sentences. Its least fixed point point collects all and only the sentences ϕ that supervene on sets of sentences of the arithmetical base language $\mathcal{L}_a$. This is Leitgeb's concept of semantic groundedness: ϕ is grounded just in case $\phi \in D_{lf}$ [Leitgeb, 2005, Def 12].

A sentence φ supervenes on a candidate interpretation X of '*T*' if there is no difference in the truth value of φ without a difference in the interpretation of '*T*'. It does not matter, however, which truth value is assigned to φ in the model that interprets '*T*' by X. In other words, Leitgeb's supervenience relation does not distinguish between truth and falsehood. Thus, among the grounded sentences there are both true and false ones. To define grounded *truth*, Leitgeb constructs a set T_{lf} such that for every sentence $\phi \in D_{lf}$, $\phi \in T_{lf}$ just in case $\mathbb{N}(T_{lf}) \models \phi$.

As with the supervaluational approach of the previous section, Leitgeb's framework requires generalization. When he writes 'no difference in the truth value of ϕ', he means its truth value in an expansion of the *standard model*. As a result, purely arithmetical sentences do not depend, or as I will say supervene, on anything. Thus it appears as if nothing was needed to generate the sentences *grounded* in arithmetic. Arithmetic comes for free.

In order to render explicit that Leitgeb's groundedness is groundedness *in* the sentences of arithmetic, I will generalize Leitgeb's definition of semantic supervenience and consider quantification over *arbitrary* $\mathcal{L}_a$-models.

Definition 19 (Generalized Semantic Supervenience)**.** ϕ semantically supervenes on X iff

$$\forall \mathfrak{M} \forall Y (\mathfrak{M}(\ulcorner Y \urcorner) \models \phi \Leftrightarrow \mathfrak{M}(\ulcorner Y \cap X \urcorner) \models \phi)$$

This concept gives naturally rise to a *generator*.

Definition 20 (Semantic Supervenience generator)**.**

$$\mathbf{L} := \{\langle X, \phi \rangle : \phi \text{ supervenes on } X\}$$

Recall Leitgeb's jump D^{-1}

Proposition 4.

$$\phi \in D^{-1}(\ulcorner X \urcorner) \Leftrightarrow \exists Y \subseteq X \langle Y, \phi \rangle \in \boldsymbol{L}$$

Proof. Right-to-left: trivial, let $Y = X$. Left-to-right: from the fact that if ϕ supervenes on X and $X \subseteq Y$ then ϕ supervenes on Y (lemma 3 in Leitgeb's 2005). □

Corollary 1. *The $\mathcal{L}_{at}$-sentences in the least fixed point of D^{-1} over $\mathcal{L}_a$ are exactly the sentences **L**-grounded in the sentences of the arithmetical base language $\mathcal{L}_a$. In other words, Leitgeb's notion of groundedness coincides with the concept of **L**-groundedness in the arithmetical sentences.*

Recall the concept of *deterministic* generators (definition 1).

Proposition 5. ***L** is not deterministic.*

Proof. Immediately from lemma 3 in Leitgeb (2005). □

2.8 LEITGEB'S GROUNDED TRUTH AS A VARIANT OF SUPERVALUATIONAL GROUNDED TRUTH

Remark 1. Γ_{lf} can be defined directly, using a single operator.

Definition 21. (Bonnay, van Vugt) Let G be an operator on sets of L[T]-sentences such that

$$G(X) = \{\phi : \phi \text{ supervenes on } X \cup \neg X \text{ and } Val_X(\phi) = 1\}$$

Lemma 13. *(Bonnay, van Vugt) For any α,*

$$\Gamma_{\alpha+1} = G(\Gamma_\alpha)$$

Proof.

$$\begin{aligned}
\Gamma_{\alpha+1} &= \{\phi : \phi \in \Phi_{\alpha+1} \text{ and } Val_{\Gamma_\alpha}(\phi) = 1\} \\
&\overset{\text{Definition of } \Phi_{\alpha+1}}{=} \{\phi : \phi \text{ supervenes on } \Phi_\alpha \text{ and } Val_{\Gamma_\alpha}(\phi) = 1\} \\
&\overset{\text{Lemma ??}}{=} \{\phi : \phi \text{ supervenes on } \Gamma_\alpha \cup \neg\Gamma_\alpha \text{ and } Val_{\Gamma_\alpha}(\phi) = 1\} \\
&\overset{\text{Definition 13}}{=} G(\Gamma_\alpha)
\end{aligned}$$

□

Lemma 14. *(Martin Fischer) G is not monotone.*

Proof. Let λ be a (non-strengthened) liar sentence, such that for every X, $\mathfrak{N}(X) \vDash \lambda \leftrightarrow T\ulcorner\neg\lambda\urcorner$. Note that λ as well as $\neg\lambda$ supervenes on $\{\lambda\}$, *a fortiori* on $\{\lambda, \neg\lambda\}$. Since obviously, $\mathfrak{N}(\{\ulcorner\lambda\urcorner\}) \vDash \neg T\ulcorner\neg\lambda\urcorner$, we have that $\ulcorner\neg\lambda\urcorner \in G(\{\ulcorner\lambda\urcorner\})$. However, $\ulcorner\neg\lambda\urcorner$ is not an element of $G(\{\ulcorner\lambda, \neg\lambda\urcorner\})$, since $\mathfrak{N}(\{\ulcorner\lambda, \neg\lambda\urcorner\}) \vDash T\ulcorner\neg\lambda\urcorner \wedge (\lambda \leftrightarrow T\ulcorner\neg\lambda\urcorner)$. □

As Leitgeb noted in [Leitgeb, 2006, p. 83], and Meadows and Bonnay-van Vugt have recently rediscovered independently, Leitgeb's theory is closely related to the Kripke truth theory in *supervaluational* logic.

Defining Leitgeb's truth predicate Γ_{lf} in terms of the operator G (definition 13) makes this easy to see.

Recall that according to definition 17 the supervenience of ϕ on X is a matter of how its truth value varies with *arbitrary* reinterpretation of 'T'.

But, we may want to add some constraints on how 'T' is to be interpreted, just as we impose *admissibility* constraints on candidate interpretations of 'T' on the supervaluational approach of section 2.6. Together with lemma **??**, this motivates an alternative definition of semantic supervenience.

Definition 22 (Semantic supervenience, qualified)**.** ϕ semantically supervenes on X with respect to an admissibility conditio$\Subset$ n just in case, for every Y, X $\Subset$ Y,

$$\forall \mathfrak{M} \forall Y(\mathfrak{M}(\ulcorner Y \urcorner) \models \phi \Leftrightarrow \mathfrak{M}(\ulcorner Y \cap X \urcorner) \models \phi)$$

For example, we may want these interpretations to include all the truths that we have already identified, and to exclude all falsehoods of which we already know. In other words, we may restrict our attention to interpretations "sandwiched between" the truths and the falsehoods. Doing so, we adopt Burgess' admissibility condition from p. 27 for Leitgeb's jump of grounded truth G.

Proposition 6. *Let $\Subset$ be an admissibility condition. For every α,*

$$\mathfrak{J}^{\alpha}_{as} = s\Gamma_{\alpha} \tag{10}$$

$$\tag{11}$$

In particular, their least fixed points coincide.

3

THE GROUNDEDNESS APPROACH TO CLASS THEORY: KRIPKEAN CLASS THEORIES

Can we make a similar move in our present situation? Can we apply Kripke's method to single out the *grounded* instances of naive class comprehension? Extant literature gives reason to be hopeful. Most prominently, Penelope Maddy has carried out a Kripkean construction over set theory.[1] The present chapter is intended as a general and systematic investigation into the prospects of grounded class theory. In the next section, I develop properties we would like such a theory to have. However, it is not guaranteed that these desiderata can all be satisfied; and maybe they need not all be. What follows are *prima facie* desirable features.

3.1 DESIDERATA FOR A THEORY OF GROUNDED CLASSES

Firstly, whichever way we approach a theory of grounded classes, we wish to answer Russell's paradox while allowing the membership relation to figure on the right-hand side of class comprehension. Thus, one desideratum is immediate. We want our theory to get us as much of comprehension as possible.

COMPREHENSION A class theory should contain many instances of class comprehension.

At this point, let me emphasize that although they pose analogous challenges, Tarski's schema and naive class comprehension differ in one respect. Whereas sentences are plugged into (T), the schema of comprehension takes open formulae; and these are universally quantified. As a result, one instance of (C) corresponds to many instances of (T). Comprehension for the formula $\phi(x)$ is grounded only if *for every* closed term a, the sentence $\phi(a)$ is grounded. In effect, as we will see, identifying grounded fragments of (C) is significantly more demanding than restricting the T schema to its grounded instances.

In order to motivate the second desideratum, allow me to ask: what do we need class theory for in the first place? After all, we already have a theory of *sets*, and it is both mathematically well developed and philosophically motivated. One way to argue that we also need a theory of classes is as follows.[2]

There are two ways of collecting some things.[3] On the one hand, we collect some things by a sequence, possibly uncountable, of independent decisions whether a given object belongs to them or not – basically, by listing them. This *combinatorial* idea of collection underlies the theory of sets.

1 Maddy [1983, 2000]. For an alternative approach, see Cantini [1996]. Mathematically, the theories also relate loosely to work by Feferman (1975a; 1975b) and Aczel (1980).
2 See Maddy (1983, §1) for the history of this line of thought.
3 Of course, from the Platonist viewpoint usually adopted, 'collection' strictly speaking is a metaphor. Much of the philosophy of set theory is devoted to explicating this metaphor, see e.g. Parsons [1977].

On the other hand, we collect some things by giving a condition which exactly they satisfy. This is the *definitional*, or *logical*, idea of collection. For example, we may use the condition of being an ordinal number to collect, well, the ordinals. On pain of contradiction, there is no set of all the ordinals. Hence, in order to fully capture the definitional idea of collection, standard set theory needs to be supplemented by a theory of classes.[4]

We would like to motivate our theory of grounded classes in this manner.

IDEA A class theory should stand to the *definitional* idea of collection as standard set theory stands to the combinatorial idea.

This desideratum is explicated naturally as follows. Defining conditions are closed under the logical connectives. Thus, we would like our classes to be closed under Boolean operations. For example, if according to our theory x is not in the class of the ϕs, then x must be in the class defined by the condition $\neg\phi$, in order for our theory to satisfy the desideratum. Further, there is a trivial condition (e.g. $x = x$) as well as one that nothing satisfies ($x \neq x$). Hence, our theory should have a universal and an empty class.

I turn to the next desideratum. By itself, the definitional idea leaves open when two conditions define the same collection. We may consider intensional identity criteria of different granularity.[5] My interest, however, is in those definitional collections the naive theory of which gave rise to Russell's paradox; and this notion of class, or concept-extension, is extensional. For example, the class of the ordinals is the class of the hereditarily transitive sets, since everything is an ordinal iff it is a hereditarily transitive set. Accordingly, our theory of grounded classes ought to make them extensional.[6]

EXTENSIONALITY A class theory should imply that the class of the ϕs is the class of the ψs just in case: everything is a member of the class of the ϕs just in case it is a member of the class of the ψs.

Finally, class talk is not peculiar to philosophers. Mathematicians speak of classes, too.[7] We would like our theory of classes to account for the usage of the notion in mathematics, at least for some of it.

4 See Øystein Linnebo [2006]. To be explicit, I do not argue that standard set theory ought to be *replaced* by a theory of classes. Thus, the class theories developed below are not intended to play the role that, e.g., Quine's *New Foundations* is meant to fulfil.

5 Intensional theories of classes have been developed within the proof-theoretic programme of *explicit mathematics* (Feferman [1975b, 1979]; Jäger et al. [2001]).

6 The set theoretic axiom of extensionality has been argued for on pragmatic, or external, grounds (Fraenkel et al. [1973], Maddy (1988, p. 483)). It seems to me that these arguments carry over to class theory.

7 See Parsons [1974] and Uzquiano [2003] for discussion of this point.

How do working mathematicians use the notion of class? I will concentrate on two observations. On the one hand, the notion of class is used generally to speak of any collection which is not a set. In particular, different kinds of things are taken to form classes. Not merely sets, but numbers, graphs and categories. Consequently, our theory of grounded classes should be equally applicable to various areas. This intuitive thought must be rendered precise, however, as there are different senses in which a theory may be thought to be generally applicable. The relevant notion of applicability is this: we would like to be able to extend any given theory by classes grounded in it. It is in this specific sense of applicability that the following desideratum is to be understood.

BASE A class theory should be applicable to a variety of base theories.

On the other hand, mathematicians reason classically. Hence, we have the following desideratum.

CLASSICALITY A class theory should be closed under classical logic.

In this section, I have collected what we would, *prima facie*, a theory of grounded classes to be like. Next, I will explore how to develop such a theory.

It can be done in two ways. On the one hand, we may develop the theory directly, giving axioms or characterizing its intended model. Maddy followed this method.[8] On the other hand, we may take a theory of grounded truth and translate it into the language of 'η'. Work by Andrea Cantini can be viewed as being of this kind.[9] The former, direct approach is arguably more natural. However, examining the latter, derivative method will illuminate challenges specific to class theory. Therefore, I will begin by exploring what can be done derivatively, and turn to the direct approach later (§3.4).

My presentation will be largely self-contained. As to notation, I will mostly follow Halbach [2011b]. Deviation from or addition to his symbolism will be made explicit.

3.2 DERIVING GROUNDED CLASSES FROM GROUNDED TRUTH

In this section I examine theories of grounded classes derived from a given theory of grounded truth. The idea is this. We translate the language of class theory into the language of truth theory, roughly by translating $a\eta\ulcorner\phi\urcorner$ as $T\ulcorner\phi(a)\urcorner$.[10] Then, we endorse as our theory of grounded classes the set of sentences whose translations follow from our favourite theory of grounded truth.

8 See Maddy [1983, 2000]. The key technical idea is found already in Brady [1971].
9 Cantini (1996) §§9-11.
10 Recall that $\ulcorner\phi\urcorner$ stands for ϕ's Gödel code or numeral, depending on the context.

To see how this works in detail, let us focus on the most popular theory of grounded truth, the theory of Kripke's least fixed point model based on Strong Kleene logic.[11] Let $\mathcal{L}$ be the language of first-order arithmetic plus 'η'.[12] When dealing with paradox, caution is needed that may otherwise seem unnecessarily circumstantial. This concerns in particular the distinction of object- and meta-language. I will use letters from the beginning of the Roman alphabet ('a', 'b' etc.), with sub- and superscripts, as meta-linguistic variables for $\mathcal{L}$-terms and variables, and letters from the end of the Roman alphabet ('x', 'y' etc.), with sub- and superscripts, as variables of the language $\mathcal{L}$. Fix a specific $\mathcal{L}$-variable x_0, and let *Fml* be the set of $\mathcal{L}$-formulae with x_0 as their single free variable.

In order to derive a class theory from Kripke's theory of truth, we translate an $\mathcal{L}$-sentence ψ into the language of truth theory. To explain just how this is done will require me to go into some detail. Readers less formally inclined need not to follow me all the way; it suffices to keep in mind the basic idea that we translate $a\eta\ulcorner\phi\urcorner$ as $T\ulcorner(\phi)^\star(a)\urcorner$, for $(\phi)^\star$ the translation of ϕ.

Usually, we define a translation by induction on syntactic complexity. The translation $(\cdot)^\star$, however, cannot be obtained in this manner, since in order to translate an atomic formula $a\eta\ulcorner\zeta \vee \xi\urcorner$ we must already have translated the complex formula $\zeta \vee \xi$. Towards an alternative definition of our translation, I propose the following notion of a formula's *rank*. Formulae of the base language have rank 0.

Also, formulas $a\eta b$ have rank 0 iff b is a variable, or a closed term that is not a Gödel numeral $\ulcorner\phi\urcorner$. The rank of a formula '$a\,\eta\,\ulcorner\phi\urcorner$' is one greater than the rank of ϕ. Complex formulae containing 'η' inherit their rank from their immediate constituents. For example, the rank of $\phi \vee \psi$ is the rank of ϕ or ψ, whichever is greater; and the rank of $\exists x\phi$ is that of ϕ. The fact that the code of $a\eta\ulcorner\phi\urcorner$ is strictly greater than that of ϕ, ensures the relation "... is of lower rank than ... " to be well-founded on the $\mathcal{L}$-formulae. Thus, we can translate the language of class theory $\mathcal{L}$ into the language of truth theory.

A central role will be played by the syntactical operation *Sb* which takes a term a and a formula $\phi \in Fml$, and outputs the substitution of a for x_0 in ϕ.[13] On the basis of our coding $\ulcorner\ldots\urcorner$, $Sb(a,\phi)$ is represented by an arithmetical formula $Sb^\bullet(x,y)$, such that first order arithmetic ('PA') proves $Sb^\bullet(\ulcorner a\urcorner,\ulcorner\phi\urcorner) = \ulcorner Sb(a,\phi)\urcorner$. I abbreviate by $\dot{x}$ a PA-representation of the function that maps a number n to its numeral $\overline{n}$. Quantification into the context $Sb^\bullet$ then is facilitated by quantification into this function, as in $\forall x\exists y\exists z\big(Sb^\bullet(\dot{x},\dot{y})\big) = z$. Occasionally, I will write $\ulcorner\phi(a)\urcorner$ for $Sb^\bullet(\ulcorner a\urcorner,\ulcorner\phi\urcorner)$.

11 Kripke [1975].
12 For simplicity, I will assume $\wedge$, $\forall$ and $\to$ to be defined in terms of the primitive symbols $\neg$, $\vee$ and $\exists$.
13 I assume that bound variables in ϕ are renamed if necessary.

Definition 23. Let $\mathcal{L}$ be the language of first order arithmetic $\mathcal{L}_0$ extended by 'η' and let ϕ be an $\mathcal{L}$-formula. We define its translation $(\phi)^\star$ by induction on the rank of ϕ.

If it is 0, then ϕ is a formula of the base language $\mathcal{L}_0$, or of the form $a\eta b$ and b is not $\ulcorner\psi\urcorner$ for some formula ψ. If $\phi \in \mathcal{L}_0$ then we set $(\phi)^\star = \phi$. If ϕ is of the form $a\eta b$, and b a variable, we set

$$(a\eta b)^\star = \begin{cases} TSb^\bullet(\ulcorner a\urcorner, \dot{b}) & \text{if } a \text{ is a closed term} \\ TSb^\bullet(\dot{a}, \dot{b}) & \text{if } a \text{ is a variable} \end{cases}$$

Finally, if b is a closed term but does not denote the code of some formula, let $(\phi)^\star$ be Tb.

Now assume that the rank of ϕ is $n+1$, and that we have defined $(\zeta)^\star$ for formulae ζ of rank $\leqslant n$. At this point, inside of the induction on rank, we run an induction on the syntactic complexity of ϕ. If ϕ is atomic, it is of the form $a\eta\ulcorner\zeta\urcorner$ for some formula ζ of rank n. We let

$$(a\eta\ulcorner\zeta\urcorner)^\star = \begin{cases} TSb^\bullet(\ulcorner a\urcorner, \ulcorner(\zeta)^{\star}\urcorner) & \text{if } a \text{ is a closed term} \\ TSb^\bullet(\dot{a}, \ulcorner(\zeta)^{\star}\urcorner) & \text{if } a \text{ is a variable} \end{cases}$$

Our induction hypothesis ensures $(\zeta)^\star$ to be defined. Now we set:

$$(\neg\phi)^\star = \neg(\phi)^\star$$

and proceed analogously for the other connectives and the quantifiers.

Using this translation we can define a theory in the language $\mathcal{L}$ as follows. Let $\mathfrak{N}(\mathrm{SK}_\infty)$ denote the standard model of arithmetic $\mathfrak{N}$ expanded by the least fixed point SK_∞ of Kripke's Strong Kleene jump.

Definition 24. $\mathrm{HSK} := \{\phi : \mathfrak{N}(\mathrm{SK}_\infty) \models_{\mathrm{SK}} (\phi)^\star\}$

I will speak of class theories using the following notation. The first letter 'H' indicates that we deal with a theory in a language containing 'η'.[14] Then follows a code denoting the analogous truth theory. In the present case, 'SK' denotes the theory of the least Strong Kleene fixed point model.

In the following, I will examine HSK and test it against the desiderata of section 3.1. For this, I connect with notions due to by Solomon Feferman.

14 Recall that in the Greek alphabet, 'H' is a capital 'η'.

Let $Cl(\ulcorner\phi\urcorner)$ be a meta-linguistic abbreviation of the formula $\forall y(y\eta\ulcorner\phi\urcorner \vee y\eta\ulcorner\neg\phi\urcorner)$.[15] The property expressed by '*Cl*' will play a central role in the following. Note that HSK contains $Cl(\ulcorner\phi\urcorner)$ just in case for every term a, the sentence $\big(\phi(a)\big)^\star$ has a classical truth value in the model $\mathfrak{N}(SK_\infty)$. A sentence is classical in the least fixed point, however, just in case it satisfies Kripke's formal definition of groundedness. Therefore, if we seek a theory of grounded classes, then we ought to be interested in which formulae satisfy *Cl*.

Moreover, if HSK contains $Cl(\ulcorner\phi\urcorner)$ then it contains $\forall x\big(x\eta\ulcorner\phi\urcorner \leftrightarrow \phi(x)\big)$, and therefore the ϕ-instance of comprehension. Due to this fact, I will say that a formula ϕ defines a class if $Cl(\ulcorner\phi\urcorner)$ holds in our theory, and will refer to *Cl* as the property of grounded class-hood.

Failure of grounded class-hood is identified fairly easily. $Cl(\ulcorner\phi\urcorner) \in$ HSK only if every for closed term a, the sentence $\big(\phi(a)\big)^\star$ is grounded. Hence, the formula '$x_0\eta x_0$' fails to define a class since its instance '$\ulcorner x_0\eta x_0\urcorner\,\eta\,\ulcorner x_0\eta x_0\urcorner$' is translated as an ungrounded *truth-teller*. Similarly, comprehension does not hold for the Russell formula $x\not\eta x$. This is how we block the paradox from page **??**.

It is good to know that the Russell formula does not define a class, but we would also like to know which formulae do so. More precisely, which formulae satisfy *Cl*, that is $\forall y(y\eta\ulcorner\phi\urcorner \vee y\eta\ulcorner\neg\phi\urcorner)$, over HSK? We can show that the theory contains all arithmetically definable classes, classes defined in terms of these, and so on. To render this precise, I introduce some terminology, again due to Feferman.[16]

Definition 25. Let ϕ and $\psi_0, \dots, \psi_n$ be $\mathcal{L}$-formulae with exactly one free variable. Call ϕ *elementary in* the ψ_i if (i) every atomic subformula in ϕ that contains 'η' is of the form $a\eta\ulcorner\psi_i\urcorner$ for some $i \leqslant n$; and (ii) in ϕ only atomic subformulae are negated.

A formula ϕ is *elementary* simpliciter if there are some ψ_i that ϕ is elementary in.

For example, $x\eta\ulcorner\psi\urcorner$ is elementary in ψ, as is $x\not\eta\ulcorner\psi\urcorner \vee \forall x\exists y(x = y + 1)$. The formula $\exists y(x\eta y)$, however, is not elementary, since it contains quantification into the range of η. This notion of elementarity allows us to give a sufficient condition on formulae ϕ for HSK to prove $Cl(\ulcorner\phi\urcorner)$.

Proposition 7. *For every* $\phi, \psi_0, ..., \psi_n \in$Fml *such that* ϕ *elementary in the* ψ_i*, if for every* $i \leqslant n$, $\mathrm{Cl}(\ulcorner\psi_i\urcorner) \in$ HSK *then*

$$\mathrm{Cl}(\ulcorner\phi\urcorner) \in \text{HSK}$$

15 See Feferman [1991], p. 28.

16 In the literature, 'elementary' usually applies to formulae of the base language. Feferman's concept is more general: elementary formulae may contain 'η'. Indeed, they are closed under iteration of η. However, if ϕ is elementary in the ψ_i then we know that in ϕ, class talk is confined to atomic formula of the form $a\eta\ulcorner\psi_i\urcorner$. Thus, ϕ can be viewed as a base-linguistic function of atomic formulae $a\eta\ulcorner\psi_i\urcorner$: it is *elementary in* the ψ_i (Feferman [1975b]).

Proposition 7 proves useful. The notion of elementarity is purely syntactic. Thus, proposition 7 allows us to sidestep the non-classical semantics of Kripke's model construction and examine the class theory HSK directly. For one, every formula of the base language is classically equivalent to an elementary formula, and the base language fragment of HSK is closed under classical logic. Thus, HSK provides comprehension for arithmetical formulae. Consequently, it also proves comprehension for formulae elementary in arithmetical formulae. And so on. For another, there are elementary formulae every instance of which is true by logic, such as the formula '$x = x$'. For other elementary formulae, every instance is false by logic. Hence, HSK has a universal and an empty class. Furthermore, every formula of the form '$a = x$', for any closed term a, is trivially elementary and defines a class over HSK. Thus, the HSK classes are closed under the singleton operation.

Further, the syntactic property of elementarity is closed under the connectives. Consequently, the HSK classes are closed under the Boolean operations. For every ϕ and ψ, HSK firstly contains $Cl(\ulcorner\phi\urcorner)$ just in case it contains $Cl(\ulcorner\neg\phi\urcorner)$. Secondly, if HSK contains $Cl(\ulcorner\phi\urcorner)$ and $Cl(\ulcorner\psi\urcorner)$ then it contains $Cl(\ulcorner\phi \wedge \psi\urcorner)$ and $Cl(\ulcorner\phi \vee \psi\urcorner)$. Recall that for every formula ϕ such that $Cl(\ulcorner\phi\urcorner) \in$ HSK, HSK contains the corresponding instances of class comprehension. Together with the observations just made, this implies that the classes of HSK are closed under complement, union and intersection. On this basis, HSK can be viewed as capturing the definitional idea of collection, as we would like our theory of grounded classes to do.

How does HSK perform with respect to the other desiderata? We would like our class theory to be closed under classical logic. How does HSK perform in this respect? Badly. Of course, the model $\mathfrak{N}(\mathrm{SK}_\infty)$ is partial and the set of sentences true in it is not closed under classical logic. Consequently, HSK is not either. Hence, the theory HSK does not satisfy our desideratum of classicality.

Fortunately, another theory of grounded truth is closed under classical logic: Burgess' theory KFB.[17] It axiomatizes the classical model $\mathfrak{N}(\mathrm{SK}^+_\infty)$, which we obtain from Kripke's partial model $\mathfrak{N}(\mathrm{SK}_\infty)$ by extending the anti-extension SK^-_∞ to the complement of the extension SK^+_∞ (the *closed off* fixed point model). Further, KFB is a theory of grounded truth as it extends the well-known theory KF by a schema to the effect that '$\forall x\,(T\,(x) \rightarrow \phi\,(x))$' is proved whenever $\phi(x)$ satisfies the left-to-right direction of the KF axioms. In this precise sense, KFB axiomatizes the *least* predicate closed under the KF axioms. Since these correspond to the inductive clauses of Kripke's Strong Kleene fixed point construction, KFB is well viewed as an axiomatization of the *least* such fixed point.

17 Burgess [2009] and Halbach [2011b], §17.

The theory KFB has various properties which we would expect of a theory of grounded truth. For example, it proves the truth-teller sentences to be neither true nor false.[18] However, some features of KFB hardly square with semantic groundedness. Most prominently, the theory proves the Liar sentence, although not its truth.[19] On this basis, it may be challenged how faithful KFB is to the idea of semantic groundedness.

I do not wish to take a stance in this debate. However, if closing off the least fixed point is incompatible with groundedness, then the desideratum of classicality can hardly be met. In this chapter, I explore the prospects of grounded class theory, and will conclude that they are limited. Thus, I should first make a good case on behalf of the friend of grounded class theory. Therefore, I will examine what would be available to her if we assumed that closing off the least fixed point is compatible with the idea of groundedness.

In sum, as I will argue that the prospects of grounded theories of classes are limited, it is fair to concede the legitimacy of closing off, since this is an assumption on behalf of my opponent. KFB axiomatizes the closed off least fixed point, and I will use it to obtain a theory of grounded classes.

Of course, I could also work with the complete theory of the closed off model $\mathfrak{N}(\mathrm{SK}^{+}_{\infty})$. Unlike it, however, KFB is an axiomatic theory of truth. For some authors, the axiomatic approach to truth has advantages over the semantical approach. Although I do not claim that much, I wish to show how to obtain class theories from axiomatic as well as semantical theories of truth. HSK was based on a semantical theory of truth. Therefore, it is the axiomatic theory KFB from which I derive a theory of grounded classes HKFB, in the following manner.

Definition 26. $\mathrm{HKFB} \coloneqq \{\phi : \mathrm{KFB} \vdash (\phi)^{\star}\}$

HKFB has all desirable properties of HSK and excels in various other respects. To begin with, HKFB, unlike HSK, is closed under classical logic. It satisfies the desideratum of classicality. What fragment of naive comprehension does HKFB prove? As with HSK, the definition of '*Cl*' implies that for every ϕ,

$$\mathrm{HKFB} \vdash Cl(\ulcorner\phi\urcorner) \to \forall x\big(x\eta\ulcorner\phi\urcorner \leftrightarrow \phi(x)\big) \tag{12}$$

Accordingly, the question again is: what formula does HKFB prove to have the property *Cl*? Above, I have found that in HSK, the set of formulae which define a class over HSK is closed under the connectives (proposition 7). The same holds for HKFB. Indeed, due to its classicality the theory proves the object-linguistic conditional.

18 Burgess (2009, §14).

19 KFB proves the theory KF+Cons [Halbach, 2011b, §§17.2,17.3], which proves $\neg T\ulcorner\lambda\urcorner$, for $\mathrm{PA} \vdash \lambda \leftrightarrow \neg T\ulcorner\lambda\urcorner$ (ibid., p. 217).

Proposition 8 (Cantini, 1996 9.7(ii)). *If ϕ is elementary in the ψ_i then*

$$\mathrm{HKFB} \vdash \mathrm{Cl}(\ulcorner\psi_1\urcorner) \wedge \ldots \wedge \mathrm{Cl}(\ulcorner\psi_n\urcorner) \rightarrow \mathrm{Cl}(\ulcorner\phi\urcorner) \tag{13}$$

Schema 13 is proved already by weaker theories, in particular the $\mathcal{L}$-theory HKF+Cons derived from the theory of truth KF+Cons.[20] Thus, proposition 8 is not optimal from a proof-theoretic point of view. However, HKF+Cons cannot be viewed as a theory of *grounded* classes since, unlike KFB, it is not intended as a axiomatization of the least, but of all consistent Strong Kleene fixed points. From the philosophical perspective taken in this chapter, therefore, the theory HKFB is of particular interest.

As a corollary to proposition 8, HKFB itself proves the same closure of class-hood that we observed, meta-theoretically, for HSK (p. 42). To this extent, HKFB captures the definitional idea of collection and satisfies the corresponding desideratum (p. 37).

The theories considered so far, HSK and HKFB, extend first order arithmetic by a theory of classes. However, we would like our theory of grounded classes to be applicable to arbitrary base theories. To some extent, this poses a problem to the present, derivative approach. Grounded theories of truth are almost all developed over arithmetic. This restriction is useful, but fortunately not essential. Occasionally, other bases are considered. Very recently, Kentaro Fujimoto examined the extension of ordinary set theory ZF by the truth axioms of KF.[21] It can be strengthened to a theory of grounded truth ZF+KFB. Translating the language of set theory plus 'η' into the language of truth over set theory, we obtain a theory of grounded classes on top of ZF. We can show that its classes, too, are closed under elementary definition.[22]

So far, the derivative theory HKFB has performed well with respect to our desiderata. However, HKFB disappoints in one important respect: it does not satisfy the desideratum of extensionality, as I will show in the next section. I will have to go into some detail. The reader who accepts, if only for the sake of the argument, that extensionality poses a problem to the derivative approach, may well skip the following and continue with section 3.4 where I develop a new, direct approach to an extensional theory of classes.

20 See Cantini [1996], p. 7, and footnote 19. Cantini provides further information about a system mutually interpretable with HKF+Cons. For example, he shows that it interprets Σ^1_1-AC [Cantini, 1996, p. 66].

21 Fujimoto [2012].

22 Based on the class-hood of elementary formulae, and generalizing a proof due to Feferman (1991), Fujimoto shows that his theory ZF+KF interprets iterations of NBG. For details, I refer the reader to Fujimoto's paper (2012).

3.3 DERIVATIVE THEORIES AND EXTENSIONALITY

I will begin by pointing out that even if distinct formulae ϕ, ψ define co-extensional classes, the theory HKFB is bound to contain $\ulcorner\phi\urcorner \neq \ulcorner\psi\urcorner$. Having noted this simple fact I will look more closely at what exactly is required for our theory of classes to satisfy the desideratum of extensionality. Based on this analysis, I will examine two routes that the friend of the derivative approach may take towards an extensional theory.

Firstly, I will discuss whether extensionality is at least achieved for the '='-free fragment of the language $\mathcal{L}$. This approach, however, puts undesirable limitations on our theory of classes.

Secondly, I pursue the thought that extensionality can be achieved by revising how we translate the language of classes into the language of truth. The idea is to translate $\ulcorner\phi\urcorner = \ulcorner\psi\urcorner$ as the statement that everything is a member of the class of the ϕs just in case it is a member of the class of the ψs. As simple as this thought is, it will require some additional machinery to implement it. However, even if we make the necessary assumptions, we will find the resulting theory of classes not to satisfy the desideratum of extensionality. I conclude that instead of further elaborating on the derivative approach, we ought to develop a theory of grounded classes directly.

Consider any two equivalent arithmetical formulae, for example $\rho =$'$x_0 = \overline{2}$' and $\sigma =$'$x_0 = \overline{1} + \overline{1}$'. By proposition 8, HKFB contains

$$Cl(\ulcorner\rho\urcorner) \wedge Cl(\ulcorner\sigma\urcorner) \wedge \forall x(x\eta\ulcorner\rho\urcorner \leftrightarrow x\eta\ulcorner\sigma\urcorner) \tag{14}$$

Whichever reasonable way we choose to arithmetize syntax, ρ and σ are assigned distinct Gödel numbers. By arithmetic alone, therefore, HKFB also contains

$$\ulcorner\rho\urcorner \neq \ulcorner\sigma\urcorner \tag{15}$$

Thus, *prima facie* our theory says that there are co-extensional but distinct classes.[23]

In view of this basic fact, let us take a step back and ask what is required for our theory of classes to satisfy the desideratum of extensionality. Recall the desideratum: we would like our theory to imply that the class of the ϕs is the class of the ψs just in case that everything is a member of the one if and only if it is a member of the other. On the present, derivative approach this means that we would like our class theory to say that the class of the ϕs is the class of the ψs iff the underlying theory of truth proves $\forall x\big(T\ulcorner(\phi)^\star(x)\urcorner \leftrightarrow T\ulcorner(\psi)^\star(x)\urcorner\big)$. This schema induces an equivalence relation E on the formulae in

23 This observation is a simple variant of known limitative results that apply to theories with comprehension for elementary formulae, or formulae of a similar syntactic property. See Gilmore [1974], Hinnion [1987], and (for a survey) Hinnion and Libert [2003].

Fml. Thus, HKFB satisfies the desideratum of extensionality only if two formulae that stand in this relation E, define one and the same class. By Leibniz' law, the class of the ϕs is identical to the class of the ψs only if one is indiscernible from the other. In the language $\mathcal{L}$, the class of the ϕs is denoted by the term $\ulcorner\phi\urcorner$. Therefore, HKFB satisfies extensionality only if for any two E-equivalent formulae ϕ, ψ, it finds $\ulcorner\phi\urcorner$ and $\ulcorner\psi\urcorner$ indiscernible. The fact that HKFB contains both (14) and (15) shows that this is not so.

Maybe we have been too demanding. What (15) shows is that $\ulcorner\phi\urcorner$ and $\ulcorner\psi\urcorner$ are discernible *qua* codes. However, these distinct codes may still stand for the same class. All that matters is the following. If ϕ and ψ stand in the relation E, $\ulcorner\phi\urcorner$ and $\ulcorner\psi\urcorner$ must not be *class-theoretically* discernible.

A natural way of rendering precise this thought is to consider the $\mathcal{L}$-fragment $\mathcal{L}^-$ without '='. Then, we ask whether it holds that for every ϕ and ψ of the original language $\mathcal{L}$, if ϕ bears E to ψ then $\ulcorner\phi\urcorner$ and $\ulcorner\psi\urcorner$ cannot be discerned within the fragment $\mathcal{L}^-$. More precisely, do we have that for every $\mathcal{L}^-$ formula ζ, the theory HKFB contains $\zeta(\ulcorner\phi\urcorner) \leftrightarrow \zeta(\ulcorner\psi\urcorner)$? No. Let ρ be as above, and let the number n be its code. Since KFB proves $T\ulcorner\ulcorner\rho\urcorner = \overline{n}\urcorner$, our derived theory of classes contains $\ulcorner\rho\urcorner\,\eta\,\ulcorner x_0 = \overline{n}\urcorner$. However, KFB also proves $\neg T\ulcorner\ulcorner\sigma\urcorner = \overline{n}\urcorner$, for σ as above. Therefore, HKFB contains $\ulcorner\sigma\urcorner \not\eta\, \ulcorner x_0 = \overline{n}\urcorner$; but '$y\,\eta\,\ulcorner x_0 = \overline{n}\urcorner$' is a formula of the '='-free fragment $\mathcal{L}^-$. Hence, $\ulcorner\rho\urcorner$ and $\ulcorner\sigma\urcorner$ are not even indiscernible with respect to this restricted language.

It may be objected that although '$y\,\eta\,\ulcorner x_0 = \overline{n}\urcorner$' is an $\mathcal{L}^-$-formula, it contains the code of an equation. When asking for indiscernibility, the thought goes, we ought to not only focus on '='-free formulae, but also disallow codes of formulae with '='. However, the proposed notion of what makes a formula *class theoretic* is excessively restrictive: it gives up on classes defined by formulae of the base language. Our theory of grounded classes over arithmetic would not be able to speak of arithmetical classes.

Fortunately, there is an alternative. The second route mentioned in the beginning of this section preserves class-definition in terms of '='. Recall that its idea is to translate $\mathcal{L}$-formulae into the language of truth in a *smart* way. In order to implement this idea, I need to modify the setting of the derivate approach in two respects.

Initially, it may be thought that we can translate $\ulcorner\phi\urcorner = \ulcorner\psi\urcorner$ in such a way that our theory proves this base language sentence whenever ϕ and ψ stand in the relation E. However, this would lead to a theory of classes that contradicts its own base theory. After all, for distinct formulae ϕ and ψ, $\ulcorner\phi\urcorner \neq \ulcorner\psi\urcorner$ is a theorem of rudimentary arithmetic.

In order to translate equations as statements of class-identity we need to disentangle the role of $\ulcorner\phi\urcorner$ as a number term and its role as standing for the formula ϕ. A natural way of doing so is to speak of the class of the ϕs by a new term $\hat{x}\phi$, and no longer rely on its

Gödel numeral $\ulcorner\phi\urcorner$. Formally, we extend the language $\mathcal{L}$ by a variable-binding, term-forming operator $\hat{\cdot}$ to the language $\mathcal{L}^\wedge$. We define the set of $\mathcal{L}^\wedge$-formulae and $\mathcal{L}^\wedge$-terms by simultaneous induction, such that $\hat{a}\phi$ is a term just in case a is a variable and ϕ an $\mathcal{L}^\wedge$-formula. $\hat{a}\phi$ has precisely the free variables of ϕ but for a. I will refer to a term $\hat{a}\phi$ as an 'abstraction-term'.

In the remainder of this chapter, I will work with this extended language $\mathcal{L}^\wedge$. In addition to making the syntax of class theory more perspicuous, I have thus carried out the first of the two changes that together will allow me to implement the *smart* way of translating class talk into truth talk.

The second modification concerns how the new, smart translation is defined. The notion of rank from the previous section is developed into the concept of a formula's *degree*. It is defined by an induction on ϕ's syntactic complexity, such that $a = b$ has degree 0 iff a or b is not an abstraction term $\hat{x}\phi$, and the degree of $\hat{x}\phi = \hat{x}\psi$ is one greater than that ϕ or ψ, whichever greater. The degree of a formula $a\eta b$ is defined just like its rank, and so is the degree of a syntactically complex formula. Since the term $\hat{x}\phi$ cannot occur in the formula ϕ, the relation "... is of lower degree than ..." is well-founded.

Having extended the language by abstraction terms, and using the new concept of degree, we are now in a position to implement the smart translation of our language of class theory into the language of truth. Let $(\phi)^\dagger$ be defined by an induction on the degree of ϕ. We proceed analogously to how we defined $(\phi)^\star$, except that now, $\hat{x}\zeta = \hat{x}\xi$, is translated as

$$\forall x\big(T\ulcorner(\zeta)^\dagger(x)\urcorner \leftrightarrow T\ulcorner(\xi)^\dagger(x)\urcorner\big) \tag{16}$$

Since the language of truth does not have abstraction terms, in other contexts $\hat{x}\zeta$ is represented by the code of the translation of ζ. In particular, $\hat{x}\zeta\eta\,\hat{x}\xi$ is translated as $T\ulcorner(\xi)^\dagger\ulcorner(\zeta)^\dagger\urcorner\urcorner$. Shortly, we will find that this is a problem.

For the new translation $(\cdot)^\dagger$, the schema (16) defines over KFB a new equivalence relation $E_\dagger$. Let †HKFB be the theory of classes derived from the theory of truth KFB, through the smart translation $(\cdot)^\dagger$.

$$\dagger\text{HKFB} \coloneqq \{\phi : \text{KFB} \vdash (\phi)^\dagger\} \tag{17}$$

By definition, whenever ϕ and ψ stand in the corresponding equivalence relation $E_\dagger$, †HKFB contains $\hat{x}\phi = \hat{x}\zeta$ just in case it contains $\hat{x}\psi = \hat{x}\zeta$, for every ζ. Thus, being smart about translating formulae $\hat{x}\phi = \hat{x}\psi$, we have come closer to our goal of an extensional theory of classes derived from KFB.

Have we succeeded? †HKFB would satisfy the desideratum of extensionality if whenever ϕ and ψ stand in the relation $E_\dagger$, we have that for every formula ζ, †HKFB contains $\zeta(\hat{x}\phi)$ just in case $\zeta(\hat{x}\psi)$. However, this is not the case. I will state the problem first, and then explain how it is rooted in the definition of $(\cdot)^\dagger$.

Let ρ, σ be as above. Since KFB proves $\forall x\big(\rho(x) \leftrightarrow \sigma(x)\big)$, we have that †HKFB contains $\hat{x}\rho = \hat{x}\sigma$. Now, let m be the Gödel code of $(\rho)^\dagger$, such that PA$\vdash$ $\ulcorner(\rho)^{\dagger}\urcorner = \overline{m}$. We have that

$$\mathrm{KFB} \vdash T^{\ulcorner\ulcorner}(\rho)^{\dagger\urcorner} = \overline{m}^{\urcorner} \wedge \neg T^{\ulcorner\ulcorner}(\sigma)^{\dagger\urcorner} = \overline{m}^{\urcorner} \tag{18}$$

Consequently, our derived theory of classes is bound to contain the following.

$$\hat{x}\rho = \hat{x}\sigma \wedge \hat{x}\rho\,\eta\,\hat{x}(x = \overline{m}) \wedge \hat{x}\sigma\,\not\eta\,\hat{x}(x = \overline{m}) \tag{19}$$

Thus, it is not the case that formulae that stand in the equivalence relation $E_\dagger$ are indiscernible over the derived theory of classes †HKFB. Therefore, although being based on a smart translation, the theory †HKFB does not satisfy extensionality.

The reason is that if an abstraction term $\hat{x}\phi$ occurs on the left-hand side of 'η', it is translated as the Gödel code of $(\phi)^\dagger$. More precisely, the formula $\hat{x}\phi\,\eta\,\hat{x}\zeta$ is translated as $T^{\ulcorner}(\zeta)^{\dagger}(^{\ulcorner}(\phi)^{\dagger\urcorner})^{\urcorner}$. However, this treatment of atomic formulae with 'η' undoes what we have gained by being smart about '='. Since, all information is lost as to what other formulae bear $E_\dagger$ to ϕ when translating $\hat{x}\phi\,\eta\,\hat{x}\zeta$ as $T^{\ulcorner}(\zeta)^{\dagger}(^{\ulcorner}(\phi)^{\dagger\urcorner})^{\urcorner}$. As we have just seen, there are formulae involving 'η' for which this information matters.

Although the translation $(\cdot)^\dagger$ is smarter than our original translation function $(\cdot)^\star$, it is not smart enough. In order to make the fact that ϕ and ψ stand in the relation $E_\dagger$ ensure the indiscernibility of $\hat{x}\phi$ and $\hat{x}\psi$, not only $\hat{x}\phi = \hat{x}\psi$, but also $\hat{x}\phi\,\eta\,b$ needs to be translated in a manner that takes into account what other formulae are co-extensional with ϕ.

One way of implementing this *smarter* approach is by translating $\hat{x}\phi\,\eta\,b$ and $\hat{x}\psi\,\eta\,b$ as the same formula of the language of truth, if ϕ and ψ are co-extensional.[24] We may for example represent a formula ϕ by the lexicographically least $\mathcal{L}$-formula ψ that bears $E_\dagger$ to ϕ. That is, let $[\phi]_\dagger$ be the lexicographically least formula ψ such that

$$\mathrm{KFB} \vdash \forall x\big(T^{\ulcorner}(\phi)^{\dagger}(x)^{\urcorner} \leftrightarrow T^{\ulcorner}(\psi)^{\dagger}(x)^{\urcorner}\big) \tag{20}$$

Thus, if distinct formulae ϕ and ψ stand in the relation $E_\dagger$, they are both represented by the same formula $[\phi]_\dagger$. Using this representation, we can define a new translation $(\cdot)^\ddagger$ just like $(\cdot)^\dagger$ except that formulae $\hat{x}\phi\,\eta\,\hat{x}\psi$ are now translated as

$$T^{\ulcorner}(\psi)^{\ddagger\ulcorner}([\phi]_\dagger)^{\ddagger\urcorner\urcorner} \tag{21}$$

Let ‡HKFB be the $\mathcal{L}$-theory derived from KFB through this new, smarter translation $(\cdot)^\ddagger$. It can be shown that whenever ϕ bears $E_\dagger$ to ψ then ‡HKFB finds $\hat{x}\phi$ and $\hat{x}\psi$ to be indiscernible.

24 I thank Sam Roberts for pointing me into this direction.

However, the smarter translation $(\cdot)^{\ddagger}$ itself induces a new equivalence relation $E_{\ddagger}$. It can also be shown that there are formulae that stand in this new relation $E_{\ddagger}$, but are not indiscernible in the theory ‡HKFB.[25] Therefore, ‡HKFB does not satisfy extensionality, either. What would be needed is a translation t such that $\hat{x}\phi\, \eta\, \hat{x}\psi$ is translated as

$$T^{\ulcorner}(\psi)^{t\ulcorner}([\phi]_t)^{t\urcorner\urcorner} \tag{22}$$

Unfortunately, it is not obvious that such a mapping can be defined. In order for it to make sense to speak of $[\phi]_t$, t must already be defined not only for ϕ, but for every other formula, too, including $\hat{x}\phi\, \eta\, \hat{x}\psi$ itself.

Independently of technical details, there is a philosophical reason not to pursue this route further. The more elaborate our translation, the less reason we have to think that the groundedness of our truth theory carries over to our derived theory of classes. After all, groundedness is a philosophical notion, and syntactic translations do not generally preserve philosophical significance. Moreover, the resources we invest in setting up a sophisticated translation we may as well use to develop a theory of class directly. I will do so in the next section.

3.4 GROUNDED MEMBERSHIP AND GROUNDED IDENTITY

In this section, I will develop a theory of grounded classes without the detour through truth theory characteristic of the derivative approach. My approach is semantical. I will define a model for the language with 'η' and abstraction terms '$\hat{x}\phi$'. The basic idea is as follows. I will extend a given base model by a relation of class membership and a relation of class identity. These relations are defined inductively using jump operators that turn satisfaction in the given model into a new model. Together, these operators reach a least fixed point. In effect, I define *grounded* membership and *grounded* identity analogously to how Kripke defines a predicate of grounded truth,

The model construction will combine elements of Penelope Maddy's theory as well as unpublished work by Hannes Leitgeb, and Leon Horsten and Øystein Linnebo.[26] However, I will go beyond this extant work.

Maddy approaches a theory of grounded classes directly, and semantically. On the basis of set theory, she constructs a model for class theory using a monotone operator similar to my membership jump $\mathcal{H}$ below. Leitgeb, in an unpublished note from 2004, proceeds similarly. In the work of both authors, class identity is defined in terms of grounded membership. The class of the ϕs is the class of the ψs if $\forall x(x\eta\hat{y}\phi \leftrightarrow x\eta\hat{y}\psi)$ holds in the least fixed point model. Effectively,

25 For example, the formulae $\hat{x}\rho\, \eta\, y$ and $\hat{x}\sigma\, \eta\, y$.
26 See Maddy [1983, 2000].

class identity is dealt with as in the theory †HKFB of the previous section (p. 47). Consequently, Maddy's theory likewise fails to satisfy the desideratum of extensionality, as noted by herself.[27] The following is intended as one way of doing better.

In order to satisfy the desideratum of arbitrary bases, I will outline the construction for any first order base language $\mathcal{L}_0$, and any $\mathcal{L}_0$-structure $\mathfrak{M}$. For simplicity, I will assume that the base language contains a constant for every object of the base domain.[28] Given the base model $\mathfrak{M}$, I proceed as follows.

Firstly, I extend the base domain M by the set *Abs* of the closed abstraction terms. In the extended model, a closed term $\hat{x}\phi$ will denote itself. These terms will be the large pool of objects from which we will abstract the classes of our theory. It is useful to think of the terms as *proto*-classes, or class-candidates. For some terms $\hat{x}\phi$, the model below will validate $Cl(\hat{x}\phi)$ – the guiding question will be how many candidates are thus elected.[29]

Secondly, I add to the base model $\mathfrak{M}$ a membership relation H and a relation of class identity I. In the extended model $\mathfrak{M}(I, H)$, the new relation symbol 'η' will be interpreted by H.[30] Accordingly, H relates objects of the full domain $M \cup Abs$ to proto-classes. I extends plain identity on the base domain M by a relation between proto-classes: $I \subseteq ID_M \cup Abs \times Abs$. Intuitively, I extends identity in the base model by *class identity*. Accordingly, in the extended model $\mathfrak{M}(I, H)$, '=' will be interpreted by the relation I, such that, for example,

$$\mathfrak{M}(I, H) \models \hat{x}\phi = \hat{x}\psi \Leftrightarrow \langle \hat{x}\phi, \hat{x}\psi \rangle \in I \tag{23}$$

My goal is a specific model $\mathfrak{M}(I, H)$, a model for a theory of grounded classes. I will define a grounded membership relation H and a grounded identity relation I, analogous to how Kripke defined a predicate of grounded truth.

My construction is based on two operators $\mathcal{I}$ and $\mathcal{H}$. Each takes one identity and one membership relation, but they differ in their output. $\mathcal{I}$, on the one hand, outputs an identity relation. I will speak of it as the 'identity jump'. $\mathcal{H}$, on the other hand, is a 'membership jump': it gives a membership relation.

There are various ways in which such jumps may be defined. I choose the *supervaluational* method, for two reasons. Firstly, doing so I explore an area not considered by Maddy.[31] Secondly, I will eventually formulate a challenge to the friend of grounded classes. Therefore, I should first make a good case on her behalf. I will argue that

27 Maddy (2000), p. 305.
28 This is not the case if we work with the language of set theory. Here, we can either, as Fujimoto does, extend the language of set theory by new constants or work not with formulae, but with pairs of a formula and parameters. See Fujimoto [2012].
29 Recall that '$Cl(\hat{x}\phi)$' abbreviates the $\mathcal{L}$-formula $\forall y(y\eta\hat{x}\phi \vee y\eta\hat{x}\neg\phi)$.
30 Recall that 'H' here is the capital Greek letter eta.
31 Maddy [1983].

the present, direct approach is unsatisfactory because it makes many natural candidates for class comprehension fail. More precisely, for many formulae ϕ that intuitively ought to define a class, the model does not validate either $a\eta\hat{x}\phi$ or $a\eta\hat{x}\neg\phi$ for every closed term a. Therefore, it is apposite to choose a semantics that maximizes the amount of sentences with classical truth value. Supervaluation fits this bill.

The basic idea behind supervaluation is to consider a range of candidate interpretations of '=' and 'η', determine which object o satisfies which formula ϕ in all these models and add $\langle o, \hat{x}\phi \rangle$ to the membership relation.[32] Analogously, we add $\langle \hat{x}\phi, \hat{x}\psi \rangle$ to the identity relation if ϕ and ψ are co-extensional in all models $\mathfrak{M}(J, K)$, for J, K extending I, H. Since my interest is in relations J that are candidates for class identity, I will restrict my attention to *equivalence* relations that are *coherent* in the sense that if $\langle o, p \rangle \in J$ then for every formula ϕ $\langle \hat{x}\phi(x, \overline{o}), \hat{x}(x, \overline{p}) \rangle \in J$. Further, since my goal is a relation of class membership that respects class identity, I focus on pairs J, K such that K respects J: for every o, p, if $\langle o, p \rangle \in J$ then for every q, $\langle o, q \rangle \in K$ if and only if $\langle p, q \rangle \in K$, and $\langle q, o \rangle \in K$ if and only if $\langle q, p \rangle \in K$. Below, this will allow me to show that grounded membership respects grounded identity, which in turn ensures the resulting theory to satisfy the desideratum of extensionality.

The more extensions we consider, the less pairs $\langle o, \hat{x}\phi \rangle$ will there be such that o satisfies ϕ in all of them. Thus, the more extensions are considered, the less new information is added to the given relations of identity and membership. Hence, the more extensions are taken into account, the weaker our resulting theory will be. Therefore, usually further conditions are imposed on the range of extensions. The more restrictive such an admissibility condition, the more terms $\hat{x}\phi$ will be such that for every o either $\langle o, \hat{x}\phi \rangle$ or $\langle o, \hat{x}\neg\phi \rangle$ is added. Thus, which condition is chosen partly determines how many instances of class comprehension are satisfied.

Exploring the prospects of grounded class theory, I wish to test the best possible case for such a theory. For my model construction, I therefore choose the strongest admissibility condition available from the literature. In the variant of Kripke's jump operator due to Andrea Cantini, an extension is admissible if and only if it is consistent.[33] Its least fixed point exceeds all other supervaluational theories in the literature.[34] Accordingly, I will use jumps that quantify over consistent extensions only.[35]. In sum, a pair J, K is an *admissible* extension of I, H

32 I use letters from the middle of the Roman alphabet ('n', 'o' etc.) as variables for objects of the extended domain $M \cup Abs$.

33 Cantini (1990) p. 250.

34 For example, Cantini's theory contains the sentences $\neg T \ulcorner \lambda \urcorner \vee \neg T \ulcorner \lambda \urcorner$, for liar sentences λ.

35 A membership relation H is consistent just in case there is no ϕ such that for any o, both $\langle o, \hat{x}\phi \rangle$ and $\langle o, \hat{x}\neg\phi \rangle$ are in H. An identity relation I is consistent if there is no ϕ such that for any o, both $\langle \hat{x}\phi, o \rangle \in I$ and $\langle \hat{x}\neg\phi, o \rangle \in I$

(in symbols: $I, H \Subset J, K$), if and only $I \subseteq J$, $H \subseteq K$, I is a *coherent* equivalence relation, K respects J, and they are both consistent.

I will now define the identity jump $\mathcal{I}$ and the membership jump $\mathcal{H}$. Firstly, the intuitive idea underlying the identity jump $\mathcal{I}$ is the following. $\mathcal{I}$ takes an identity relation I and a membership relation H and identifies all pairs $\langle \hat{x}\phi, \hat{x}\psi \rangle$ such that ϕ and ψ are co-extensional in all admissible extensions of the model $\mathfrak{N}(I, H)$.[36]

Definition 27 (Identity Jump)**.**

$$\mathcal{I}(I, H) = \Big\{ \langle \hat{x}\phi, \hat{x}\psi \rangle : \forall J \forall K \Big(I, H \Subset J, K \Rightarrow \mathfrak{M}(J, K) \models \forall x \big(\phi(x) \leftrightarrow \psi(x) \big) \Big) \Big\}$$

I now turn to the *membership* jump $\mathcal{H}$. Its definition is based on the following idea. Given an identity relation I and a membership relation H, $\mathcal{H}$ outputs just the pairs $\langle o, \hat{x}\phi \rangle$ such that o satisfies the formulae ϕ in all models compatible with the identity and membership facts encoded in I and H. This intuitive idea is implemented by the supervaluational method which I have used already to obtain the identity jump $\mathcal{I}$. In order to identify the right pairs $\langle o, \hat{x}\phi \rangle$ we consider all admissible extensions of the pair I, H.[37]

Definition 28 (Membership Jump)**.**

$$\mathcal{H}(I, H) = \big\{ \langle o, \hat{x}\phi \rangle : \forall K \forall J \big(I, H \Subset J, K \Rightarrow \mathfrak{M}(J, K) \models \phi(\overline{o}) \big) \big\}$$

I record some useful facts as to how $\mathcal{I}$ and $\mathcal{H}$ interact. Firstly, for any identity relation I and membership relation H, $\mathcal{I}(I, H)$ is an identity relation and $\mathcal{H}(I, H)$ is a membership relation. Secondly, if I and H are consistent, then so are $\mathcal{I}(I, H)$ and $\mathcal{H}(I, H)$. Finally, $\mathcal{I}(I, H)$ is an equivalence relation. Note, however, that neither is $\mathcal{H}(I, H)$ ensured to respect $\mathcal{I}(I, H)$, nor $\mathcal{I}(I, H)$ to be coherent.

Identity jump $\mathcal{I}$ and membership jump $\mathcal{H}$ together induce an operator on the pairs I, H. This operator is monotone with respect to the ordering of one pair of relations being extended pointwise by another. Therefore, it has a least fixed point IH_∞. We can show the following key fact.

Lemma 15. IH_∞ *is admissible extension of itself.*

Proof. Firstly, of course, I_∞ and H_∞ extend themselves. Secondly, we have already observed that $I_\infty = \mathcal{I}(IH_\infty)$ is an equivalence relation (p. 52). Thirdly, we need to show that H_∞ respects I_∞, i.e.

A. for every $o, p \in M \cup Abs$, if $\langle o, p \rangle \in I_\infty$ then

36 In an unpublished manuscript, Leon Horsten and Øystein Linnebo use a similar jump to construct a model of Frege's Basic Law V. However, they keep the underlying second order logic predicative, as in Heck [1996]. In effect, their work corresponds to using $\mathcal{I}$ with a fixed membership relation H that captures satisfaction of base language formulae in the base model.

37 Recall that o has a name in our language $\mathcal{L}^\wedge$, that I will denote by '$\overline{o}$'.

1. for all $q \in M \cup Abs$, $\langle q, o \rangle \in H_\infty$ iff $\langle q, p \rangle \in H_\infty$.
2. for all $q \in M \cup Abs$, $\langle o, q \rangle \in H_\infty$ iff $\langle p, q \rangle \in H_\infty$.

Finally, we need to show the coherence of I_∞, that is,

B. for every $o, p \in M \cup Abs$, if $\langle o, p \rangle \in I_\infty$ then for all formulae $\zeta(z, x)$ with exactly the free variables displayed, $\langle \hat{z}\zeta(z, \overline{o}), \hat{z}\zeta(z, \overline{p}) \rangle \in I_\infty$

For simplicity, I focus on the case discussed in the main text, the fixed point over the natural numbers $\mathfrak{N}$.

(A1) If o or p is a base domain object, that we know not to be in the range of H_∞, the claim is vacuously true. So let o be $\hat{x}\phi$ and p be $\hat{x}\psi$ for some ϕ, ψ. By the definition of $\mathcal{H}$ we know that they have exactly one free variable x.

Assume that $\langle \hat{x}\phi, \hat{x}\psi \rangle \in I_\infty$, such that

$$\forall J \forall K \Big(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J, K) \models \forall z \big(\phi(z) \Leftrightarrow \psi(z) \big) \Big) \tag{24}$$

Let q be any object. If $\langle q, \hat{x}\phi \rangle \in H_\infty$ then $\phi(\overline{q})$ holds at every admissible extension. By (24) and logic, we have $\mathfrak{N}(J, K) \models \psi(\overline{q})$ for every admissible extension J, K. Hence $\langle q, \hat{x}\psi \rangle \in H_\infty$, as desired. And analogously vice versa.

(A2) The claim is vacuously true unless q is a closed abstraction term $\hat{z}\zeta$. Since IH_∞ is a fixed point, it suffices to show that

$$\forall J \forall K \big(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J, K) \models \zeta(\overline{o}) \big) \Longleftrightarrow \forall J \forall K \big(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J, K) \models \zeta(\overline{p}) \big)$$

I show the left-to-right direction, the other is just analogous, swapping 'p' and 'o'. So assume the antecedent, and let J, K be any admissible extension of IH_∞. I show that $\mathfrak{N}(J, K) \models \zeta(\overline{p})$ by induction on the positive complexity of ζ.[38]

Firstly, ζ of the form $z = \overline{r}$ or $z \neq \overline{r}$, for some r, are taken care of by the *transitivity* of J together with our assumption that $\langle o, r \rangle \in I_\infty \subseteq J$. Secondly, let ζ be of the form $\hat{y}\xi(y, z) = \overline{r}$ or $\hat{y}\xi(y, z) \neq \overline{r}$. Since we assume that $\hat{y}\xi(y, \overline{o}) = \overline{r}$ holds in every admissible extension of $\mathfrak{N}(IH_\infty)$, the *coherence* of J ensures that $\langle \hat{y}\xi(y, \overline{p}), r \rangle \in J$ respectively $\langle \hat{y}\xi(y, \overline{p}), r \rangle \notin J$, and $\mathfrak{N}(J, K) \models \hat{y}\xi(y, \overline{p}) = \overline{r}$, as desired.

Thirdly, let ζ be of the form $z \,\eta\, a$ or $a \,\eta\, z$, respectively their negations. Then, $\mathfrak{N}(J, K) \models \zeta(\overline{p})$ follows from our assumption that $\mathfrak{N}(J, K) \models \zeta(\overline{o})$ and the fact that K respects J, which contains $\langle o, p \rangle$.

Finally, if $\zeta(\overline{o})$ is an η-literal such that $\overline{o}$ occurs within an abstraction term $b(x)$, we observe that the coherence of J ensures that $\langle b(\overline{o}), b(\overline{p}) \rangle \in$

38 See definition 15.9 in Halbach [2011b]. Careful examination shows that the attempt to prove the claim by induction on regular syntactic complexity breaks down at the induction step, at the clause for negations.

J.[39] Then, $\mathfrak{N}(J, K) \models \zeta(\overline{o})$ implies $\mathfrak{N}(J, K) \models \zeta(\overline{p})$ by the fact that K respects J.

At the induction step, we exploit the induction hypothesis. For example, let $\zeta(z)$ be of the form $\exists x(\xi(x, z))$. Then for some object of the domain q, $\mathfrak{N}(J, K) \models \xi(\overline{q}, \overline{o})$. Now, $\xi(\overline{q}, z)$ is of lower complexity than $\zeta(z)$. Hence, our induction hypothesis ensures that $\mathfrak{N}(J, K) \models \xi(\overline{q}, \overline{p})$. Consequently, $\mathfrak{N}(J, K) \models \exists x(\xi(x, \overline{p}))$, as desired.

(B) Again, by the fixed point character of IH_∞, it suffices to show that

$$\forall J \forall K \Big(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J, K) \models \forall z(\zeta(z, \overline{o}) \leftrightarrow \zeta(z, \overline{p})) \Big)$$

So let J, K be such that $IH_\infty \Subset J, K$, and let q be any object of the domain. We show $\mathfrak{N}(J, K) \models \zeta(\overline{q}, \overline{o}) \leftrightarrow \zeta(\overline{q}, \overline{p})$ by induction on the complexity of ζ. Recall that by the definition of the identity jump ζ is ensured to have exactly two free variables. I confine myself to the left-to-right direction as again, the other direction is just analogous.

We reason much like in the case of (A2). If $\zeta(\overline{q}, \overline{o})$ is of the form $\overline{q} = \overline{o}$ or $\hat{x}\xi(x, \overline{q}) = \overline{o}$ for some ξ, the claim follows from the transitivity of J. If it is of the form $\hat{y}\xi(y, \overline{o}) = \overline{q}$ we recall that J is coherent and contains $\langle o, p \rangle$. Finally, for atomic formulae containing η we note that K respects the coherent J, as before. The induction step is taken care of by the induction hypothesis and logic. For example, if $\zeta(\overline{q}, \overline{o})$ is of the form $\exists x \xi(x, \overline{q}, \overline{o})$ we reason as at the end of the argument for (A2). □

I will denote the identity relation of the fixed point pair IH_∞ by 'I_∞', and the membership relation by 'H_∞'. Note, however, that the interplay of the operators $\mathcal{I}$ and $\mathcal{H}$ is essential. It can be shown that I_∞, which is obtained starting from the empty identity and the empty membership relation, is distinct from the least fixed point of $\mathcal{I}$, for the empty membership relation.

I now examine what theory of classes this model construction provides. For a fair comparison with the theories of the previous sections, I focus on first-order arithmetic as our base theory. Thus, we extend the standard model of arithmetic $\mathfrak{N}$ by the least fixed point pair of relations IH_∞, obtained on the basis of arithmetic. The complete theory of this model I call 'HC', since the construction is based on the admissibility condition of *consistency*.

Definition 29.

$$HC := \{\phi : \mathfrak{N}(IH_\infty) \models \phi\}$$

I examine HC against the desiderata from section 3.1. Firstly, $\mathfrak{N}(IH_\infty)$ is a classical model. Hence, HC meets the desideratum of classicality. So did the derivative theory HKFB considered above. Unlike in

39 Our definition of the identity jump operator $\mathcal{I}$ ensures ζ to have exactly one free variable, which in this case implies that $b(x)$, too, has just the free variable displayed.

HKFB, however, every classically tautological formula defines a class over HC. In this precise sense, HC may be viewed as being more classical than HKFB. Of course, this additional degree of classicality is paid for. For example, it is not the case that $x\eta\hat{y}\phi$ or $x\eta\hat{y}\psi$ whenever $x\eta\hat{y}(\phi \vee \psi)$. This fact is due to the choice of supervaluational operators, and corresponds to the failure of compositionality in supervaluational truth theory. However, in the present class-theoretic context I consider it less problematic. We are interested not in single sentences of the form $x\eta\hat{x}(\phi \vee \psi)$, but in formulae $\phi \vee \psi$ of which we know that they define classes. And in this respect, a supervaluational class theory is not inferior to a closed-off Strong Kleene theory such as HKFB – neither proves downwards closure of classes under the connectives, e.g. $Cl\big(\hat{x}(\phi \vee \psi)\big) \to Cl(\hat{x}\phi) \vee Cl(\hat{x}\psi)$.

Secondly, class theories should stand to the *definitional* idea of collection as standard set theory stands to the *combinatorial* idea. How does the theory HC do with respect to this desideratum? A natural sharpening of the definitional idea was that classes make up a Boolean algebra. Brief reflection on the fixed point character of the model $\mathfrak{N}(\mathrm{IH}_\infty)$ and its classicality shows that the theory HC is closed under complement, union, intersection and iteration of membership.[40]

Proposition 9.

$$\begin{aligned}\mathfrak{N}(\mathrm{IH}_\infty) \models \forall x \forall y \Big(\mathrm{Cl}(x) \wedge \mathrm{Cl}(y) \to\, & \exists z\big(\mathrm{Cl}(z) \wedge \forall w(w\eta z \leftrightarrow r\not\eta x)\big)\\ & \wedge \exists z\big(\mathrm{Cl}(z) \wedge \forall w(w\eta z \leftrightarrow r\,\eta\, x \vee w\eta y)\big)\\ & \wedge \exists z\big(\mathrm{Cl}(z) \wedge \forall w(w\eta z \leftrightarrow r\,\eta\, x \wedge w\eta y)\big)\\ & \wedge \mathrm{Cl}\big(\hat{z}(z\eta x)\big)\Big)\end{aligned}$$

It is easy to see that every base language formula ϕ defines a class.[41] More precisely, for every $\mathcal{L}_0$-formula ϕ with a single free variable we have that

$$\mathfrak{N}(\mathrm{IH}_\infty) \models Cl(\hat{x}\phi) \tag{25}$$

Thus, HC is a theory of classes based on first-order arithmetic and closed under natural operations. This makes HC a good candidate for a formal theory of definitional collections.

Thirdly, the desideratum of extensionality was met by none of the derivative theories. HC performs considerably better in this respect. On the one hand, since I_∞ is a fixed point of the identity jump $\mathcal{I}$, two terms $\hat{x}\phi$ and $\hat{x}\psi$ stand in the relation I_∞ just in case they are co-extensional in $\mathfrak{N}(\mathrm{IH}_\infty)$.

40 Propositions 10 and 11 rely on lemma 15.

41 For $o \in Abs$, $\mathfrak{N}(\mathrm{IH}_\infty) \models \neg\phi(\overline{o})$ such that $\langle o, \hat{x}\neg\phi\rangle \in \mathrm{H}_\infty$; and for every $o \in \omega$, $\mathfrak{N} \models \phi(\overline{o}) \vee \neg\phi(\overline{o})$ such that $\langle o, \hat{x}\phi\rangle$ or $\langle o, \hat{x}\neg\phi\rangle$ enters the membership relation at the very first stage.

Proposition 10. *For all* $o, p \in \mathrm{Abs}$,

$$\mathfrak{N}(\mathrm{IH}_\infty) \vDash \overline{o} = \overline{p} \leftrightarrow \forall z(z\eta\overline{o} \leftrightarrow z\eta\overline{p})$$

Proof. The left-to-right direction of the claim follows directly from the fact that H_∞ respects I_∞ (lemma 15). For the right-to-left direction, assume that for every q, $\langle q, \hat{x}\phi\rangle \in H_\infty$ iff $\langle q, \hat{x}\psi\rangle \in H_\infty$. Hence, $\forall J \forall K\big(\mathrm{IH}_\infty \Subset J, K \Rightarrow \mathfrak{N}(J,K) \vDash \phi(\overline{q})\big)$ iff $\forall J \forall K\big(\mathrm{IH}_\infty \Subset J, K \Rightarrow \mathfrak{N}(J,K) \vDash \psi(\overline{q})\big)$ By logic, $\forall J \forall K\Big(\mathrm{IH}_\infty \Subset J, K \Rightarrow (\mathfrak{N}(J,K) \vDash \phi(\overline{q}) \Leftrightarrow \mathfrak{N}(J,K) \vDash \psi(\overline{q})\Big)$. Since this holds for every q, $\forall J \forall K\Big(\mathrm{IH}_\infty \Subset J, K \Rightarrow \mathfrak{N}(J,K) \vDash \forall z\big(\phi(z) \leftrightarrow \psi(z)\big)\Big)$. Hence, $\langle \hat{x}\phi, \hat{x}\psi\rangle \in I_\infty$, as desired.

□

On the other hand, every two class terms that stand in the relation I_∞ are also ensured to be *indiscernible* in the model $\mathfrak{N}(\mathrm{IH}_\infty)$.[42] I abbreviate a list of variables $x_0, \dots, x_n$ as '$\overrightarrow{x_n}$'.

Proposition 11. *For every* $o, p \in M \cup \mathrm{Abs}$, *if* $\langle o, p\rangle \in I_\infty$ *then we have that for every* $\mathcal{L}^\wedge$*-formula* $\phi(\overrightarrow{x_{n+1}})$,

$$\mathfrak{N}(\mathrm{IH}_\infty) \vDash \forall \overrightarrow{x_n}\big(\phi(\overline{o}, \overrightarrow{x_n}) \leftrightarrow \phi(\overline{p}, \overrightarrow{x_n})\big)$$

Proof. Let $\langle o, p\rangle \in I_\infty$. A basic theorem of model theory says that if two objects are indiscernible in a first-order model $\mathfrak{M}$ in terms of the primitive relation symbols of the $\mathfrak{M}$ signature, then they are indiscernible in $\mathfrak{M}$ with respect to every formula ϕ of this language.[43] It is proved by induction on the complexity of ϕ.

The following is a slight modification of that standard proof, for the non-standard language $\mathcal{L}^\wedge$ with its abstraction terms. It has two primitive relation symbols, '=' and 'η'. Therefore, we have to show that every two o and p that stand in the relation I_∞ are indiscernible in terms of '=' and 'η'. Since $\overline{o}$ and $\overline{p}$, however, may occur within open abstraction terms, six cases need to be distinguished.

I. $\mathfrak{N}(\mathrm{IH}_\infty) \vDash \forall \overrightarrow{x_n}\big(a(\overrightarrow{x_n})\,\eta\,\overline{o} \leftrightarrow a(\overrightarrow{x_n})\,\eta\,\overline{p}\big)$, for every $a(\overrightarrow{x_n})$.

II. $\mathfrak{N}(\mathrm{IH}_\infty) \vDash \forall \overrightarrow{x_n}\big(\overline{o}\,\eta\,a(\overrightarrow{x_n}) \leftrightarrow \overline{p}\,\eta\,a(\overrightarrow{x_n})\big)$, for every $a(\overrightarrow{x_n})$.

(I) and (II) follow directly from the fact that H_∞ respects I_∞ (lemma 15). From the coherence of I_∞ we know that it contains $\langle b(\overline{o}, \overline{q_0}, \dots, \overline{q_n}), b(\overline{p}, \overline{q_0}, \dots, \overline{q_n})\rangle$ for every term b with $n+1$ free variables and every sequence of objects $q_0, \dots, q_n$ with their canonical $\mathcal{L}^\wedge$-terms $\overline{q_0}, \dots, \overline{q_n}$. From the fact that H_∞ respects I_∞ it follows that

42 Proposition 11 strengthens an early result by Ross Brady, who shows, modulo notation, that the schema $\forall y(y\eta\hat{x}\phi \leftrightarrow y\eta\hat{x}\psi)$ over his theory defines a *congruent* relation of co-extensionality. Unlike the language $\mathcal{L}$ of the theory HC, however, Brady's language does not contain '=' Brady [1971].

43 See e.g. Ketland (2011, lemma 3.5).

III. $\mathfrak{N}(\mathrm{IH}_\infty) \models \forall\overrightarrow{x_n}\big(a(\overrightarrow{x_n})\,\eta\,b(\overline{o},\overrightarrow{x_n}) \leftrightarrow a(\overrightarrow{x_n})\,\eta\,b(\overline{p},\overrightarrow{x_n})\big)$, for every $a(\overrightarrow{x_n})$ and $b(\overrightarrow{x_{n+1}})$.

IV. $\mathfrak{N}(\mathrm{IH}_\infty) \models \forall\overrightarrow{x_n}\big(b(\overline{o},\overrightarrow{x_n})\,\eta\,a(\overrightarrow{x_n}) \leftrightarrow b(\overline{p},\overrightarrow{x_n})\,\eta\,a(\overrightarrow{x_n})\big)$, for every $a(\overrightarrow{x_n})$ and $b(\overrightarrow{x_{n+1}})$.

Since I_∞ is an equivalence relation, in particular transitive, we have that

V. $\mathfrak{N}(\mathrm{IH}_\infty) \models \forall\overrightarrow{x_n}\big(a(\overrightarrow{x_n}) = \overline{o} \leftrightarrow a(\overrightarrow{x_n}) = \overline{p}\big)$, for all terms $a(\overrightarrow{x_n})$.

Finally,

VI. $\mathfrak{N}(\mathrm{IH}_\infty) \models \forall\overrightarrow{x_n}\big(a(\overrightarrow{x_n}) = b(\overline{o},\overrightarrow{x_n}) \leftrightarrow a(\overrightarrow{x_n}) = b(\overline{p},\overrightarrow{x_n})\big)$, for all terms $a(\overrightarrow{x_n})$ and $b(\overrightarrow{x_{n+1}})$.

is true because firstly I_∞ is coherent, such that for every sequence of objects $q_0, \dots, q_n$ we have that $\langle b(\overline{o},\overline{q_0},\dots,\overline{q_n}), b(\overline{p},\overline{q_0},\dots,\overline{q_n})\rangle \in I_\infty$; secondly, the transitivity of I_∞ ensures that for every r, $\langle b(\overline{o},\overline{q_0},\dots,\overline{q_n}), r\rangle \in I_\infty$ iff $\langle b(\overline{p},\overline{q_0},\dots,\overline{q_n}), r\rangle \in I_\infty$, as desired.

Based on (I) to (VI), we show by an ordinary induction on the syntactic complexity of ϕ that

$$\mathfrak{N}(\mathrm{IH}_\infty) \models \forall\overrightarrow{x_n}\big(\phi(\overline{o},\overrightarrow{x_n}) \leftrightarrow \phi(\overline{p},\overrightarrow{x_n})\big)$$

thus completing the proof. □

Result 11 is highly desirable, and distinguishes HC from all other theories considered in this chapter (see §3.3 above).

So far, the theory HC of the fixed point model $\mathfrak{N}(\mathrm{IH}_\infty)$ has performed well. I now turn to the desideratum of comprehension. How much of the comprehension schema does HC contain? As it has been the case with the derivative theories of the previous sections, HC contains comprehension for a formula ϕ just in case it contains $Cl(\hat{x}\phi)$. Above, we have seen that every base language formula ϕ defines a class. However, the goal of a grounded theory of classes is to recover as much comprehension as possible for formulae that contain 'η'.

The derivative theory HKFB proved class-hood, and thus comprehension, for every *elementary* formula. Unfortunately, this positive result does not carry over to the present, direct approach. Over the class theory HC, elementarity no longer suffices for class-hood.

To see this, consider any formula ϕ elementary in the ψ_i, and assume that HC proves these ψ_i to define classes. In the old setting, this sufficed for ϕ, too, to define a class, even if ϕ contains an atomic formula of the form $\ulcorner\zeta\urcorner = a$, for some ζ not among the ψ_i. It only mattered which terms occur in the range of η. Formulae such as '$\ulcorner\zeta\urcorner = x_0$' did not incur presuppositions about ζ. However, the very point of the present model construction was a more sophisticated treatment of identity statements. As a consequence, however, elementarity as in definition 25 no longer suffices for a formula to define a class. In the following proposition, we also assume it not to contain class identity statements.

Proposition 12. *Let ϕ be elementary in the ψ_i, $i \leqslant n$, and assume that it does not contain any subformula of the form $a = b$ for a or b an abstraction term or variable. We have:*

$$\mathfrak{N}(IH_\infty) \models Cl(\hat{x}\psi_0) \wedge \ldots \wedge Cl(\hat{x}\psi_n) \rightarrow \forall y\big(y\eta\hat{z}\phi \leftrightarrow \phi(y)\big)$$

Proof. Recall definition 25 of elementarity (41). Let the ψ_i be arbitrary, and ϕ elementary in them. Further assume that '=' occurs in ϕ only flanked by terms of the base language. Assume that $\mathfrak{N}(IH_\infty) \models Cl(\psi_0) \wedge \ldots Cl(\psi_i)$.

Let o be an arbitrary object from $\omega \cup Abs$. I need to show that

$$\langle o, \hat{x}\phi \rangle \in H_\infty \Leftrightarrow \mathfrak{N}(IH_\infty) \models \phi(\overline{o})$$

Note that the left-to-right direction follows from the fact that the pair IH_∞ is an admissible extension of itself (lemma 15). So it suffices to show that if $\mathfrak{N}(IH_\infty) \models \phi(\overline{o})$ then $\langle o, \hat{x}\phi \rangle \in H_\infty$.

We reason by induction on the positive complexity of ϕ. So assume firstly that $\phi =$'$a = b$' and $\mathfrak{N}(IH_\infty) \models \phi(\overline{o})$. By our assumption about ϕ and without loss of generality, a is the variable x and b is a base language term denoting in $\mathfrak{N}$ a natural number n. Hence, if $\mathfrak{N}(IH_\infty) \models \phi(\overline{o})$ then $o = n$, and every admissible J contains $\langle o, n \rangle$. Consequently, $\langle o, \hat{x}\phi \rangle \in H_\infty$, as desired.

Secondly, let ϕ be an atomic formula with the relation symbol 'η'. Since it is assumed to be elementary in the ψ_i, we know that ϕ is of the form $x\,\eta\,\hat{y}\psi_i$. We assume $\mathfrak{N}(IH_\infty) \models \overline{o}\,\eta\,\hat{y}\psi_i$. Hence, $\langle o, \hat{y}\psi_i \rangle \in H_\infty$, such that for every admissible extension J, K of IH_∞, $\mathfrak{N}(J, K) \models x\,\eta\,\hat{y}\psi_i$, as desired.

Still at the base of our induction on positive complexity, we now turn to negations ϕ. By the elementarity of ϕ, however, this implies that it is either (i) of the form '$x \not\eta\, \hat{y}\psi_i$', for some $i \leqslant n$, or (ii) of the form '$x \neq a$' (without loss of generality).

If (ii) then $\mathfrak{N}(IH_\infty) \models \phi(\overline{o})$ only if $\overline{o}$ and a are both terms of the base language and we reason as just as with atomic equationhe before. So assume (i) that ϕ is of the form '$x \not\eta\, \hat{y}\psi_i$', and assume that $\mathfrak{N}(IH_\infty) \models \phi(\overline{o})$. Let (J, K) be any admissible extension of IH_∞. I need to show that $\mathfrak{N}(J, K) \models \overline{o} \not\eta\, \hat{y}\psi_i$. Since $\mathfrak{N}(IH_\infty) \models Cl(\hat{y}\psi_i)$, $\mathfrak{N}(IH_\infty) \models \forall x(x\,\eta\,\hat{y}\psi_i \vee x\,\eta\,\hat{y}\neg\psi_i)$. But we assume that $\mathfrak{N}(IH_\infty) \models \phi(\overline{o})$, i.e. $\mathfrak{N}(IH_\infty) \models \overline{o}\,\eta\,\hat{y}\neg\psi_i$, hence $\langle o, \hat{x}\neg\psi_i \rangle \in H_\infty$. Therefore, $\langle o, \hat{x}\neg\psi_i \rangle \in K$, too. Since it is consistent, however, , $\langle o, \hat{y}\neg\psi_i \rangle \notin K$, hence $\mathfrak{N}(J, K) \models x \not\eta\, \hat{y}\psi_i$, as desired.

Having thus completed the base case of our induction, we proceed to the case of disjunctions ϕ. Assume that $\mathfrak{N}(IH_\infty) \models \phi(\overline{o})$, i.e. $\mathfrak{N}(IH_\infty) \models \zeta(\overline{o})$ or $\mathfrak{N}(IH_\infty) \models \xi(\overline{o})$, for some ζ,ξ. Assume, without loss of generality, that $\mathfrak{N}(IH_\infty) \models \zeta(\overline{o})$. Note that ζ, too, is elementary in the ψ_i. Hence, by our induction hypothesis, $\langle o, \hat{x}\zeta \rangle \in H_\infty$. Therefore, for every admissible extension (J, K) of IH_∞, $\mathfrak{N}(J, K) \models \zeta(\overline{o})$. By logic, for every such (J,K), $\mathfrak{N}(J,K) \models \zeta(\overline{o}) \vee \xi(\overline{o})$. Hence, $\langle o, \hat{x}\phi\rangle \in H_\infty$, as desired.

Finally, assume that $\phi =$'$\exists y\big(\zeta(x,y)\big)$' and that there is some p such that $\mathfrak{N}(IH_\infty) \models \zeta(o,\overline{p})$. Then, by our induction hypothesis, $\langle o, \hat{x}\zeta(x,\overline{p})\rangle \in H_\infty$. By reasoning just analogous to before, we conclude that $\langle o, \hat{x}\phi\rangle \in H_\infty$. □

A formula $x = \hat{y}\psi$, however, is not ensured to define a a class whenever ψ does. In fact, the situation is even worse.

Proposition 13. *For every formula* $\phi(\overrightarrow{x_{n+1}})$

$$\mathfrak{N}(IH_\infty) \models \forall \overrightarrow{y_n}\Big(Cl(\hat{x}\phi, \overrightarrow{y_n}) \to \neg Cl\big(\hat{y}(y = \hat{x}\phi(x, \overrightarrow{y_n}))\big)\Big)$$

Thus, the natural way of defining the singleton of a given class fails. We would both like our class theory to recover a significant fragment of naive class comprehension and its classes to be extensional. The direct approach of the present section has solved the problem of extensionality, but its theory HC violates the desideratum of comprehension badly.

Proof.

Fact 2. *Let s be the $\mathcal{L}^\wedge$-term $\hat{x}(x\eta x)$. We have that neither $\langle s, \hat{x}(x\,\eta\,x)\rangle \in H_\infty$ nor $\langle s, \hat{x}(x\,\not\eta\,x)\rangle \in H_\infty$. Hence, $x\eta x$ does not define a class.*

Let ϕ be any formula and $q_0, \ldots, q_n$ any sequence of objects from the domain, with their canonical $\mathcal{L}^\wedge$-names $\overline{q_0}, \ldots, \overline{q_n}$ such that $\mathfrak{N}(IH_\infty) \models Cl\big(\hat{x}\phi(x, \overline{q_0}, \ldots, \overline{q_n})\big)$ and let ψ be the formula $x = x \wedge s\,\eta\,s$, for s as in fact 2. I show that $\langle \hat{x}\psi, \hat{y}\big(y = \hat{x}\phi(x, \overline{q_0}, \ldots, \overline{q_n})\big)\rangle \notin H_\infty$ and $\langle \hat{x}\psi, \hat{y}\big(y \neq \hat{x}\phi(x, \overline{q_0}, \ldots, \overline{q_n})\big)\rangle \notin H_\infty$, hence $\mathfrak{N}(IH_\infty) \models \neg Cl\big(\hat{y}(y = \hat{x}\phi)\big)$. I suppress the parameters $q_0, \ldots, q_n$ for the rest of the proof.

To show the first conjunct assume, for contradiction, that $\langle \hat{x}\psi, \hat{y}(y = \hat{x}\phi)\rangle \in H_\infty$. Then $\hat{x}\psi = \hat{x}\phi$ must be true in every admissible extension of $\mathfrak{N}(IH_\infty)$. In particular, the pair $\langle \hat{x}\psi, \hat{x}\phi\rangle$ must be in the fixed point identity relation I_∞ (cf lemma 15). By its fixed point character, we have

$$\forall J \forall K\Big(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J,K) \models \forall x\big(x = x \wedge s\eta s \leftrightarrow \phi(x)\big)\Big) \tag{26}$$

Now let o be any object in the domain of $\mathfrak{N}(IH_\infty)$; that is, let o be a number or an abstraction term. Since $\mathfrak{N}(IH_\infty) \models \forall y(y\eta\hat{x}\phi \vee y\eta\hat{x}\neg\phi)$, we can assume that either $\langle o, \hat{x}\phi\rangle \in H_\infty$ or $\langle o, \hat{x}\neg\phi\rangle \in H_\infty$. Firstly, assume that $\langle o, \hat{x}\phi\rangle \in H_\infty$. Hence $\forall J \forall K\big(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J,K) \models \phi(\overline{o})\big)$. Then by (26) and logic, $\forall J \forall K(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J,K) \models s\eta s)$ Hence, $\langle s, s\rangle$ must be in H_∞, contrary to fact 2.

Secondly, assume that $\langle o, \hat{x}\neg\phi\rangle \in H_\infty$ such that $\forall J \forall K(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J,K) \models \neg\phi(\overline{o}))$. Then by (26) and logic, $\forall J \forall K(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J,K) \models \neg\, s\,\eta\, s)$. This requires, again contrary to fact 2, that $\langle s, \hat{x}(x\,\not\eta\,x)\rangle \in H_\infty$.

To show the second conjunct assume, for contradiction, that $\langle \hat{x}\psi, \hat{y}(y \neq \hat{x}\phi)\rangle \in H_\infty$. By the fixed point character of the pair IH_∞, $\forall J \forall K(IH_\infty \Subset J, K \Rightarrow \langle \hat{x}\psi, \hat{x}\phi\rangle \notin J)$ This is the case only if either (i) $\langle \hat{x}\psi, \hat{x}\neg\phi\rangle \in I_\infty$ or (ii) $\langle \hat{x}\neg\psi, \hat{x}\phi\rangle \in I_\infty$.

Assume (i), such that

$$\forall J \forall K\Big(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J, K) \models \forall x\big(x = x \land s\eta s \leftrightarrow \neg\phi(x)\big)\Big) \quad (27)$$

Let o be any object of the domain. As before, since $\mathfrak{N}(IH_\infty) \models Cl(\hat{x}\phi)$, we have that either $\langle o, \hat{x}\phi\rangle \in H_\infty$ or $\langle o, \hat{x}\neg\phi\rangle \in H_\infty$. Firstly, assume the former. Then

$$\forall J \forall K(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J, K) \models \phi(\overline{o})) \quad (28)$$

Fact 2 ensures that there is a pair $(J_0, K_0) \sqsupseteq IH_\infty$ such that $\langle s, s\rangle \in K_0$. By (27), $\mathfrak{N}(J_0, K_0) \models \overline{o} = \overline{o} \land s\eta s \leftrightarrow \neg\phi(\overline{o})$ and by (28) and logic, $\mathfrak{N}(J_0, K_0) \models \neg\, s\eta s$ which contradicts our assumption that K_0 contains $\langle s, s\rangle$.

Secondly, assume that $\langle o, \hat{x}\neg\phi\rangle \in H_\infty$. Now choose $(J_1, K_1) \sqsupseteq IH_\infty$ such that K_1 does not contain $\langle s, s\rangle$ (IH_∞ itself is such a pair). By (27), $\mathfrak{N}(J_1, K_1) \models x = x \land s\eta s \leftrightarrow \neg\phi(\overline{o})$. By our assumption and logic $\mathfrak{N}(J_1, K_1) \models s\eta s$ contrary to our choice of K_1.

Now assume (ii), such that

$$\forall J \forall K\Big(IH_\infty \Subset J, K \Rightarrow \mathfrak{N}(J, K) \models \forall x\big(\neg(x = x \land s\eta s) \leftrightarrow \phi(x)\big)\Big) \quad (29)$$

We reason just conversely. For any o we firstly assume $\langle o, \hat{x}\phi\rangle \in H_\infty$ and choose a J_0 containing $\langle s, s\rangle$. (29) implies that $\mathfrak{N}(J_0, K_0) \models \neg\, s\eta s$, contradiction. Secondly, we assume $\langle o, \hat{x}\neg\phi\rangle \in H_\infty$, choose J_1 *not* containing $\langle s, s\rangle$, which contradicts our assumption and (29). □

3.5 CONCLUSION

In this chapter, I examined the prospects of class theory inspired by theories of grounded truth. I asked how to restrict the schema of class comprehension to grounded formulae, just as Kripke restricted Tarski's T-schema to grounded sentences.

Having laid out desiderata, I first explored the derivative approach. I translated "x is in the class of the ϕs" as "$\phi(x)$ is true" (p. 39). Through this translation, a theory of grounded truth induces a corresponding theory of grounded classes. The desiderata of section 3.1 suggested to start from the theory of truth KFB. The resulting class theory HKFB is closed under classical logic, allows for arithmetical as well as set-theoretical base theories, and proves comprehension for every *elementary* formula (p. 44). However, HKFB does not satisfy the desideratum of extensionality (§ 3.3).

In section 3.4 I turned to developing a theory of grounded classes directly. I described the extension of an arbitrary base model by a relation of grounded class membership and a relation of grounded class identity. The resulting model provides a theory HC whose classes are extensional in the strict sense that firstly, HC identifies $\hat{x}\phi$ and $\hat{x}\psi$ just in case $\forall y(y\,\eta\,\hat{x}\phi \leftrightarrow y\,\eta\,\hat{x}\psi)$. Secondly, classes that HC identifies are indiscernible in the theory. However, these positive results are blighted by a severe deficiency: according to the theory HC, whenever ϕ defines a class, $x = \hat{y}\phi$ does not.

Prima facie, we would like classes to be extensional, and their theory to provide natural ways of defining classes. My findings cast doubt on whether both can be achieved by the groundedness approach to class theory.

4

THE GROUNDEDNESS APPROACH TO CLASS THEORY: FINE'S CLASS MEMBERSHIP

The chapter is structured as follows. First, I explain Fine's ideas in informal terms (§4.1.1). Section 4.1.2 presents his model-theoretic construction as well as the class-theoretic axioms that it validates. Along the way, I will fill minor gaps in Fine's own presentation (§4.1.2.3).

Then, I discuss in more detail Fine's underlying account of well-founded definitions (§**??**). I argue that it needs be supplemented by an account of grounded membership.

Finally, I show that Fine's class-membership is indeed grounded in set-theoretic elementhood in the sense of chapter 1.

4.1 FINE'S THEORY OF CLASSES

4.1.1 *Philosophical Motivation*

Fine's idea is a shift in perspective: it is not the universe that is extended – all objects are given at the outset of the construction. Instead, the membership relation is developed in a step by step manner, whereby the members of more and more classes are 'revealed'.

To motivate this 'Copernican revolution', as he calls it, Fine invites the reader to think of classes as boxes [Fine, 2005, p. 548]. As long as the lid is closed, the members are hidden. But the containers may be opened, or made transparent, and their members be identified. In this picture, it seems natural that the classes are given, if only as black boxes, and the membership relation is developed later.

It is in how Fine 'opens the boxes' that he makes another characteristic turn. We do not take a box and open it – rather, we choose some things and then find the appropriate box, namely the container that they are in. Clearly, though, these things cannot be referred to as the members of this class. Instead, a predicate[1] is used to specify the collection of things that it is true of. For example, the formula 'x is an ordinal' opens the box that contains all the ordinals. It provides the class ON. In effect, the classes of Fine's theory coincide with predicate-extensions. The notion of class that drives his construction is what at the end of the paper he calls the 'logical conception' [Fine, 2005, p. 568]. That is, Fine's classes derive from concepts: the members of any class are just those which satisfy a certain condition.

In view of this, the question of paradox naturally becomes pressing. As the failure of naive set theory and Frege's *Grundgesetze* show, concepts must not carelessly be mapped into the first order domain. However, Fine's approach motivates an elegant solution.

Traditionally, the class-theoretic antinomies have been blamed on naive comprehension. Fine suggests a different analysis. Below, I will discuss his approach in more detail (§**??**). For the time being it suffices to point out that Fine's treatment of the paradoxes is based on the 'reversal in the roles of the predicate of membership and the ontology of

1 Fine himself speaks of 'conditions'.

sets' (op. cit.). If the membership relation is not any longer assumed to be unequivocal and given at the outset but on the contrary seen as the result of a step-by-step construction, the heedless use of *naive membership* may be seen as the culprit.

This new diagnosis also suggests a natural repair. If the membership relation is developed in stages then the phrase 'x is a member of y' expresses different concepts at different stages, as do complex formulae built up from it. Moreover, this re-interpretation proceeds in a way such that formulae which on their usual, *naive* interpretation lead to paradox now give rise to concepts that can coherently be taken to have definite extensions. And these are the classes of Fine's theory.

It needs to be added that Fine motivates his construction by a story about God and Archangel Gabriel. Both Gabriel and God know all proper classes but only God knows their members. In a heavenly dialogue God reveals to Gabriel which classes have which members. However, I think this epistemic vocabulary serves a merely heuristic purpose. Central to Fine's construction, to sum up the above, are his logical understanding of class, and the shift in focus from the classes to the membership relation.

4.1.2 *Technical Implementation*

Fine does not leave it at the philosophical motivation as described in the preceding section. He develops his theory of classes in more technical terms – he provides a model and axioms. I will try to set things out somewhat more explicitly. This will allow me to clear up certain ambiguities in Fine's presentation.

4.1.2.1 *The Ground Model: Set Theory with Urelemente*

Since this theory is meant to extend the set-theoretic universe **V** and imply, for instance, a class of all sets, Fine cannot literally define a model for it. What he can do instead, however, is to define a *set*-model. Its existence he can prove in ordinary ZFC extended by the assumption of an inaccessible cardinal. In addition, though, this construction is also a model in the scientist's sense. From truth in the set-model we can generalize to truth in the real world of classes. Thus, Fine's set-theoretic construction serves to motivate a theory that vastly exceeds standard set theory.

Fine starts from a set of urelemente C. On the intended interpretation, these urelemente are the proper classes. Fine aims for a class theory closed under complementation, hence for every set there needs to be a class of its non-elements. Thus, there must be at least as many urelemente as sets. Since Fine works in ZFC+'There is an inaccessible κ', the size of C is taken to be this κ.

The universe of classes now is modelled by the cumulative hierarchy on the basis of C. However, this hierarchy must not, as usual, be based on the power-set operation. If it was then already at the first stage there would be, contrary to Fine's intention, many more sets than classes. The 2^κ-many sets $\mathcal{P}(C)$ could never each have its own complement class in C. Fortunately, this difficulty is avoided as follows. At any stage $\alpha + 1$, instead of the full power set $\mathcal{P}(V_\alpha(C))$ we confine ourselves with the subsets of size less than κ ($\mathcal{P}^\kappa(V_\alpha(C))$). Then, $V_\kappa(C)$ has cardinality κ itself.[2]

$V_\kappa(C)$ now is the domain of Fine's models. It will remain unchanged all through the construction. On it, though, larger and larger membership relations e are defined. The starting point is ordinary set-theoretic elementhood e_0. It gives rise to a first model,

Definition 30. $\mathfrak{M}_0 = \langle V_\kappa(C), e_0 \rangle$

The range of e_0 contains only the pure sets of V_κ. At this first stage, the elements of C are not yet in the range of the membership relation.

4.1.2.2 *Mapping Urelemente to Formulae (1)*

However, many predicates of the language of set theory define proper classes, among which $\ulcorner x = x \urcorner$, or $\ulcorner \exists z(z \neq \emptyset \wedge \subseteq x \forall y \in z(y \cap z \neq \emptyset) \urcorner$ ('x is ill-founded'). The core of Fine's construction are what he calls *intensional* assignments: functions Δ that map the urelemente into these conditions. In fact, Fine allows for conditions formulated in an $\mathcal{L}_{\kappa,\kappa}$ extension of the first order language of set-theory that also contains a constant for every urelement [p. 551]. Interpreted in $\mathfrak{M}_0$, such a condition ϕ defines an extension $|\phi|_0 \subseteq V_\kappa(C)$. Thus, a new membership relation e_1 can be defined whose range now covers proper classes in C, too.

$$x\, e_1\, y \text{ iff } x\, e_0\, y \text{ or } x \in |\Delta(y)|_0$$

e_1 includes the set-theoretic elementhood relation (left disjunct). In addition, though, its range now also contains urelemente. On the intended interpretation, some black boxes have been opened and the members of some classes have been revealed. Formally, if a class c is mapped to a condition ϕ (i.e. $\Delta(c) = \phi$) and this predicate, according to $\mathfrak{M}_0$, is true of some objects, then $y\, e_1\, c$ if and only if ϕ, according to $\mathfrak{M}_0$, is true of y.

For example, some c will be mapped to the formula 'Sx' ($\Delta(c) =$'Sx'). If this formula is interpreted in the model $\mathfrak{M}_0$ then it defines a non-empty subset of the domain $V_\kappa(C)$, in fact quite a large one, namely $V_\kappa(C) \backslash C$. Therefore, c is in the range of e_1, and '$x \in c$' will be true in $\mathfrak{M}_1$ for every set x. Thus, c represents the class of all sets.

2 $|\mathcal{P}^{<\kappa} V_\alpha(C)| = \kappa^{<\kappa}$ which on the assumption of κ's inaccessibility is just κ.

Other 'boxes', however, remain opaque. There are classes whose extension cannot be expressed in terms of set-theoretic elementhood. In the model-theoretic construction, this is reflected by the fact that Δ maps some urelemente to formulae which do not 'deliver' if interpreted in $\mathfrak{M}_0$ – there are no objects that they are true of. One example is being in the complement of some particular set t.[3] The predicate

$$\exists y(x \in y \wedge \forall u(u \in y \leftrightarrow u \notin t)) \tag{30}$$

has an empty extension if interpreted in $\mathfrak{M}_0$. If Δ maps some urelement to the condition (30), that is, if there is to be a class of all set-complements, the range of the new membership relation e_1 cannot yet exhaust C.

The members of some more classes will only be revealed in the next step, when a new membership relation is defined in terms of the function $|\Delta(c)|_1$. If this procedure is iterated transfinitely many times, it gives rise to a sequence of models.

$$\mathfrak{M}_{\alpha+1} = \langle V_\kappa(C), e_{\alpha+1}\rangle \text{ where } x\, e_{\alpha+1}\, y \text{ iff } x\, e_\alpha\, y \text{ or } x \in |\Delta(y)|_\alpha$$
$$\mathfrak{M}_\gamma = \langle V_\kappa(C), e_\lambda\rangle \text{ with } e_\lambda = \bigcup_{\beta<\gamma} e_\beta, \text{ for limit ordinals } \gamma$$

4.1.2.3 *Indeterminate Membership*

Fine defines the *order* of a class as the stage where its members are revealed [p. 554].

Definition 31. $\text{order}(c) = \mathit{min}\{\alpha : c \in \text{rn}(e_\alpha)\}$

Since c enters the range of $e_{\alpha+1}$ just in case that $|\Delta(c)|_\alpha \neq \varnothing$, we can alternatively set

$$\text{order}(c) = \mathit{min}\{\alpha + 1 : |\Delta(c)|_\alpha \neq \varnothing\}$$

Thus, to use the picture again, the order of a class is the stage when the box has been opened and its content been determined. Clearly, this interpretation of $|\Delta(c)|_\alpha$ (for $\alpha = \text{order}(c)$) as the members of c makes sense only if once an urelement has been mapped to an extension of $V_\kappa(C)$, this extension does not change at higher stages.

Unfortunately, the construction as described so far does not provide the urelemente with unique extensions. There are formulae ϕ and stages α such that $\varnothing \neq |\phi|_\alpha \subsetneq |\phi|_{\alpha+1}$. Thus, on Fine's account, there will be a c such that at different stages, different members are ascribed to c.

3 That is, let t be a definite description in the language of set theory that picks out a unique set when interpreted in $\mathfrak{M}_0$.

An example is the formula 'x is membered'.

$$\exists u(u \in x) \tag{31}$$

At the outset of the construction, when '$\in$' is interpreted as the ordinary set-theoretic elementhood relation, 31 is true of all and only the pure sets, i.e. $|(31)|_0 = V_\kappa(C) \backslash C$. Already at the next stage, though, some urelemente have been added to the range of the membership relation such that (31) is now true not only of the sets but also of some of these. Formally, $|(31)|_1 = V_\kappa(C) \backslash (C \backslash rn(e_1)$. In general, for any stage α $|(31)|_\alpha = V_\kappa(C) \backslash (C \backslash rn(e_\alpha) \subsetneq |(31)|_{\alpha+1}$. For all these λ many different extensions, however, there is just one urelement c such that $\Delta(c) = (31)$; Fine explicitly wants Δ to be one-one [Fine, 2005, p. 553]. What, now, are the members of c? Fine's construction as he describes it does not determine the extension for all of its classes.

To show why this is a direct consequence of how Δ is defined, let me picture Fine's construction by a two-dimensional diagram (see figure 9). The vertical axis corresponds to the increasing membership relation and the horizontal lists the κ many formulae. The result is a two-dimensional table mapping formulae to their extensions for increasing interpretations of the relation symbol '$\in$'.

On Fine's account, Δ maps the urelemente one-one to formulae [Fine, 2005, p. 553]. In consequence, every urelement corresponds to a column of the table. Therefore, as soon as one formula is mapped to more than one non-empty extension, there are more non-empty fields in the table than classes. This picture shows why Fine's construction must *undergenerate*: it fails to provide enough classes for all the $\kappa \times \lambda$ many extensions.

Fortunately, this way of looking at the problem already suggests a solution. If you wish to retain Fine's basic idea of a step-by-step reinterpretation of the membership relation as well as continue interpreting the urelements as concept-extensions, then you must no longer map the urelemente to formulae but to pairs of one formula and one stage. In other words, an urelement no longer corresponds to a *column* of the table, but to one of its *cells*. In the next section I will suggest a way to spell out this intuitive idea.

4.1.2.4 *Mapping Urelemente to Formulae (2)*

First, using some ordinal enumeration of the urelemente, and encoding of pairs, define a bijection $\mu : C \mapsto \kappa \times \lambda$. Figuratively speaking, μ maps every urelement to a cell of figure 9, represented by a pair of two ordinals. On this basis, enumerating the formulae according to their lexicographical order, define

Definition 32. $\Delta'(c) = \phi_\alpha$ iff $\mu(c) = \langle \alpha, \beta \rangle$

e_λ	$V_\kappa(C)\backslash C$	$\overline{g} \cup \overline{h}$	$V_\kappa(C)\backslash(C\backslash\bigcup_{\alpha<\lambda} rn(e_\alpha))$
$\vdots$			
$e_{\omega+1}$	$V_\kappa(C)\backslash C$	$\overline{g} \cup \overline{h}$	$V_\kappa(C)\backslash(C\backslash rn(e_\omega))$
$\vdots$			
e_ω	$V_\kappa(C)\backslash C$	$\overline{g} \cup \overline{h}$	$V_\kappa(C)\backslash(C\backslash\bigcup_{\alpha<\omega} rn(e_\alpha))$
$\vdots$			
e_2	$V_\kappa(C)\backslash C$	$\varnothing$	$V_\kappa(C)\backslash(C\backslash rn(e_1))$
e_1	$V_\kappa(C)\backslash C$	$\varnothing$	$V_\kappa(C)\backslash(C\backslash rn(e_0))$
e_0	$V_\kappa(C)\backslash C$	$\varnothing$	$V_\kappa(C)\backslash C$

"1, 2""$_\lambda$10, 7"

$$\begin{tikzcd}
\ulcorner Sx \urcorner & (32) & \ulcorner \exists u(u \in x) \urcorner & \cdots \\
c \arrow[u, "\Delta"'] & d \arrow[u, "\Delta"'] & e \arrow[u, "\Delta"'] & \cdots
\end{tikzcd}$$

Figure 9: Fine's Δ: Formulae and membership relations

e_λ	"	"	m	"
$\vdots$				
$e_{\omega+1}$	"	l	"	"
$\vdots$				
e_ω	"	k	"	"
$\vdots$				
e_2	j	"	"	"
e_1	g	h	i	"
e_0	c	d	e	f

"1, 2""$_\lambda$10, 7"

$\ulcorner Sx \urcorner$	(32)	$\ulcorner \exists u(u \in x) \urcorner$	(**??**)

Figure 10: Δ': Formulae and membership relations

Importantly, and this is how it differs from Fine's original Δ, the intensional assignment Δ' is not bijective. Instead, for every formula ϕ there are $|\lambda|$ many urelemente c such that $\Delta'(\phi) = c$. In figure (10), every formula corresponds to a column of different urelemente.

At first, this modification seems already to have solved the problem of undergeneration. The increasing extensions of (31) now each make up a separate class. Generally, the bijectivity of μ ensures that

Fact. *For any stage α and any formula ϕ_β there is a $c \in C$ such that for any x*

$$x\, e_{\alpha+1}\, c \textit{ iff } x \in |\phi_\beta|_\alpha$$

However, Δ' gives rise to a new problem. For many formulae, their extensions remain constant from a certain stage on, for example the following ('x is in the complement of g or h').

$$\exists y(x \in y \land \forall u(u \in y \leftrightarrow u \notin g \lor u \notin h) \tag{32}$$

Assume that at the first stage (i.e. in the model $\mathfrak{M}_0$) none of the two classes c and d are revealed. This means, $|\Delta(c)|_0 = |\Delta(d)|_0 = \varnothing$. Therefore, both '$x \notin c$' and '$x \notin d$' are true of every object in the domain of $\mathfrak{M}_1$, i.e. $|x \notin c|_1 = |x \notin d|_1 = V_\kappa(C)$. But there is no object in the range of e_1 (which interprets '$\in$' in $\mathfrak{M}_1$) that is co-extensive with the $V_\kappa(C)$ (κ is inaccessible). Hence, there is no witness for the existential quantification in (32) – interpreted in $\mathfrak{M}_1$, (32) is *false* of every x. For this reason, $|(32)|_1 = \varnothing$.

But assume that at stage 1, at least c (but not d) is mapped to some set of objects such that $|\Delta(c)|_1 \neq \varnothing$. At the subsequent, second stage of the construction, '$x \notin c$' therefore is no longer vacuously true of everything. The formula '$\exists y(x \in y \land \forall z(z \in y \leftrightarrow z \notin c))$' therefore will have a non-empty interpretation in $\mathfrak{M}_2$. However, $|x \notin d|_2$ still is $V_\kappa(C)$ such that (32) is false in $\mathfrak{M}_2$, too. Therefore, (32) still has an empty interpretation in $\mathfrak{M}_2$: $|(32)|_2 = |(32)|_1 = \varnothing$. Only when both $\Delta(c)$ and $\Delta(d)$ have been mapped to non-empty interpretations, say at the third stage, '$x \notin c \lor x \notin d$' is false of some objects in the universe and (32) again true in $\mathfrak{M}_3$. In this case, however, the extension of (32) has been fixed also for any stage $\alpha > 3$.

This example shows that in the table of figure 10, there will be cells of the same content. Thus, since on the present approach every cell is considered to represent a class, the modified assignments Δ' map different urelemente to the same extension. Whereas Δ was not able to reflect differences, Δ' now *overgenerates*. However, this difficulty can be resolved if the definition of membership is carefully modified, as I will explain in the next section.

4.1.2.5 *Restricted Membership*

The modification I would like to propose can also be motivated from Fine's heavenly dialogue, or from a modest development of his story. When Gabriel has submitted a condition, and God opened a box that contains just those objects of which the predicate is true, She commands Gabriel to look back at all the boxes they have opened so far. She lets him check if any of these contains the same objects as the one just opened. If so, God closes it again. Only when Gabriel has done so, may he continue with the next condition. Thus, God ensures that at the end of their game, no two open boxes have the same content.

Let me now formulate this idea within the framework of Fine's set-theoretic models. More precisely, I will add to Fine's definition of $e_{\beta+1}$ a constraint that corresponds to Gabriel's checking all previously opened boxes.

First, notice that the function μ induces a natural ordering of the urelemente, when the pairs of ordinals are arranged in the reverse lexicographical order.

Definition 33. For $c, d \in C$, $c \ll d$ iff $\mu(c) = \langle \alpha, \beta \rangle$, $\mu(d) = \langle \gamma, \delta \rangle$ and $\beta < \delta$, or $\beta = \delta$ and $\alpha < \gamma$

By means of the relation '$\ll$' we can express that some boxes are opened earlier than others. Thus, it allows me to sharpen the idea of looking back at the boxes opened so far, and take a first shot at the condition I wish to add to Fine' construction. An urelement d that we have just for the first time mapped to a collection of objects is added to the range of the membership relation only if there is no $c \ll d$ of the same extension.

Due to the definition of '$\ll$', the condition 'there is no c: $c \ll d$' excludes those urelemente c that have been assigned the same extension at some earlier stage of the construction (formally, $\text{order}(c) < \text{order}(d)$). Nonetheless, these stages are again referred to when we compare extensions $|\Delta'(c)|_\alpha$ which are functions of $\Delta'(c)$ *and* some stage α. Therefore, to fully formalize the idea intended we also need to quantify over the stages α.

In sum, I propose the following definition of models $\mathfrak{M}'_\alpha$

Definition 34.

$$\begin{aligned}\mathfrak{M}'_{\alpha+1} =& \langle V_\kappa(C), e_{\alpha+1} \rangle \text{ where } x\, e_{\alpha+1}\, y \text{ iff } x\, e_\alpha\, y, \text{ or } x \in |\Delta'(y)|_\alpha \\ & \text{and for any } \gamma \leqslant \alpha \text{ there is no } c \in C \text{ such that } c \ll y \text{ and } |\Delta'(y)|_\alpha = |\Delta'(c)| \\ \mathfrak{M}'_\gamma =& \langle V_\kappa(C), e_\lambda \rangle \text{ with } e_\lambda = \bigcup_{\beta<\gamma} e_\beta, \text{ for limit ordinals } \gamma\end{aligned}$$

Henceforth, I will use 'membership sequence' strictly in the sense of this definition and will mean by 'membership relation' some e_α as it occurs in such a sequence of $\mathfrak{M}_\alpha$s.

This slight modification of Fine's construction solves the problems of the original proposal. On one hand, the use of Δ' ensures that each class is ascribed a definite membership (see proposition 4.1.2.4 above). On the other hand, the restriction now imposed on the definition of $e_{\beta+1}$ rules out that two different urelemente are assigned the same collection of objects. To consider the example from above, at the third stage, some urelement u such that $\Delta'(u) = 32$ is mapped to the union of the complements of c and d. From now on, any urelement will only be added to the range of the membership relation *only if* it is *not* assigned this extension $|(32)|_3$. In other words, u is guaranteed to remain the unique urelement that represents the class-union of the complement of c and the complement of d.

Another instructive example is found in the two formulae 'x is a set' (Sx) and 'x has a member' ($\exists y(y \in x)$). Interpreted at stage

0, these predicates have the same extension, namely the pure sets $V_\kappa(C)\backslash(C)$. However, there will be two different urelemente c and d such that $\Delta'(C) = \ulcorner Sx \urcorner$ and $\Delta'(d) = \ulcorner \exists y(y \in x)\urcorner$ but $\mu(c) = \langle \alpha, 0\rangle$ and $\mu(c) = \langle \beta, 0\rangle$. In other words, there will be two different urelemente corresponding to the two cells of the bottom row of figure 10 that contain the same extension $V_\kappa(C)\backslash(C)$. Fortunately, though, due to the lexicographical, i.e. strict linear ordering of the formulae we can assume, without loss of generality, that $c \ll d$. Therefore, d will not satisfy the condition imposed on membership in definition 34 ($|\Delta'(d)| = |\Delta(c)|$ *and* $c \ll d$), hence c witnesses the second conjunct).

This reasoning can be generalized to a proof that the construction does not *overgenerate*.

Proposition 14. *The classes of the models $\mathfrak{M}'_\alpha$ are extensional. For any* $c, d \in C$ *and any* α*, if* c, d *ascribed members* ($\exists x(x\, e_\alpha\, c \wedge x\, e_\alpha\, d)$) *then*

$$\forall x(x\, e_\alpha\, c \leftrightarrow x\, e_\alpha\, d) \rightarrow c = d$$

Proof. Argue by induction on α. If $\alpha = 0$ then the claim is vacuously true since no urelement is in the range of e_0. For α limit ordinal, $rn(e_\alpha) = \bigcup_{\gamma<\alpha} rn(e_\gamma)$ such that $(x\, e_\alpha\, c \leftrightarrow x\, e_\alpha\, d)$ only if $(x\, e_\gamma\, c \leftrightarrow x\, e_\gamma\, d)$ for some $\gamma < \alpha$, but then $c = d$ by the induction assumption.

Assume that $\alpha = \beta + 1$, $\exists x(x\, e_\alpha\, c)$ and $\forall x(x\, e_{\beta+1}\, c \leftrightarrow x\, e_{\beta+1}\, d)$. It cannot be that $x\, e_\beta\, c$ but not $x\, e_\beta\, d$. Namely, for x not to be in the range of e_β there would have to be a $\gamma < \beta$ such that $|\Delta'(c)|_\gamma = |\Delta'(d)|_\beta$ which contradicts the assumption that $d \in rn(e_\alpha)$ ($x\, e_\beta\, c \rightarrow x\, e_{\beta+1}\, c \leftrightarrow x\, e_{\beta+1}\, d$). Hence, there are two cases. Either (i), $x\, e_\beta c \leftrightarrow x\, e_\beta d$ and this implies, together with the induction assumption, $c = d$. Or (ii), $x \in |\Delta'(c)|_\beta \leftrightarrow x \in |\Delta'(d)|_\beta$ such that $|\Delta'(c)|_\beta = |\Delta'(d)|_\beta$. Assume that $c \neq d$ and without loss of generality $c \ll d$ – by the strengthened definition of $e_{\alpha+1}$ now x cannot be $e_{\beta+1}\, d$, contrary to the assumptions $x\, e_{\beta+1}\, c$ and $x\, e_{\beta+1}\, c \leftrightarrow x\, e_{\beta+1}\, d$. □

Moreover, the recursive definition of the $\mathfrak{M}_\alpha$ captures the intuitive idea of the construction as a *step-by-step* process in the course of which *more and more* classes are defined.

Proposition 15. *For any membership sequence, the range of the membership relation increases monotonically, in the sense that for every* $\alpha, \beta < \lambda$*,*

$$\textit{If } rn(e_\alpha) \subseteq rn(e_\beta) \textit{ then } rn(e_{\alpha+1}) \subseteq rn(e_{\beta+1})$$

Proof. The claim follows easily from modest observations. For one, the case of $\alpha \geqslant \beta$ is trivial (observe that the range of e increases). If $\alpha < \beta$, the following reasoning by nested induction suggests itself. First, notice that for any α and limit β, $rn(e_\beta) = \bigcup_{\alpha<\gamma<\beta} rn(e_\gamma) \supseteq rne_{\alpha+1}$. For $\alpha = 0$, assume that $x \in e_1$. If $\beta = 1$ then clearly, $x \in e_{\beta+1}$. If $\beta = \gamma + 1$ and $x \in rn(e_\gamma)$ then $x \in rne_{\beta+1}$, too. For successor α and $\beta = \alpha + 1$, $rn(e_{\alpha+1}) \subseteq rn(e_\beta) \subseteq rn(e_{\beta+1})$ follows straightforwardly. The case for $\beta = \gamma + 1$ is established just like before. □

This fact also ensures the existence of *terminal* membership relations e_λ that exhausts all of C. How large this terminal ordinal λ really is depends on how quickly the urelemente C are used up. This again is a matter of which predicates $\phi(x)$ the classes are mapped to, and therefore depends on Δ.

Fine, however, prefers to fix the terminal ordinal directly. For this, he introduces the notion of *class-inaccessibility*. λ is class-inaccessible if there is no ordinal $\alpha < \lambda$ such that for any membership sequence $\mathfrak{M}_\alpha$ defines a well-ordering of order-type λ (if there is such an ordinal α, λ is *accessible*).[4]

Fine proposes to focus on the *least* such class-inaccessible ordinal [Fine, 2005, p. 556]. Fine motivates this choice from the heavenly dialogue by which he had already illustrated the construction of the $\mathfrak{M}_\alpha$. God leaves it to Gabriel to decide how long their question-and-answer game continues.

Clearly, though, Gabriel cannot overview the construction as a whole. Nonetheless, there is a way for him to fix the length of the dialogue from 'within'. The membership relation e_α allows to formulate well-orderings on the universe. Each of these fixes an ordinal (their order type) and thus may be used by Gabriel to request a membership relation e_α.

If the construction proceeds up to the least class-inaccessible ordinal, therefore, it is continued as long as Gabriel may possibly wish. Membership sequences of this length thus reflect God's '... well-known love of freedom' [p. 556].

This is a nice picture. A more mundane reason to let λ be the least class-inaccessible ordinal is found in the following remark.

> Just as set-inaccessibility represents a natural closure condition for the formation of sets, so class-inaccessibility represents a natural closure condition for the definition of classes. [p. 557]

Class-inaccessibility, Fine suggests, transposes the usual set-theoretic notion into the class-theory of the models $\mathfrak{M}_\alpha$, and thus allows for the following analogy. Just as the least inaccessible cardinal is a natural upper bound to the cumulative hierarchy, the model of the least *class*-inaccessible cardinal completes the genesis of class-membership.

In Fine, the properties of the terminal models also depends on the cardinality or 'height' of the universe $V_\kappa(C)$. Namely, since the range of the membership relation keeps increasing, the size of $V_\kappa(C)$ constitutes an upper bound to the terminal ordinal λ.[5] However, since above the cumulative hierarchy has been constructed by means of the

4 In fact, Fine's notion of class-accessibility is somewhat weaker since it quantifies only over *regular* models in the sense of §**??** below.

5 That is, let t be a definite description in the language of set theory that picks out a unique set when interpreted in $\mathfrak{M}_0$.

restricted power-set operation $\mathcal{P}^{<\kappa}$ (see above), this complication may be neglected. In the present context, the size of the universe just is κ.

4.1.3 *Regularity*

Eventually, Fine focuses on a specific family of models.

> I wish to propose the *regular* terminal models M_λ, for lambda class- inaccessible, as the intended models for the theory of classes. [Fine, 2005, p. 557, my emphasis]

The relevant notion of regularity is defined in terms of the relation that one object x bears on another y if the canonical term of x occurs in the condition that Δ maps to y. This relation '$\overline{x}$ occurs in $\Delta(y)$' Fine labels the *dependence* of y on x [Fine, 2005, p. 554]. However, since I use term 'dependence' for the more general relation defined on p. 8 above, I will adopt an alternative terminology suggested by Fine, and speak of a class *being defined in terms of* an object. Note that this relation only makes sense to speak of relative to a given assignment Δ.

Definition 35 (Regularity)**.** Δ is regular iff for every $c \in C$, the relation '$\overline{\ldots}$ occurs in $\Delta(\ldots)$' is well-founded on the urelemente C. In other words, Δ is regular iff there is no infinitely descending chain of one class being defined in terms of another.

Although Fine does not make this connection explicit, his Regularity requirement follows from a general requirement on real definition that he develops in the final section of his paper.

(Requirement) Real definitions must be well-founded. More precisely, the relation '... is used to define ...' is well-founded on the objects defined.

When applied to Fine's class-theory, (Requirement) becomes his regularity requirement. First, recall that it is an function Δ that fix how the classes are defined. Accordingly, in this special case (Requirement) becomes a requirement on Δ. Further, Fine's class definitions are real definitions.

For some such Δ, now, the relation '... is used to define ...' is just the relation '$\overline{\ldots}$ occurs in $\Delta(\ldots)$'. Finally, the objects defined are the urelemente in C. In sum, Δ satisfies (Requirement) if the relation '$\overline{\ldots}$ occurs in $\Delta(\ldots)$' is well-founded on C; that is, if and only if it is *regular*.

4.2 THE GROUNDEDNESS OF FINE'S CLASSES

Fine draws an analogy between his regular models and the well-founded models of ZF [Fine, 2005, p. 554].

> It is the regular models, under our approach, which correspond to the well-founded models of ZF.

In section 1.4 above, I have shown that the well-foundedness of the sets is an example for the general concept of groundedness from chapter 1. Can the classes of Fine's regular models also be viewed as grounded? In this section I will answer this question in the affirmative.

I will identify a Finean class generator **F** such that for every object x from C, there is some α such that $x \in rn(e_\alpha)$ if and only if x is **F**-grounded in the standard sets.

However, just as Fine really presents a family of models, each induced by its own regular assignment Δ', it will only make sense to speak of a Finean class generator relative to a given Δ'.

Definition 36 (Fine's Class Generator)**.** Let Δ' be a regular assignment. x is **F**-generated from yy iff x is defined in terms of yy, relative to the assignment Δ'.

Proposition 16. *Let Δ' be any regular intensional assignment. Then for every object $x \in C$, x is **F**-grounded in the pure sets V iff for some α, $x \in rn(e_\alpha)$.*

Proof. □

5 WHAT IS THE PHILOSOPHICAL SIGNIFICANCE OF GROUNDEDNESS?

In chapter 1 I presented a general, formal concept of groundedness. Subsequently, I discussed applications of this general concept: the iterative conception of sets (§1.4), grounded truth (§2) as well as approaches to a grounded theory of classes 3. Now, I take a step back and ask for the philosophical significance of this formal concept of groundedness.

This question is not easy to answer. Many instances of the general concept lack philosophical content, and for others, it is at least controversial to claim that they have such. I will give examples in sections 5.1 to 5.3 below. Together, I take them to be evidence that the general, formal concept of groundedness from chapter 1 is in need of philosophical supplementation. One way of accounting for the philosophical significance of certain cases of groundedness, and explaining why others lack such, I will outline in the second half of this chapter, sections 6.1 to **??**.

5.1 FORSTER'S ITERATIVE CONCEPTION OF CHURCH-OSWALD CLASSES

I motivated my general concept of groundedness as a further generalization of Forster's generalized iterative conception Forster [2008]. However, his main example of a construction that falls under the iterative conception is a case of groundedness whose philosophical significance is controversial. It is the Church-Oswald construction of models for class theories with a universal class [Forster, 2008, §§2,5]. Their classes can be viewed as grounded in the sense of my formal definition, but it is not obvious whether this case of groundedness is philosophically significant. I briefly rehearse the simplest Church-Oswald construction in the usual, set-theoretic setting before I explain how it exemplifies groundedness.[1] Then, I will argue that the philosophical significance of this instance of groundedness is contentious.

Take any model $\langle M, E \rangle$ of Zermelo-Fraenkel set theory ZF and attach labels, say 0 and 1, to the objects of its domain. For example, this is implemented by taking pairs $\langle x, 0 \rangle$, $\langle x, 1 \rangle$ for $x \in M$. Choose a bijection c that maps every set in M to exactly one such pair, not necessarily containing this set itself. That is, $c(x)$ is a pair $\langle y, 0 \rangle$ or $\langle y, 1 \rangle$ for some $y \in M$. We assume that the rank of $c(x)$ is greater than that of x.

This function c allows us to define a relation F on M. Together with the domain we started from, M, this new relation gives rise to a model $\langle M, F \rangle$ of a theory of classes with a universal class.

1 My exposition follows closely Forster's [Forster, 2008, §5], but see also Oswald [1976].

Definition 37. For $x, y \in M$

$$xFy :\Leftrightarrow \exists z \begin{cases} c(y) = \langle z, 0 \rangle \text{ and } x \in z \\ \text{or} \\ c(y) = \langle z, 1 \rangle \text{ and } x \notin z \end{cases}$$

Let $\mathcal{L}$ be a basic language of class theory, the language of first-order logic extended by the relation symbol 'η' (see chapters 3 and 4). $\langle M, F \rangle$ is an $\mathcal{L}$-structure. F functions as a relation of class membership, and the M as classes. In particular, the object $u \in M$ such that $c(u) = \langle \emptyset, 1 \rangle$ functions as a universal class in the sense of this model: every $x \in M$ bears F to u. It can be shown that the model $\langle M, F \rangle$ validates extensionality with respect to the membership relation F.

In his 2008 article, Forster provides an alternative characterization of these classes (§2). It based on two "wands", or in the terminology of my chapter 1, two generators. The first one is well known. It is simply the set-generator **S** from section 1.4 (definition 9). The other is rather unusual: it allows us to generate from some things xx the class of everything that is *not among* xx.

Definition 38 (Forster's Complement Class Generator)**.** Let xx **Ƨ** y iff y has as its members all and only the z such that $z \not\prec xx$

The classes of the model $\langle M, F \rangle$ can be viewed as generated from their elements (in the sense of the relation F), by **S**, or from those objects that are *not* their elements, by **Ƨ**. To see this, recall that for every $y \in M$, there is a $z \in M$ such that either $c(y) = \langle z, 0 \rangle$ or $c(y) = \langle z, 1 \rangle$. In the first case, the model says that x is an element of y, for every x, if and only if $x \in z$. That is, the theory of $\langle M, F \rangle$ takes x to be an element of y if and only if x is among those things that we, in the meta-theory, know to be an element of z. In other words, y is the class of the things in z. z collects the objects zz from which y is generated by **S**. Thus, the sets of the model $\langle M, F \rangle$ represent objects of a new kind, classes that are generated through a generator quite unlike the standard generator of sets. I will use 'Church-Oswald class' to refer to these objects stipulated by Forster's new interpretation of the Church-Oswald models.

In the second case ($c(y) = \langle z, 1 \rangle$), the model says that x is an element of y if and only if $x \notin z$. In other words, the object theory takes x to be an element of y if and only if x is *not* among the things that the meta-theory knows to be in z. This time, therefore, z represents the objects zz from which y is generated by **Ƨ**. Note that unlike a standard set, a Church-Oswald class may not be grounded in its elements, but in those things which are precisely not its elements.

Forster argues that the class of Church-Oswald models, which are **ƧS**-grounded, are as legitimate as the standard sets, which are **S**-grounded. The fact that he considers it necessary to add philosophical argument to his groundedness characterization of the Church-Oswald classes, already suggests that the philosophical content of this characterization is not obvious. In the remainder of this section I will discuss whether Forster succeeds in establishing that **ƧS**-groundedness is as legitimate as **S**-groundedness. I will begin with a series of indirect arguments that Forster gives, as they will help clarifying what is at stake.

Forster discusses three worries one may have about his **ƧS**-groundedness characterization of the Church-Oswald classes [Forster, 2008, §4]. These worries are not of mathematical nature. It is not questioned that there are Church-Oswald models, nor that the classes of these models are **ƧS**-grounded. Instead, these worries are reasons to doubt Forster's contention that his two-wands picture is as philosophically significant as the standard, one-wand picture of the cumulative hierarchy [Forster, 2008, Horn 1 on p. 108]. Thus, they are worries about its philosophical significance.

Forster phrases these objections as arguments that the **ƧS**-grounded objects are not sets. I do not think that this is the most felicitous way of putting it. After all, it is trivial that among the **ƧS**-grounded things there are objects that are not sets (in the standard sense). As we have observed above, there is an object $u \in M$ of which the theory of our simple Church-Oswald model $\langle M, F \rangle$ thinks that it is a universal class, and there is no universal set. Unless, of course, by 'set' we no longer mean the **S**-grounded objects of the standard cumulative hierarchy, but allow for a more liberal use of this expression; in particular, unless we start calling the Church-Oswald classes 'sets'. However, this is not what is disagreed on. Forster does not engage in a merely verbal dispute. Therefore, I understand the objections that Forster considers as arguments that whereas **S**-groundedness provides a philosophical case for sets, **ƧS**-groundedness does not do the same thing for Church-Oswald classes. Does Forster succeed in fending off these objections?

The first argument goes as follows [Forster, 2008, §4.1]. **S**-grounded sets are constituted from their elements, but **ƧS**-grounded classes are not. Due to this difference between standard sets and Church-Oswald classes, the latter are not as legitimate as the former.

Forster responds to this argument in two steps. Firstly, he argues that to say that sets are constituted from their elements is just to say that sets are extensional. Secondly, he points out that the Church-Oswald classes are extensional, too. Therefore, **S**-grounded sets and **ƧS**-grounded classes do not differ after all in the relevant sense.

I do not think that Forster's response is conclusive. There is a relevant sense in which sets are constituted from their elements, a sense which is not exhausted by the extensionality of sets. In his seminal 1971 article, Boolos explicitly contrasts two characteristics of sets: on the one hand, their extensionality, on the other hand, the fact that

'[...] the elements of a set are "prior" to it' (p. 216). To say that a set is constituted from its elements may mean that it has both of the characteristics mentioned by Boolos, only that its elements are prior to it, or finally just that it is extensional. Forster's response addresses this latter sense in which a set is constituted from its elements, but not the others. The argument against the philosophical significance of **ƧS**-groundedness, however, can equally be formulated based on the other two senses. In particular, it is plausible to say that the standard set generator **S** tracks the priority of some things to their set, while Forster's complement generator **Ƨ** does not. Further, it can well be argued that the philosophical significance of **S**-groundedness stems from the fact that a set is **S**-grounded in presicely the things that are prior to it [Potter, 2004, §3.3]. In section 6.1 below I will pick up this line of thought and develop it further.

The second argument Forster considers is based on the following observation. Given some things *zz*, the condition of not being among *zz* only defines a plurality if the universe is already given as a definite collection. Otherwise, it is not definite which members a Church-Oswald class has. Note that I elaborate slightly on Forster's own exposition, in that I explicate his temporal metaphor of "the end of time" in terms of whether or not the universe is definite.

However, it is contentious to assume that the universe is a definite plurality.[2] In fact, it conflicts with our assumption about the set generator **S**. To see how, let uu be all the things there are and use **S** to generate from uu the set of all things {uu}, contradiction. Therefore, *prima facie* what members an **Ƨ**-generated class has is not a definite matter. In this sense, a Church-Oswald class are *intensions*, not *extensions*.

Standard, **S**-grounded sets are extensions. The elements of a given set are just those things from which it is generated, and therefore always ensured to be definite – otherwise, the set could simply not have been generated. Therefore, whereas **S**-groundedness ensures having a definite range of elements, **ƧS**-groundedness does not. Hence, the legitimacy and philosophical significance of **S**-grounded sets does not carry over to the **ƧS**-grounded Church-Oswald classes.

In Forster's discussion of this objection [§4.2] I discern two distinct, indeed possibly conflicting, responses. On the one hand, Forster accepts that the fact that unlike standard sets, a Church-Oswald class generated through **Ƨ** has definite members only if the universe is definite, marks '[...] an important difference' between **S**-groundedness and **ƧS**-groundedness [Forster, 2008, p. 105]. It is not a mathematical difference, since the relevant notion of definiteness (captured by Forster's metaphor of the "end of time") is of philosophical nature. Hence, Forster acknowledges at least one aspect in which his generalized iterative conception does not ensure philosophical significance.

2 See the extensive literature on *absolute generality*, e.g. in Rayo and Uzquiano [2006].

On the other hand, Forster argues that the objection overshoots. It does not only cast doubt on the legitimacy of **ƧS**-groundedness, but on the legitimacy of inductive, or in Forster's terms, recursive constructions. Thus, the objection contradicts what Forster labels *Conway's principle*, that 'objects may be created from earlier objects in any reasonably constructive fashion' [Forster, 2008, p. 99].[3]

Why should Forster's opponent be moved by Conway's principle? It depends on what we take it to mean. If the principle says that every inductive definition ensures philosophical significance, then for Forster to uphold it, is not to provide an argument for, but simply to repeat his conviction that his iterative conception of Church-Oswald classes is as legitimate as the standard iterative conception of sets.

A more charitable reading of Conway's principle is as giving expression to a feature of mathematical reasoning, namely that inductive definition of a collection of things licences reference to them. On this reading, Forster's response becomes that disallowing Church-Oswald classes contradicts mathematical practice. This would certainly be unacceptable.

However, the objection is not that **ƧS**-groundedness does not licence *mathematical* reasoning with Church-Oswald classes. It is already accounted for by the standard Church-Oswald model construction (definition 37). The objection is that Church-Oswald classes do not have the same philosophical significance as standard sets. More precisely, the objection is that Forster's **ƧS**-groundedness does not provide Church-Oswald classes with the legitimacy that standard sets acquire from the received iterative conception, **S**-groundedness in my framework. Therefore, the objection does not contradict Conway's principle as suggested by Forster.

In sum, Forster's reference to Conway's principle either merely restates his view that **ƧS**-groundedness is philosophically as good as standard **S**-groundedness, or it reminds us of the fact that in mathematical reasoning, inductive definition licences reference. The former is not an argument, while the latter does not conflict with denying its philosophical significance. Therefore, Forster's argument that the objection from definiteness overshoots, is not conclusive.

The third objection that Forster considers is a slippery slope argument. It goes as follows. If we accept that Forster's iterative conception of Church-Oswald classes, based on the set generator **S** as well as the complement generator **Ƨ**, is as significant as the received iterative conception of sets, based on **S** alone, then any other generator has equal claim to produce legitimate objects.

One philosopher's modus tollens is the other's modus ponens. Forster is ready to accept that for any generator Φ, Φ-groundedness is philosophically significant. More importantly, however, he points out that even if we did not accept every generator, the threat of regress

3 Forster cites Conway [2001].

> ... is not *by itself* an argument for drawing the line so close to home that the two-constructor case [i.e. **2S**-groundedness] is excluded.

I agree. By a similar thought, however, the fact that we cannot as suggested argue against **2S**-groundedness is no reason to agree with Forster. The burden of proof is on him to show that his conception of Church-Oswald classes is philosophically as good as the standard iterative conception of sets. After all, it is the received view that standard set theory Z, possibly ZF, receives good motivation from the iterative conception, and that in this respect it is superior to alternative theories, such as that of the Church-Oswald models. Forster claims that the Church-Oswald theory is just as well motivated. However, unless Forster provides positive reason for this, methodology requires us to adhere to the standard view.

So far, I have only presented Forster's indirect arguments, by which he responds to objections likely to be put forward against his unconventional view. In fact, however, Forster also provides a positive argument. Indeed, the first two sections of his paper are well viewed as arguing that his two-generator iterative conception of Church-Oswald classes provides them with as good philosophical motivation as does the one-generator iterative conception, i.e. **S**-groundedness, for standard set theory.

Forster's argument rests on the following assumption [Forster, 2008, p. 98].

(Q) The appeal of the cumulative hierarchy lies precisely in its neat response to Quine's challenge.

By 'the cumulative hierarchy' Forster refers to what I call **S**-groundedness. The 'appeal' that Forster ascribes to it is its appeal to philosophers, hence, at least partly, its philosophical significance.

By 'Quine's challenge', Forster means Quine's famous insight that a theorist can use first-order logic with identity to reason about things of a certain kind only if she has an identity criterion for them. The cumulative hierarchy, or rather the view that every set is found at some stage of it (i.e. is **S**-grounded), satisfies this necessary condition, that is responds to Quine's challenge, because it provides an identity criterion for sets. Finally, this response is 'neat' in the precise sense that the identity criterion provided comes with a '[...] recursive algorithm for deciding identity' [Forster, 2008, p. 98].

In sum, Forster's assumption (Q) is well paraphrased as follows.

(Q′) The philosophical significance of **S**-groundedness is that it gives a recursive identity criterion for sets.

Forster points out that **2S**-groundedness, too, provides a recursive identity criterion for Church-Oswald classes. Based on the assumption (Q′), he concludes that **2S**-groundedness is philosophically as significant as **S**-groundedness. I do not accept this conclusion, because I reject Forster's premise (Q′). Before I argue against (Q′), however, let me follow Forster and explain how **ƧS**-groundedness provides us with a recursive identity criterion for Church-Oswald classes.

For this, it is useful to remind ourselves how for standard, **S**-grounded sets x, y the question whether $x = y$ is answered. By their extensionality, we know that $x = y$ if the elements of x are the elements of y. Thus, the question whether $x = y$ reduces to questions whether $u = v$, for $u \in x$ and $v \in y$. This allows us to proceed as follows.

...

In sum, **ƧS**-groundedness provides a recursive identity criterion for Church-Oswald classes. This much is undeniable. Forster, however, deploys this fact to argue that **ƧS**-groundedness is philosophically as significant as the well-foundedness of standard sets, i.e. their **S**-groundedness. This inference relies on his assumption (Q′). In the remainder of this section, I will argue that this premise is false, and conclude that Forster's positive argument for the significance of **ƧS**-groundedness does not go through.

...

5.2 FRIEDMAN AND SHEARD'S MODELS OF TRUTH

In the previous section I have argued that the philosophical significance of Forster's iterative conception of Church-Oswald classes is controversial. The formal concept of groundedness from my chapter 1 generalizes Forster's liberalized iterative conception. In particular, the Church-Oswald classes are grounded, as I have explained in the previous section.

Thus, I have found reason to be sceptical about the philosophical significance of the general, formal concept of groundedness. In this section, I will give another case of groundedness whose philosophical significance is not obvious. In fact, I now turn to a case that, unlike Forster's iterative conception of Church-Oswald classes, was never intended as philosophically significant.

Recall, from chapter **??** Kripke's notion of *grounded truth*. It comes in several variants, each based on a distinct monotone evaluation scheme. The *Strong Kleene* variant (§2.4) has received the most attention, but truth-theoretic groundedness based on Weak Kleene, or on a *supervaluational* scheme have also been discussed.

I would like to emphasize that these cases of groundedness are philosophically significant. They have been discussed in philosophical journals and books, and not merely so in the wake of Kripke's seminal paper, but repeatedly over the past four decades. Today, Kripke's theory of truth, or family of theories to be precise, has become the standard theory of self-referential truth, to the extent that such consensus is found among philosophers. In particular, it is considered to have advantages over its revision-theoretic contenders. The appeal that Kripke's theory has to the majority of philosophers is at least partly due to that it is motivated from his notion of groundedness, which is an instance of the general formal concept from chapter 1. Therefore, truth-theoretic groundedness, in its Strong or Weak Kleene variant, or in one of its supervaluational variants, is philosophically significant.

However, there are further variants of truth-theoretic groundedness of which this cannot be said. They are found in another seminal piece of formal theory of truth, Harvey Friedman and Michael Sheard's [1987]. Friedman and Sheard provide an impressive array of results as to which axioms and rules, each of which embodies some aspect of naïve truth, are mutually consistent. For this purpose, they construct models very similar to Kripke's. However, these models themselves are not intended to capture an aspect of truth. They are merely technical devices to show that certain axioms are consistent. Nevertheless, their predicates of truth are grounded much like Kripke's (see p. 16).

For example, Friedman and Sheard construct a model $\mathfrak{N}(\ulcorner \mathrm{Th}_\infty \urcorner)$ whose truth predicate Th_∞ is the union of a sequence of sets Th_n, where Th_0 is true first-order arithmetic, and Th_{n+1} is the set of sentences ϕ such that [Friedman and Sheard, 1987, §3, G]

$$\{T\ulcorner \forall x\psi(x)\urcorner : \forall x T\ulcorner \psi(\dot{x})\urcorner \in \mathrm{Th}_n\} \cup \{T\ulcorner \psi \urcorner : \psi \in \mathrm{Th}_n\} \cup N \vDash_\omega \phi \quad (33)$$

Here, $\vDash_\omega$ is consequence in ω-logic, that is classical logic, in the language of arithmetic, enhanced by the following rule.

$$\frac{\phi(\bar{0}) \qquad \phi(\bar{1}) \qquad \dots}{\forall x \phi(x)}$$

This model validates the following axiom system whose consistency is thereby proved. [4]

$$\text{T-Intro}\ \frac{\phi}{T\ulcorner\phi\urcorner} \qquad \text{T-Elim}\ \frac{T\ulcorner\phi\urcorner}{\phi} \qquad \frac{\neg T\ulcorner\phi\urcorner}{\neg\phi}\ \neg\text{T-Elim}$$

$$\text{T-Rep}\ T\ulcorner T\ulcorner\phi\urcorner\urcorner \to T\ulcorner\phi\urcorner$$

$$\text{U-Inf}\ \forall x T\ulcorner\phi(\dot{x})\urcorner \to T\ulcorner\forall x\phi(x)\urcorner$$

The sentences in Th_∞, now, can be shown to be *grounded* in the truths of arithmetic N, in a manner very similar to how, say, Kripke's Strong Kleene theory of truth is grounded in them. More precisely,

4 The construction above is simpler than what Friedman and Sheard literally do. At every stage, they do not only add every ω-logic consequence of $N \cup \{T\ulcorner\psi\urcorner : \psi \in \mathrm{Th}_n\}$, but also every instance of the schemata T-Rep and U-Inf. However, for their purpose, i.e. to prove the consistency of the axiom system given above, the simpler construction discussed suffices.
Also, I suppress the fact that the base theory of Th_∞ is not just first-order arithmetic PA, but also includes basic truth-theoretic principles [Friedman and Sheard, 1987, p. 4]. Such details do not matter for the general point I intend to make, that the construction exemplifies the general concept of groundedness from chapter 1.

there is a generator Φ, not unlike the generators underlying Kripkean groundedness, such that $\phi \in \mathrm{Th}_\infty$ iff ϕ is Φ-grounded in $\mathbb{N}$.

On the *fine* understanding of semantic groundedness that I have found advantageous (see §2.3 above), the sentences in the least fixed point of, say, the Strong Kleene Kripke jump, are viewed as grounded in the the truths of arithmetic, through the combination of the truth generator **T** (p. 20) and the Strong Kleene logic generator **SK**. Similarly, the sentences of Friedman and Sheard's theory Th_∞ can be viewed as generated from the truths of arithmetic through three generators.

The definition of Th_{n+1} above (equation 33) has two key components. On the one hand, the relation $\vDash_\omega$ of consequence in ω-logic; on the other hand, the step from $\forall x T\ulcorner\psi(\dot{x})\urcorner$ to $T\ulcorner\forall x\psi(x)\urcorner$, and the step from ψ to $T\ulcorner\psi\urcorner$. Accordingly, we can view Th_∞, too, as grounded through the combination of a logic- and a truth-generator.

The truth generator, on the one hand, I will call it **GT**, is given by two rules. The first rule is T-Intro, in terms of which we have also characterized Kripke's truth generator **T** (p. 20). The second rule allows us to infer that it is true that for everything it is the case that ϕ, from the assumption that for everything it is true that ϕ.

$$\frac{\forall x T\ulcorner\phi(\dot{x})\urcorner}{T\ulcorner\forall x\phi(x)\urcorner}$$

Note that **GT**, just like Kripke's truth generator **T**, is deterministic in the sense of definition 1.

The logic generator, on the other hand, is simply the generator **C** of classical logic; recall that it allows us to generate universal quantifications from an infinity of sentences (p. **??**).

Proposition 17. *The sentences in Th_∞ are **GTC**-grounded in the truths of first-order arithmetic* $\mathbb{N}$.

$$\phi \in \mathrm{Th}_\infty \Rightarrow \phi \textit{ \textbf{GTC}-grounded in } \mathbb{N}$$

Proof. Since $\mathrm{Th}_\infty = \bigcup_{n<\omega} \mathrm{Th}_n$, it is natural to reason by induction on n. $\mathrm{Th}_0 = \mathbb{N}$, hence $\phi \in \mathrm{Th}_0$ is trivially **GTC**-grounded in $\mathbb{N}$.

For $\phi \in \mathrm{Th}_{n+1}$, we reason by cases.

$$\phi \in \mathrm{Th}_{n+1} \Leftrightarrow \left\{\begin{array}{ll} \phi \in \mathbb{N} \Leftrightarrow & \\ \phi = T\ulcorner\psi\urcorner, \psi \in \mathrm{Th}_n \Leftrightarrow \mathrm{Th}_n\ \mathbf{GT}\phi & \overset{\text{I.H.}}{\Leftrightarrow} \\ \phi = T\ulcorner\forall x\psi(x)\urcorner, \forall x T\ulcorner\psi(x)\urcorner \in \mathrm{Th}_n \Leftrightarrow \mathrm{Th}_n\ \mathbf{GT}\phi & \overset{\text{I.H.}}{\Leftrightarrow} \\ \Gamma \vDash_\omega \phi, \Gamma \subseteq \mathrm{Th}_n \Leftrightarrow \phi\ \mathbf{C}\text{-grounded in } \mathrm{Th}_n & \overset{\text{I.H.}}{\Leftrightarrow} \end{array}\right\} \phi\ \mathbf{GTC}\text{-grounded i}$$

□

So, the truth predicate of Friedman and Sheard's model $\mathfrak{N}(\ulcorner \mathrm{Th}_\infty \urcorner)$ satisfies the general concept of groundedness of chapter 1. Formally, Th_∞ is as much a predicate of *grounded* truth as is the least fixed point of Kripke's Strong Kleene jump (section 2.4). However, its groundedness is not philosophically significant. As mentioned before, Friedman and Sheard do not present their model construction as such. As to their paper, they are explicit that the approach is primarily logical, and that they do not intend to make a philosophical point [Friedman and Sheard, 1987, p. 2].

> We are not solving a problem in philosophy, but rather a problem in logic with a philosophical motivation.

As to the model constructions in section 3 of the paper, their sole purpose is to prove consistent certain collections of axioms and rules governing 'T'. No further role is mentioned nor any aspect of these constructions is discussed.

Moreover, even if we went beyond how Friedman and Sheard use their models, and sought to take them seriously as philosophers, this would still not render significant the groundedness of the sentences in Th_∞.

Firstly, when I presented the model $\mathfrak{N}(\ulcorner \mathrm{Th}_\infty \urcorner)$ above (33), I defined the set of sentences Th_∞ in a manner that renders it easy to see their groundedness, starting from N and step by step adding sentences with 'T'. Friedman and Sheard, however, define it explicitly as the least set containing those axioms and closed under those rules whose consistency they want to prove. Only in passing they remark that Th_∞ can also be defined as I did above. Therefore, even if Friedman and Sheard's construction of the model $\mathfrak{N}(\ulcorner \mathrm{Th}_\infty \urcorner)$ had philosophical significance, it would not obviously carry over to its groundedness.

Secondly, $\mathfrak{N}(\ulcorner \mathrm{Th}_\infty \urcorner)$ is merely one of a list of models each of which validates a specific axiomatic system. We have as little reason to believe in the philosophical significance of $\mathfrak{N}(\ulcorner \mathrm{Th}_\infty \urcorner)$ as in the relevance of any of the others. However, many of these other models do not exhibit groundedness. For example, Friedman and Sheard use, under the heading of "converging" truth, *revision-theoretic* means to construct a model that validates the inference from ϕ to $T\ulcorner\phi\urcorner$ and back [Friedman and Sheard, 1987, §3.D].[5] Its truth predicate Th'_∞ cannot be read as a predicate of grounded truth in the same way as I have found Th_∞ to be grounded. Therefore, even if we had reason to believe in the philosophical significance of Th_∞, it would not automatically be reason to take its *groundedness* to be significant.

I conclude that Friedman and Sheard's model $\mathfrak{N}(\ulcorner \mathrm{Th}_\infty \urcorner)$ is a case of groundedness that is not intended to be philosophically significant, that there is no reason to assume it is, and that the attempt of arguing

5 More precisely, Friedman and Sheard show the consistency of what has become known as the theory FS, see also [Halbach, 2011b, §14.3].

for its significance will face difficulties. Thus, I have given additional evidence that the general, formal concept of groundedness from chapter is in need of philosophical supplementation.

5.3 HOW TO GROUND ANYTHING

In the previous two sections I have presented cases of groundedness each of which resembles a paradigmatic, and philosophically significant, instance of the general concept (sets, truth) but whose philosophical significance is at least contentious. I now turn to present cases that satisfy the general theory, but do not even resemble anything philosophically significant. I show how to, speaking informally, *cook up* groundedness, and thus produce many cases of groundedness that clearly lack philosophical content.

Firstly, consider the following way in which the natural numbers are grounded. Take some numbers, say 4, 17 and 105, and compute their cross sum, $4 + 17 + 105 = 126$. Thus, we have given a way of generating a natural number from some others, and a way of viewing 126 as grounded in 4, 17 and 105. Of course, this case of groundedness is not interesting. This is not to say that cross sums are uninteresting. They may well be for pupils in primary school who have a particular leaning towards basic arithmetic. However, no point is made by calling 126 *grounded* in 4, 17 and 105.

Contrast the vacuity of cross sum groundedness with the case of the ordinals, that are grounded by Cantor's number generator (section 1.3). The generation of transfinite ordinals from the finite plays an important role in Cantor's case for the actual infinite, put forward in his 1883 *Grundlagen*. In particular, he writes that it his principles of generation contribute to providing the new numbers with 'the same [...] objective reality as the earlier ones' [Cantor, 1883, p. 911]. These principle I have captured in a generator **O** (definition 8 on p. 11). On this reading, Cantor thus he ascribes metaphysical significance to the generator **O**. The generation of cross-sums, in contrast, is not philosophically significant.

It may be thought that the cross sum generator is deficient because it is not deterministic (recall definition 1). Of course, 126 is the cross sum of many distinct collections of numbers. However, being deterministic is neither sufficient nor necessary for a generator to be philosophically significant. For one, the logic generators of Kripkean groundedness are not deterministic. For another, the truth generator **GT** of the previous section is deterministic, but arguably not philosophically relevant.

At any rate, it is easy to cook up deterministic generators. My second example of a clearly insignificant case of groundedness is one such. Consider arbitrary, countably many xx. Enumerate them: $x_0, x_1, \ldots$. Now every $x \propto xx$ is grounded in x_0 through the generator

$\mathbf{E}$ such that $y\mathbf{E}z$ iff there is an n such that $y = x_n$ and $z = x_{n+1}$. Thus, z is generated from y if y precedes z in the enumeration. Since it, however, is completely arbitrary, so is $\mathbf{E}$-groundedness of z in y. Note that $\mathbf{E}$ is deterministic: it is exactly x_n from which we generate x_{n+1}.

However, it is absurd to assume that this case of groundedness has philosophical significance. For one, we may begin to enumerate xx at any arbitrary y among them. That is, for every y of xx we may choose an enumeration E_y such that $x_0 = y$. Therefore, for every $y \propto xx$ there is a generator E_y such that every $x \propto xx$ is E_y-grounded in y. Every x of xx is somehow grounded in every y that is among them. Even if there were philosophical reasons to single out a specific $y \propto xx$ as the ground, these reasons could not lie in the general notion of groundedness but would have to be external to it.

For another, the observation may be strengthened. Any two objects whatsoever are some things xx, and indeed countably many. Therefore, for any two objects x and y whatsoever, there is a generator by which x is grounded in y (counting from y to x), as well as a generator to ground y in x (counting backwards).

Thus, I have given a recipe how to ground anything, in anything. This shows that the general formal concept of groundedness from chapter 1 is excessively weak: everything is grounded in some way. However, not everything is philosophically significant, fortunately so, as otherwise philosophical inquiry would be impossible. Hence, it cannot by itself be philosophically significant if some things satisfy the general concept.

Nonetheless, the cases of groundedness I discussed in the previous chapters, such as the iterative conception of set, or Kripke's theories of truth, have philosophical content. It is not accounted for by the general theory of chapter 1. Therefore, the theory needs to be supplemented by an account as to why certain cases of groundedness have philosophical content. In the remainder of this chapter, I will outline such an account.

For this, I return to the paradigmatic cases of groundedness presented in previous chapters. This time, I will look more closely at their philosophical content, in order to answer the question: what renders them philosophically significant?

6

PRIORITY AND THE ITERATIVE CONCEPTION OF SETS

For the time being, I focus on the iterative conception of sets (section 1.4 above). Not only is it a pleasingly simple instance of groundedness, it also is arguably the most extensively discussed among philosophers.[1] In particular, the philosophical content of the iterative conception has been debated Parsons [1977]; Potter [2004]; Incurvati [2012]. It is reasonable to hope that these discussions shed some light on how to account for the philosophical significance of the general concept.

6.1 THE PHILOSOPHICAL CONTENT OF THE ITERATIVE CONCEPTION OF SETS

I continue working within the formal framework of chapter 1. In particular, by the iterative conception of set I understand the view (see section 1.4) that sets are obtained by iterated application of the set generator $\mathbf{S}$, where $xx\mathbf{S}y$ iff xx are the elements of y. The pure sets are generated starting from nothing, such that the first stage is the empty set $\varnothing$. Sets are $\mathbf{S}$-grounded in nothing.

I have already touched on the philosophical content of this view. Above, I contrasted $\mathbf{S}$-groundedness with Forster's iterative conception of two "wands", $\mathbf{2S}$-groundedness. Among others, I observed the following difference (p. 80). Following Boolos, the standard set generator $\mathbf{S}$ may be motivated by saying that '[...] the elements of a set are "prior" to it' [1971, p. 216]. Using $\mathbf{S}$ we generate sets from precisely those things that are prior to it. It is this what makes $\mathbf{S}$-groundedness philosophically significant.

Admittedly, this thought is imprecise as it stands. Nonetheless, it is usually taken as starting point when philosophers ask for the content of the iterative conception of sets.[2] The challenge is to explicate the relevant notion of priority.

An early, influential discussion of different attempts at an explication is found in Parsons [1977]. First, he examines an intuitionist understanding [§2]. According to it, a set is literally constructed from its elements. Constructed by whom? Orthodox intuitionism would hold that a set is constructed by [Parsons, 1977, p. 339, my emphasis]

> [...] an idealized *finite* mind which is located at some point in time [...]

However, this approach is quickly seen not to succeed since it cannot account for infinite sets.

1 Of course, Kripke's theory of truth is also widely appreciated. However, the philosophical content of semantical groundedness is seldom discussed at another than the intuitive level already found in Kripke.

2 Thus, Parsons writes that '[...] one can state [...] what is essential to the 'iterative' conception: sets form a well-founded hierarchy in which the elements of a set precede the set itself' [1977, p. 336]. See also [Wang, 1977, p. 310], [Shoenfield, 1977, p. 321] and [Potter, 2004, p. 36].

Michael Potter, in a recent discussion [2004, §3.2] that summarizes nicely much of Parsons' 1977 contribution, suggests the following response on behalf of the intuitionist. Countably infinite sets may be viewed as constructed by an idealized finite subject, if we let her carry out supertasks, that is [Potter, 2004, p. 37]

> [...] tasks which can be performed an infinite number of times in a finite period by the device of speeding up progressively [...]

A set y whose elements are enumerated $x_0, x_1, \ldots$ thus can be viewed as constructed from them in a finite amount of time, by an idealized subject who has added x_0 after one second, x_1 after $1^1/2$ seconds, x_2 after $1^1/4$ seconds and so on through all the elements of y. After two seconds, the thought goes, she will have completed this supertask and constructed y.

However, it has been debated whether an intuitionist may allow for constructions carried out as supertasks Weyl [1949]. There is reason to believe that their possibility conflicts with rejecting the actual infinite. Moreover, as Potter remarks [ibid.], even if the concept of supertasks is available to account for the construction of countable infinite sets, it cannot help us to understand how *uncountable* sets are constructed from their elements. Therefore, the priority of a set to its elements cannot lie in it being constructed from them.

Having concluded that the priority of a set to its elements cannot be understood as its construction from them, Parsons develops an alternative account. He proposes to understand the priority in *modal* terms.[3] A set could not exist without its elements. For every set x, necessarily, there is x only if every element of x exists. However, the modal operator, '$\Box$' as a symbol, can be used in various distinct ways. Therefore, a modal account of how a set is constituted from its elements, is only useful if the modality at work is explicated. In his 1977 article, Parsons does not specify how he intends his claim to be understood, that a set could not have existed without its elements. Elsewhere, however, he does [Parsons, 1983, p. 316].

On the one hand, Parsons provides an argument that the modality of "a set could not exist without there being its elements" is not metaphysical modality. Metaphysically, all pure sets exist necessarily. In order to account for the priority of elements to their sets, however, it is essential that a set is *contingent* on its elements,

> [...] since when the elements are given the set is initially given only *in potentia*.

On the other hand, he outlines a positive account about how else to understand the modality, if not as metaphysical.

3 I will return to the connection between groundedness and modal logic in chapter 11.

> In saying that a multiplicity of objects can constitute a set, I mean that they can do so without changing anything at "lower" levels, that is, without changing the structure of the individuals or of the sets that might have entered into the constitution of the objects making up the multiplicity in question. It is this strong possibility that the modal operator [. . .] is meant to express.

It is helpful to draw an analogy with another modality that we know better.[4] While it is physically necessary that I do not leap skyscrapers, the laws of physics do not have to change for me to jump across a bench. It is in this sense that I can jump across this bench while I cannot leap that skyscraper. Analogously, some things xx do not have to change for their set y to be formed. In this sense, xx *can* constitute their set.

Parsons suggests one way of modelling this modality in terms of possible worlds. Just as we may analyze the physical necessity that ϕ by saying that it is the case that ϕ at every world where the laws of physics hold, we can paraphrase "necessarily, there is the set y" as "at every stage higher up in the cumulative hierarchy there is y". On this basis, Parsons glosses the necessity of ϕ as it being '[. . .] true "from there on" [. . .]' [Parsons, 1983, p. 317].

However, we cannot understand the modality of set constitution in terms of the cumulative hierarchy, if our goal is to explain why **S**-groundedness is philosophically significant. Since, the sets of the cumulative hierarchy just are all and only the **S**-grounded pure sets. Thus, explicating the modality as suggested by these remarks of Parsons would render our attempt at explanation circular. The philosophical content of **S**-groundedness is that a set could not exist without its elements, but all we have in order to understand this 'could', is **S**-groundedness itself.

Fortunately, Parsons has more to say about how the modality of set constitution, central to his modal account of **S**-groundedness, is to be understood. Later in his 1983 article [p. 328f.], he proposes to understand the priority of some things to their set as a modality distinct from, but related to metaphysical modality in that both specify, albeit in different ways, a general mathematical modality.

This notion of mathematical modality is not developed in detail, and may be found not sufficiently clear. For the purpose of explicating the modality of set constitution, however, it suffices to note that in this general sense of possibility, mathematical entities are fully contingent. They do not all necessitate one another, as they do in the case of metaphysical necessity. Thus, Parsons' notion of mathematical modality allows us to speak of it being possible that some, but not all, sets exist.

4 Note, however, that what follows is a charitable reconstruction of Parsons' remarks.

The modality of set constitution is then viewed as a specification of this general mathematical modality. From a world w with some sets xx, all and only those worlds are accessible at which each of xx has just the elements that it has at w. Consequently, the modality renders elementhood $\in$ *rigid* and vindicates Parsons' principles (R$\in$) and (R$\notin$) [Parsons, 1983, p. 209].

(R$\in$) $x \in y \rightarrow \Box x \in y$

(R$\notin$) $x \notin y \rightarrow \Box x \notin y$

In other words, v accesses w if and only if w end-extends v, with respect to the relation of set elementhood $\in$. That is, if x is an element of y at v then $x \in y$, too, at w.

We have thus been provided with an explication of the notion of modality in terms of which Parsons proposes to understand the priority of some things to their set. Moreover, this explanation does not refer directly to **S**-groundedness. Has Parsons thus succeeded and explained the philosophical content of the iterative conception of sets? I do not think so. Above, we have found that Parsons' first account of modality of set constitution made us attempt to explain the significance of **S**-groundedness in terms of **S**-groundedness. I think that the revised explication of the previous section also leads us into a circle, as follows.

Our starting point is that **S**-groundedness is philosophically significant because the generator **S** captures the priority of some things to their set (cf. p. 80). This priority we are now invited to understand modally: a set *could* not exist without its element. The relevant modality, however, is explicated in terms of one world accessing another just in case the latter end-extends the former.

This implies that a world w is accessible from v only if no set x that exists at w but not at v is **S**-prior to some set y at v.

...

6.2 THE ITERATIVE CONCEPTION OF SET

The core of the iterative conception is this: we can only combine some objects to their set if these objects are already available. In the previous section, I have discussed Parsons' approach of explicating this intuitive idea of availability in modal terms. Which notion of modality is the right one for this task? I have found that only a primitive notion of *sui generis* set-theoretic modality appears promising.

In view of this, however, we may as well return to our point of departure, and take the notion of priority, as in Boolos' early remark (p. **??**bove), as the *primitive* of conception of set.

...

Leaving aside metaphor: A set is *constituted* from its elements.

This idea motivates to think of the sets as coming in stages. At the base, there is the empty set, since it presupposes nothing. At the next stage, there is already the singleton of the empty set, and so in indefinitely.

Let's define the *rank* of a set inductively as one larger than the highest rank of any set that it presupposes.

Thus we can formulate a restricted principle of set comprehension to replace naive set comprehension.

The restriction to sets of lower rank ensures that constituency is well-founded on the sets. Assume otherwise. Then there is some infinitely descending chain of sets $(x_n)_{n\in\omega}$ such that x_i presupposes x_{i+1}, for every $i \in \omega$. But then for such set x_i, must be of a higher rank than x_{i+1}, which contradicts the well-foundedness of less-than on the ordinals.

In particular, we know that no set contains itself. Since assume that some set of rank alpha contains itself, then it must itself be of rank less than alpha, contradiction.

More generally, On the basis of the concept of constituency we have thus developed a response to the paradoxes of naive set comprehension.

6.2.1 *Constituency*

But what is this concept of constituency? What does it mean to say that a set is *constituted* from its elements?

Well, one response would be: We form the set from its elements. Or, the mathematics do. Or, some idealized subject does.

But of course, as platonists, we can't take this seriously.

How about a modal characterization? The set *could not* exist without its elements. Necessarily, if the set exists then so do its elements. But of course, such a modal understanding of constituency is vacuous, since sets exist of metaphysical necessity.

Such considerations lead the platonist to settle with a *primitive* concept of constituency. It is not defined in terms of other, more basic concepts. But we do understand it!

Consider these examples:

1. The Kingdom of Norway is constituted from the Norwegians.

2. The meaning of '+' is constituted from the usage of this symbol.

3. The quadrangle is constituted from two triangles.

These are substantial claims. We accept or reject them. So we understand them.

In addition, we can characterize the *formal* properties of constituency. For this purpose I choose a plural meta-language. In this framework we can formalize constituency as a relation that takes at its first place singular as well as plural terms. An object thus can be constituted from a single or from several objects.

We then fix the following principles. First, existence. If the xx constitute y then the xx and y exist.

Second, uniqueness. If the yy constitute x, and the zz constitute x, then the yy are the zz.

Finally, constituency is non-circular.

This is constituency, and it is on this concept that the iterative conception of set is based.

7 EXPOSITION GROUNDING

7.1 BOLZANO ON GROUNDING

7.1.1 *Introduction*

As to notation, I will use

- capital Roman letters from the beginning of the alphabet as variables that range over propositions,
- small letters from the beginning of the alphabet, mostly `b' andc'`, as variables over ideas,
- the expression 'A(c/b)' to denote the proposition that differs from A only in that where A involves b, A(c/b) involves c.

As usual, I will refer to Bolzano's *opus magnum*, the Wissenschaftslehre (1837), by 'WL'.

7.1.2 *Propositions, Ideas, Variation*

...

7.1.3 *Bolzano's Theory of Grounding (*Abfolge*)*

7.1.3.1 *An Obscure Notion?*

All concepts developed so far apply equally to true as to false propositions. Bolzano has more to offer: a special system for *truths*. True propositions are ordered by what Bolzano calls the relation of *Abfolge*. Let me translate it by 'grounding'.

Bolzano motivates his theory of grounding from examples of the following kind (WL §198).

(1) It is warmer in Palermo than in New York.

(2) The thermometer stands higher in Palermo than in New York.

Both propositions are true but (**??**) grounds (2) and not vice versa. Grounding stands out from Bolzano's system in that it is not defined in terms of variation. In particular, the fact that (1) grounds (2) and not vice versa cannot be captured by *derivability*: (1) can be derived from (2). Therefore, a stronger concept is needed: (1) *grounds* (2).

For a long time, interpreters have found this part of Bolzano's work 'obscure' (Berg 1962, 151). Nothing in a modern logic textbook corresponds to Bolzanian grounding. Nonetheless, the concept has a long and venerable tradition. Bolzano connects with Aristotle's distinction between *why*-proofs and mere *that*-proofs (Aristotle 2006, 1051b; Betti 2010). The fact that it is warmer in Palermo than in New York is *why* the thermometer stands higher in Palermo than in New York. Generally, the *grounds* of A is *why* A.

Moreover, very recently formal systems of grounding have been developed, prominently by Kit Fine, that is well viewed as resonating Bolzano's concept of *Abfolge*. I will survey this recent literature in the next section. Firstly, however, I present Bolzano's own theory of grounding.

Bolzano gives further examples (WL §§ 162.1, 201).

(3) The proposition that the angles of a triangle add up to 180 degrees grounds the proposition that the angles of a quadrangle add up to 360 degrees.

(4) The proposition that in an isosceles triangle opposite angles are identical grounds the proposition that in an equilateral triangle all angles are identical.

(5) The proposition that God is perfect grounds that the actual world is the best of all worlds.

If A grounds B then it is the case that B *because* it is the case that A. Bolzano's grounding is a concept of objective explanation. However, it must *not* be conflated with epistemic notions, such as justification. For one, just as derivability, grounding concerns how propositions, that do not have spatio-temporal location, are ordered independently of any subject. For another, justification suffers from the same shortcoming as derivability, in that it does not respect the asymmetry between the truths (1) and (2). If you know that the thermometer stands higher in Palermo than in New York, then you are justified in believing that it is warmer in Palermo than in New York.

Bolzano discusses whether grounding can be defined in terms of derivability, and possibly other notions (WL §200); his conclusion is that such a definition is not available. Therefore, Bolzano introduces grounding as a primitive concept and characterizes it by a system of principles, analogously to how he characterized his notion of proposition.

7.1.3.2 *Principles of Grounding*

Grounding is a relation between single or collections of propositions. I will use Greek capital letters ('$\Gamma, \Delta, \ldots$') as variables ranging over pluralities of propositions. Note that Bolzano assumes grounds to be always *finite* collections of propositions (WL §199). I use the symbol '$\lhd$' for grounding such that '$\Gamma \lhd A$' reads: the propositions Γ ground the proposition A.

FACTIVITY If $A_0, A_1, \ldots \lhd B_0, B_1, \ldots$, then $A_0, A_1, ..., B_0, B_1 \ldots$ (WL §203).

For example, (5) is a case of grounding only if the actual world really is the best of all worlds. Generally, only true propositions stand in the relation of grounding. Turning to our toy science of the Corlenos, we therefore know that

(6) Every sister is male.

does not ground

(7) Francesca is male.

The ground of A is *why* it is the case that A; it explains the fact that A. The sense of explanation at work here is objective and exhaustive. This allows us to draw two conclusions about the formal properties of grounding. Fristly, what grounds a proposition does not involve this proposition itself, neither directly or indirectly.

NON-CIRCULARITY There is no chain $\Gamma_0, ..., \Gamma_n$ such that for every $i < n$, Γ_i ground Γ_{i+1} and there's an A that is among the Γ_0 as well as among the Γ_n. (WL §§204, 218)

Secondly, the grounds of A are *unique*.

UNIQUENESS If Γ ground Δ and E ground Δ then Γ=E. (WL §206)

These principles describe the relation of grounding *formally*. For example, we know that *if* the truth that

(8) Vito is Michael's male parent.

grounds that

(9) Michael is son of Vito.

then it is not the case that (9) is grounded in

(10) Sonny, Frede and Michael are Vito's sons.

However, we would like to know more. Does (8) in fact ground (9)? More generally, what cases of grounding are there? Bolzano gives examples (such as (3) - (5)), but not many general principles. One such principle, however, is that every truth A grounds the proposition that it is true that A. Formally,

$$A \lhd \mathit{Tr}(A)$$

By the same principle we have that the proposition that A is true itself grounds the proposition that it is true that A is true. In symbols,

$$\mathit{Tr}(A) \lhd \mathit{Tr}(\mathit{Tr}(A))$$

Recall that truth is an idea. Since propositions are identified by how they are build up from which ideas, the proposition that A therefore is not identical with the proposition that A is true. Hence, *uniqueness* ensures it not to be the case that $A \lhd Tr(Tr(A))$. Generally, grounding is notion of *complete, immediate* objective explanation. It is natural, however, to consider its transitive closure, *mediate* grounding. Bolzano also considers the *partial* relation which can be defined from grounding (WL §198).

DEFINITION Γ *partially* ground Δ if there are some H such that the Γ are among the H, and H ground Δ ('Γ *left-partially* ground Δ'). Analogously we speak of Δ being a *partial* consequence of Γ ('Γ *right-partially* ground Δ).

The relation of *mediate left-partial* grounding Bolzano calls 'dependence'.

DEFINITION We say that Γ *depend* on Δ if the Δ are among some propositions that stand in the transitive closure of grounding to Γ.

If A depends on B, then Bolzano calls B an 'auxiliary truth' for A.

7.1.3.3 *Ascension Trees*

> If someone starting from a given truth M asks for its ground, and if finding this in [...] the truths $[A, B, C \ldots]$ he continues to ask for the [...] grounds, which [...] these have, and keeps doing so as long as grounds can be given: then I call this *ascension from consequence to grounds*. (WL §216)

Bolzano's idea is neatly captured as a *game*.

DEFINITION Let the ascension game $G(A)$ for a true proposition A be played as follows. Player 1 starts by playing the true proposition A. 2 responds by playing the propositions $B, C, \ldots$ such that $B, C, \ldots \lhd A$. In response, 1 chooses one of $B, C, \ldots$, and so on. A player wins if his opponent can't make a move. If a run continues indefinitely, 1 loses.

Note that an ascension game corresponds to a tree of true propositions T(A) whose top node is A and every other node of which represents an auxiliary truth (see figure ??). I say "represents" and not "is", since one and the same truth may be played several times during a run of $G(A)$. The nodes may well be thought of as *tokens* of A's auxiliary truths.

Bolzano realizes and makes use of this tree structure of the collection of auxiliary truths (WL §220).

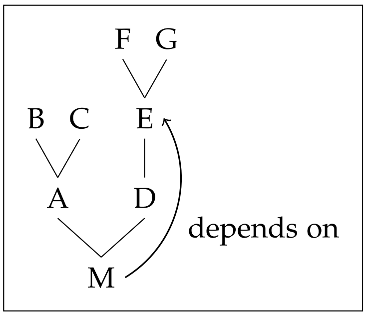

Figure 11: Ascension from *A*

7.1.3.4 *Basic Truths*

- *Dependence* does *not well-order* the truths: some true propositions have *infinite* ascension trees.
- For others again, ascension *bottoms out*:
- Bolzano argues that there are *basic truths*.

1. There are only finitely many simple *concepts*.
2. If A depends on B then A contains at least as many simple ideas than B.

Assume some conceptual truth A heads an infinite sequence of grounding. Then A depends on infinitely many truths. But from (1) we know that A involves only finitely many concepts, say n. Then by (2), every B that A depends on involves less than n+1 concepts. Since every finite set has only finitely many subsets, and any conceptual truth is uniquely determined by the set of concepts involved in it, we get that there are only finitely many conceptual truths that A depends on, contradiction.

- This works *only* for conceptual truths.

7.2 RECENT WORK ON METAPHYSICAL GROUNDING

The following is a focused survey on five recent papers, that all deal with groundedness in a metaphysical framework. On one hand, this means that they discuss how the relation of grounding may be used to answer metaphysical questions. A prime example is Fine's 2001 paper. On the other hand, they examine the notion by standards of contemporary metaphysics. For example, principles of grounding are proposed on the basis of a reflective equilibrium between intuitions and desired applications.

My interest in this literature is specific. I am interested in grounding because I want to defend certain foundational theories as grounded. Accordingly, I hope to find that what philosophers discuss under this label, proves to be a robust notion of sufficient precision that provides the philosophical justification I am looking for.

My goal is a unified and philosophically attractive response to both the class-theoretic and the semantic paradoxes. Therefore, I look for a notion of sufficient generality to apply to a range of different cases.

Finally, if it is to be used as a treatment of paradox, grounding itself better be a coherent notion. In this respect, another paper by Fine (2010) poses a challenge.

7.2.1 *Fine (2001) Question of Realism*

Fine offers grounding as a general tool to discuss realist positions. This offer is motivated from a discussion of how the dispute is usually phrased. Thus, Fine's notion of groundedness is embedded into a wider metaphysical project.

Fine points out that the anti-realist must account for the felicity of ordinary existence claims, since otherwise her position collapses into skepticism. Hence, the anti-realist needs to distinguish between two conceptions of reality. According to the *ordinary* conception, there are, say, prime numbers between 2 and 6. But, the anti-realist holds, this is not really the case. On the proper *metaphysical* conception, namely, there are no numbers.

This metaphysical reality has been understood in two different ways.

If realism about a proposition ϕ is understood in the *factual* sense, the realist holds that $|\phi|$ is true or false in virtue of how the world is like.[1] Conversely, anti-realism is the view that $|\phi|$ does not state any fact. As examples Fine lists expressivism in meta-ethics, formalism about mathematics and instrumentalism about science. In short, if

1 In the literature I am dealing with, propositions, facts and sentences are referred to in various ways. I will anticipate a notation introduced in Fine's 2011 (p. **??** below). - '$[\phi]$' denotes the fact that ϕ - '$|\phi|$' denotes the proposition that ϕ - '$\ulcorner\phi\urcorner$' denotes the sentence, well, $\ulcorner\phi\urcorner$

reality is understood the fundamental way, anti-realism about $|\phi|$ says that $|\phi|$ fails to 'perspicuously represent the facts' (p. 3).

The alternative understanding of metaphysical reality is to think of it as the basis to which can be reduced what is said to be real in a merely ordinary manner. Thus, anti-realism about $|\phi|$ is the view that $|\phi|$ is not fundamental but reduces to different propositions, whereas the realist holds that $|\phi|$ itself is irreducible. Prominent anti-realist positions in this sense are the view that that mathematical statements reduce to logic (*logicism*), and naturalism about ethics, according to which ethical truths reduce to facts about the physical domain.

Fine now argues in considerable detail that neither reading of anti-realism is intelligible. For the present purpose, I need not follow his discussion too closely. Eventually, Fine explains why any attempt to define reality in terms of factuality or reduction is bound to fail. Anti-realism, namely, is supposed to be compatible with ordinary discourse. More generally, non-skeptical anti-realism needs to be compatible with any non-metaphysical statements. Consequently, it must not be formulated in any non-metaphysical terms.

The problem remains: Instead of distinguishing between metaphysical and ordinary reality, now we need to separate metaphysical from merely ordinary *facts*. Fine suggests that this is just as hard.

This insight gives rise to *quietism* (p. 12): as the question of realism cannot be discussed but in purely metaphysical terminology, it is a pointless endeavour. Fine sets out to fend off this view. Equally well, he argues, we may explain the independence of realism from the substantial but *unique* nature of the issue.

Nonetheless, the quietist challenge persists as a methodological problem.

Even if realism can be discussed in terms of meaningful, although metaphysical, notions, it remains obscure how any dispute about these notions could be settled. Since the question of realism is supposed to be independent of ordinary statements, any notions involved in its discussion themselves show this independence. Consequently, whether they apply to a given case or not, cannot be settled on the basis of statements of ordinary discourse, on which realist and anti-realist could agree.

It is this methodological problem that Fine sets out to solve in the remainder of his paper. Accordingly, Fine's goal is to provide a general, purely metaphysical way of adjudicating between realist and anti-realist positions. It is here that the notion of ground is put to work.

Fine does not offer grounding as a criterion for factuality respectively reducibility. Instead, metaphysical reality is taken as a primitive concept. However, whether certain discourse reflects metaphysical reality (the questions of realism about this discourse) "turns on" questions in terms of grounding. This relation of one question turning on another is weaker than that of one question being explicated in terms of the other. It is a methodological relation.

Fine paraphrases the grounding relation as follows. The propositions $|\phi|$, $|\psi|$ collectively ground the proposition $|\chi|$ if its being the case that χ consists in nothing more than its being the case that ϕ and ψ. A *partly* grounds $|\chi|$ if $|\phi|$ is one of the propositions that collectively ground $|\psi|$.

Grounding is a relation between propositions. Fine distinguishes it from other such relations that are usually taken to bear on the question of realism.

First, Fine notes that the grounding relation is more liberal than that of reduction.

> A statement of reduction implies the unreality of what is reduced, but a statement of ground does not (p. 15).

Grounding statements do not have 'anti-realist import' (p.). Therefore, the realist and the anti-realist may agree on grounding statements. It is this feature that renders the notion a device to adjudicate between both positions.

Second, whether or not ϕ grounds ψ is independent of their logical relation – ϕ need not *analyze* ψ. Logical analysis is a linguistic matter, whereas grounding is of essentially metaphysical nature (p. 15).

Third, if ϕ grounds ψ, ϕ is a way of accounting for ψ. In fact, grounding is a special kind of explanation. If ϕ grounds ψ then ϕ is the *ultimate* explanation for ψ. Such remarks suggest a measure of comparing explanations. What does Fine have in mind?

Later, Fine puts it this way: ϕ explains ψ in the '... most metaphysically satisfying manner...' (p. 22).

Thus, Fine's notion of grounding seems stricter than that of Correia (see below) and other philosophers Rosen [2010], who use the term as a synonym for 'holds in virtue of'.

Fine also notes that the statement "The fact that ϕ grounds the fact that ψ" does not, contrary to first appearance, commit to facts or a substantial notion of truth. The grounding relation may equally well be expressed by the simple "ψ because ϕ". Unfortunately, Fine does not elaborate on this deflationary conception of grounding. It is not clear to me how thin the notion really can be in view of the heavy metaphysical work that Fine puts it to.

How can this notion of ground be used to adjudicate between the realist and the anti-realist? In section 6 of his paper, Fine shows that the realist about $|\phi|$ and the anti-realist disagree about the grounds of this proposition.

More relevant for the present purpose are Fine's methodological remarks from section 7, as to how disagreement about grounding relations is settled.

First, Fine attributes to philosophers reliable intuitions about matters of groundedness.

Further evidence for a statement of the form '$|\phi|$ grounds $|\psi|$' can be found in the candidate ground $|\phi|$ itself. This is because grounds are explanations, in fact explanations of superior character, and as such can be identified by their '... simplicity, breadth, coherence, or non-circularity'. Since such aspects are good evidence, $|\phi|$ is a good candidate for a ground to the extent that it is a good explanation. Therefore, grounding claims should be assessed according to standards of explanatoriness. Explanatoriness, however, should be assessed in context. Accordingly, questions of grounding cannot be properly answered in isolation but only in context.

Notice that in sum, Fine takes groundedness facts to have an a priori status.

7.2.2 *Batchelor 2010 'Grounds and consequences'*

For Batchelor (2010), grounding is relation between facts, that holds independent of epistemological considerations. Accordingly, for him the canonical statement of grounding is

> 'The fact that ϕ grounds the fact that ψ'

and 'ψ because ϕ' a mere paraphrasis.

On Batchelor's view, grounding takes a position between related, but merely empirical or merely logical relations.

On one hand, grounding is not restricted to the spatio-temporal. It makes perfect sense to talk about the grounds for, say, mathematical facts. This is how grounding differs from causation, and it is in this sense that Batchelor calls grounding a *logical* relation. An interesting question that is left open by these remarks is whether causation is a special, empirical case of grounding. Recall, at this point, that for Fine, causation may not be grounding (2001:15).

On the other hand, Batchelor distinguishes between grounding and mere implication. Grounding is the stricter relation. If the fact that ϕ grounds the fact that ψ then $[\phi]$ implies $[\psi]$. Q: But: although phi grounds phi&psi, phi doesn't imply phi & psi? Or does Batchelor mean *total* ground only?

However, grounding is not necessary for implication. For example, '$\phi \wedge \psi$' implies each conjunct, but the fact that ϕ *and* ψ does not ground ϕ nor ψ. What differs grounding from implication is its asymmetry.[2]

Batchelor notes that Bolzano (1837) was the first to examine the grounding relation, and anticipated many ideas of the contemporary discussion. Different from Bolzano, however, Batchelor sets out to define grounding. This definition is based on a classification of situa-

2 Batchelor speaks of anti-symmetry. However, he explicitly rejects the grounding relation to be reflexive (p. 70, second paragraph).

tions. A situation either obtains, in which case it is a fact, or it does not and is a mere *counter-fact*.

Batchelor sketches a hierarchy of factual and counter-factual situations. At its base, mereologically simple properties and individuals constitute *atomic* situations, some of which are atomic facts.

Complex situations are built up from what Batchelor calls 'factuality-functions'. First, there is negation, which maps facts to counter-facts and vice versa. The atomic facts and their negations Batchelor calls the 'elementary facts'. Second, conjunction maps the situation that ϕ and the situation that ψ to the situation that ϕ and ψ.

On this basis, now, Batchelor defines the relation of immediate grounding as follows.

- Elementary facts don't have any immediate grounds.
- A conjunctive fact is immediately grounded by any of its conjuncts.
- A fact of the form $[\neg\neg\phi]$ has only one immediate ground: $[\phi]$.
- The negation of a conjunctive fact is immediately grounded by any negation of a conjunct.

The grounds of a fact are its immediate grounds but also those facts that are linked to it through a chain of immediate grounds. In other words, the relation of grounding is defined as the ancestral of immediate grounding.

Notice that Batchelor's grounding is what Fine calls 'partial' ground; if $[\phi]$ grounds $[\psi]$ there may be $[\chi] \neq [\phi]$ that also grounds $[\psi]$.

This setting allows Batchelor to define a number of useful notions.

First, a fact is a *mediate* ground of another if there is a chain of immediate grounds between them of length greater than 1. Second, the *ultimate* ground of some fact is one that itself has no further ground. Ultimate grounds all are elementary facts. Finally, Batchelor identifies certain families of the grounds of a given fact. Some grounds of $[\phi]$ are *sufficient*, on one hand, if their conjunction implies $[\phi]$. The *complete* grounds of $[\phi]$, on the other hand, are simply all of them.

Having defined grounding and these auxiliary notions, Batchelor notes that

> (...) all these notions of grounding concern, as we may say, not the grounds of the being of facts, but rather the grounds of their factuality (p. 70).

Presumably, Batchelor refers to the circumstance that his definition implies any grounded situation to be a fact.[3]

3 Assume that $[\phi]$ immediately grounds $[\psi]$, and $[\psi]$ is a counter-fact. Then $[\psi] = [\neg\neg\phi]$, in which case $[\phi]$ is a counter-fact, too. Similar reasoning applies to the cases of $[\psi]$ having other forms. By induction we get that the elementary facts don't obtain, which is a contradiction.

What the definition presupposes, however, is the hierarchy of facts. It involves a different sense of groundedness, which Batchelor calls 'ontic': it is the sense in which the fact that $\neg\phi$ is grounded in its constituents, negation and the fact that ϕ. An ordinary, 'factive' ground is a constituent, for example $[\phi]$ is also an ontic ground of the conjunction $[\phi \wedge \psi]$. The converse does not hold, however, since a constituent may not even be a fact, for example the factuality-function $\wedge$.

(I skip Batchelor's remarks on pure necessity. For one, I don't find them overly lucid. Also, they do not seem to bear on his conception of grounding.)

In the third section of the paper, Batchelor elaborates on his earlier observation that grounding ensures implication, but not vice versa. If $[\phi]$ grounds $[\psi]$ then $[\phi]$ implies $[\psi]$, which validates the inference of $[\psi]$ from $[\phi]$. Thus, groundedness facts give rise to a system of proofs. These *canonical* proofs, as Batchelor calls them, reflect the grounding relations between facts.

7.2.3 *Hofweber 2009*

In his 'Ambitious, Yet Modest, Metaphysics', Thomas Hofweber raises worries about the notion of grounding.

Hofweber's motivating question is whether ontology can be modest, yet ambitious.

Hofweber argues against Fine's case for the intelligibility of the grounding notion (pp. 269n).

In the third section of his contribution, Hofweber

7.2.4 *Audi (forthcoming) 'A Clarification and Defense of the Notion of Grounding'*

In his unpublished 'Clarification and Defense of the Notion of Grounding', Paul Audi not only develops his own account of grounding but gives an original argument for the legitimacy, indeed indispensability, of the notion.

Audi, too, takes grounding to be irreflexive, asymmetric and transitive, although for different reasons than Fine and Batchelor.

Recall that Batchelor, in addition to the grounding relation on facts, also speaks of *ontic* groundedness of individuals and properties on their *constituents* (see above). Audi, now, draws a sharp distinction between grounding and constituency. This allows him to characterize grounding on the basis of property theory (see below). However, it also separates grounding from ontological dependence, and differs Audi's notion from how the term is used elsewhere in the literature Rosen [2010].

Audi ties grounding closely to how *properties* are related among each other. The fact that Fa grounds the fact that Gb only if the properties F and G are *essentially connected* (p. 10). Audi hastens to add that this terminology is not meant to be committed to a "thick" account of essence. All he requires is that

$\forall x(Fx \rightarrow Gx)$ (p. 4).

However, he assumes that which properties are essentially connected is a matter of necessity; it does not vary, so to speak, across worlds.

$\Box\forall x(Fx \rightarrow Gx)$

Audi derives that whether the condition on grounding holds, too, is a necessary matter.

Finally, Audi argues that grounding is not preserved by conjunction. If the fact that ϕ grounds the fact that ψ, the fact that ϕ and χ may still fail to be a ground of ψ (Audi calls this *non-monotonicity*).

In sum, Audi's understanding of grounding adds aspects to its conception in Fine or Batchelor. Audi's notion is therefore stricter than that of Fine or Batchelor. Also, he takes more seriously the challenge of justifying the notion.

His main argument is from the explanatory force of grounding statements. Fine, too, uses this idea to justify his use of grounding (see above). But Audi develops it into an explicit argument.

The fact that ϕ explains the fact that ψ can be used to explain the fact that ψ only if the first determines the latter. By 'determines' Audi simply means that the one fact 'makes it the case' that ψ. It is for this reason that *causes* are satisfactory explanations. However, Audi argues, many good explanations cannot be taken to state causation. He gives several examples, all of which I find convincing, and concludes that there is non-causal determination: grounding.

Later in his paper (§7) Audi turns to the interesting question as to how grounding differs from reduction. He considers this to be an important difference. In fact, Audi argues that reducibility does not even imply groundedness, since reduction implies identity, grounding, however, is irreflexive.

In a footnote (fn 56), Audi states that his partial grounding is a derivative notion, defined in terms of total grounding: the fact that ϕ *partially* grounds the fact that ψ if the fact that ϕ is one of the facts that ground the fact that ψ.

7.2.5 *Questions*

- Could Audi's notion play the methodological role Fine wants grounding to play?
- Audi understands grounding from the essential connection of the properties involved. Does this imply that if $[Fa]$ grounds

[Ga] then necessarily so? What more is required of essential connection than $\forall x(Fx \rightarrow Gx)$? More precisely, isn't this sufficient for grounding? In which case the grounding statement is equivalent to a necessity, hence necessary itself.

- Is there a weaker, yet relevant sense of reducibility on which groundedness does imply reducibility?

7.2.6 *Correia 2011 'Grounding and truth-functions'*

In this paper, Correia develops a formal theory of facts and grounding that answers to the metaphysical notion as discussed by the other authors, too.

He considers the definition of grounding in modal terms, but in view of criticism in the literature Audi [2010]; Correia [2005]; Rosen [2010] follows a different route. Correia develops an axiom system for a primitive grounding relation.

A basic assumption of Correia's is that grounding is expressed in any of the following forms:

1. The fact that ϕ is grounded in the fact that ψ, the fact that χ, ...
2. ϕ in virtue of the fact that ψ, the fact that χ ...
3. ϕ because ψ, χ, ...
4. The fact that ϕ is explained by the fact that ψ, the fact that χ,...

Correia distinguishes between two views on the logical form of grounding statements. On the *predicational* view, the basic form of grounding statements is (1). On the *operational* view, the logical form of grounding is captured by statements like (3), and grounding expressed by the sentential operator 'because'.

Recall at this point that Fine (2001) endorsed the operational view because it incurs less commitments (see p. 107 above). Correia takes the same stance.

Correia's grounding relation is many-one. This means, its left-hand side takes a plural term. As indicated by the canonical statements (1) to (4), a grounded fact has usually more than one ground. Moreover, Correia is clear that plural reference to facts cannot be reduced to singular reference to a conjunctive fact, since this would contradict the irreflexivity of the notion.[4]

---

4 To see why, consider the truth that
the fact that ϕ and ψ is grounded in the fact that ϕ and the fact that ψ. If the plural term on the right hand side was reducible to singular reference to a conjunctive fact, we would obtain 'The fact that ϕ and ψ is grounded in the fact that ϕ and psi$'. This contradicts the assumption that grounding is irreflexive.

Thus, Correia's grounding relation is *total*, different from Batchelor's *partial* grounding (p. 109 above), but in line with Audi's theory (p. 111).

Correia points out that the predicational view, according to which the logical form of grounding statements is (1), leads to the question for the truth-conditions of statements 'the fact that ϕ is the fact that ψ'

As this issue, however, is not directly relevant for my present concern I move on to Correia's formal theory of grounding. It is a classical first order system extended by propositional quantifiers. However, Correia does not allow for *plural* quantification over facts, despite his commitment to grounding as a many-one relation.

Further, Correia defines two relations:

1. The fact that ψ is the disjunction of some fact equivalent with ϕ ('$\phi \geqslant^{d} \psi$').

2. The fact that ψ is the conjunct of some fact equivalent to some disjunct of the fact that ϕ ('$\phi \geqslant^{cd} \psi$').

On this basis, Correia proposes axioms for the primitive operator $\mathcal{B}$ (read: 'because'). Interestingly, Correia prefers to have transitivity and asymmetry as theorems.

In §6, Correia provides bridge principles between grounding and propositional logic. He points out that certain intuitive principles (t1-3) require a conceptualist account of facts, according to which the fact that ϕ may not be identical with the fact that ψ. Other candidates, again, contradict the principle that

N* It's not the case that ϕ because ϕ and ψ.

(N*) follows from the transitivity and irreflexivity of grounding, on the assumption that ϕ grounds $\phi \wedge \psi$. Correia does not say so, presumably because he does not want, at this point, to invoke this assumption which relates grounding to conjunction. Instead, he prefers to derive (N*) from other bridge principles (see below).

Therefore, Correia eventually settles with the following bridge axioms.[5]

TF1 If $\phi \not\geqslant^{cd} (\psi \vee \phi)$ and ϕ then (ϕ or ψ) because ϕ.

TF2 If $\phi \not\geqslant^{cd} (\psi \wedge \phi)$ and $\psi \not\geqslant^{cd} (\phi \wedge \psi)$ and ϕ as well as ψ, then (ϕ and ψ) because ϕ and ψ.

In §6.3, then, Correia lists without further discussion the remaining bridge principles TF3 to TF6.

TF3 If ϕ because Δ, then (ϕ or ψ) because Δ.

TF4 If ϕ because Δ and (ψ or χ), and ψ, then ϕ because Δ, ψ.

TF5 If ϕ because Δ, and ψ because Γ, then (ϕ and ψ) because Δ, Γ.

TF6 If ϕ because Δ,(ψ and χ), then ϕ because Δ, ψ , χ.

5 In order to parse these axioms, recall that '$\phi \not\geqslant^{cd} \psi$' means that the fact that ϕ is no conjunct of any fact equivalent to some disjunct of the fact that ϕ.

One more axiom is added to Correia's system. Despite its suggestive title, this *reduction* axiom does not answer to the intriguing question to which extent groundedness implies reducibility. It states that grounding has a 'disjunctive nature', in the sense that if the fact that ϕ is grounded in the some facts Δ, ϕ is equivalent to a disjunction, one of whose disjuncts is the conjunction of all these facts.

In the final section of his paper, Correia proves his theory complete with respect to certain algebras. Since he himself rejects the metaphysical picture drawn by these structures, I do not think this result to be of great philosophical significance.

	Strict	Weak
Full	$<$	$\leqslant$
Partial	$\prec$	$\preceq$

Table 1: Fine's concepts of grounding.

7.3 fine's *pure logic of ground*

Fine sets up a calculus which comprises four concepts of grounding. This allows him to accommodate a range of views proposed in the metaphysical literature as well as bring out how these notions interact. One of them, *strict full* grounding, will emerge as the appropriate interpretation of semantic grounding.

First, Fine distinguishes between a weak and a strict sense of grounding. On the one hand, to say that $\phi, \psi, \ldots$ weakly ground χ is to say that for it to be the case that χ is for it to be the case that $\phi, \psi, \ldots$ [Fine, 2012a, p. 3]: In particular, any truth weakly grounds itself: weak grounding is reflexive.

Strict grounding, on the other hand, is irreflexive. Adopting a useful metaphor of Fine's, strict grounding moves us '... down in the explanatory hierarchy', where weak grounding has moved us merely 'sideways' [Fine, 2012a].

Second, Fine distinguishes between *full* and *partial* grounding. This distinction is made often [Audi, 2010, fn2010]. Fine draws it in terms of sufficiency: $\phi, \psi, \ldots$ fully ground χ just in case $\phi, \psi, \ldots$ are sufficient to ground χ. *Partial* grounds $\phi, \psi, \ldots$, on the other hand, merely help grounding χ: there are other $\xi, \ldots$ such that $\phi, \psi, \xi \ldots$ fully ground χ.

Fine presents his pure logic of ground as a system to derive sequents of the form "Δ ground A". Since I intend to apply Fine's system to enrich given theories, say of truth, by the resources to speak of grounding relations, I will formulate it as a theory in a language $\mathcal{L}_g$ which extends some first-order language $\mathcal{L}$. $\mathcal{L}_g$ is a *plural* language, with plural variables $xx, yy, \ldots$, a plural-term-forming operator ',' (such that x, y is a plural term) and the primitive relation symbol $\propto$, reading "is among" or "is one of".

Now add four connectives as in table 1. Thus, within $\mathcal{L}_g$ grounding is expressed in a way analogous to how in English, we express it by the connective "because". This approach allows us to make do without additional resources to speak of the fact that that ϕ, or the proposition that ϕ in a first-order setting.

Now let the *pure logic of ground* be formulated in the language $\mathcal{L}_g$ by the following rules.

Definition 39 (Pure Logic of Grounding)**.** Let $\mathcal{L}_g$ be a language of the pure logic of ground, i.e. extend some first-order language $\mathcal{L}$ by the grounding connectives. Say that a $\mathcal{L}_g$-sentence ζ is derived from a set of $\mathcal{L}_g$-sentences Γ in the *pure logic of ground*, if there is a proof of ζ from Γ by the following rules.

Subsumption

$$(\tfrac{<}{\leqslant})\ \frac{\phi_0, \phi_1, \ldots < \psi}{\phi_0, \phi_1, \ldots \leqslant \psi y} \qquad \frac{\phi \prec \psi}{\phi \preceq \phi}\ (\prec / \preceq)$$

$$(\tfrac{<}{\prec})\ \frac{\phi, \zeta_0, \zeta_1, \ldots < \psi}{\phi \prec \psi} \qquad \frac{\Gamma, \phi \leqslant \psi}{\phi \preceq \psi}\ (\tfrac{\leqslant}{\preceq})$$

$$\mathrm{Cut}(\tfrac{\leqslant}{\leqslant})\ \frac{xx_0 \leqslant y_0 \qquad xx_1 \leqslant y_1 \qquad\qquad y_0, y_1, \ldots \leqslant z}{xx_0, xx_1, \ldots \leqslant z}$$

Transitivity

$$(\preceq / \preceq)\ \frac{\phi \preceq \psi \qquad \psi \preceq \chi}{\phi \preceq \chi} \qquad \frac{\phi \preceq \psi \qquad \psi \leqslant \chi}{\phi \prec \chi}\ (\prec / \preceq) \qquad (\preceq / \prec)\ \frac{\phi \preceq \psi \qquad \psi \prec \chi}{\phi \prec \chi}$$

$$\text{Identity}\ \frac{}{x \leqslant y} \qquad \frac{x \prec x}{\perp}\ \text{Non-Circularity}$$

$$\text{Reverse Subsumption}\ \frac{x_0, x_1, \ldots \leqslant z \qquad x_1 \prec z \qquad x_2 \prec z \qquad \ldots}{x_0, x_1, \ldots < z}$$

Notice that Non-Circularity makes strict full grounding ('<') *non-monotone* in the following sense.

Lemma 16. *It is inconsistent with Fine's rules of grounding to assume that for every* $\phi, \psi_0, \ldots, \zeta_0$,,

$$\frac{\psi_0, \ldots < \phi}{\psi_0, \ldots, \zeta_0, \ldots < \phi}$$

Thus, if the X are full, strict grounds for ϕ then we cannot in general assume ϕ to be grounded in any extension of X. This should not surprise. Presumably, explanation generally is non-monotone. At any rate, non-monotonicity holds for the strong sense of explanation which grounding is thought to be. Assume that my being in pain is fully accounted for by the fact that my C-fibres are firing. Then it is not the case that my pain is equally fully explained by the fact that my C-fibres are firing and the fact that 1 plus 1 equals 2.

8 GROUNDING GROUNDEDNESS

In this chapter I use the notion of grounding to account for the philosophical significance of the Kripkean theories of truth from chapter 2. I will develop an interpretation of the groundedness of truth (chapter **??** that supplements the formal notion with philosophical significance.

Recall the *fine* concept of semantic groundedness from chapter **??**pp. 21ff.). I analyzed Kripke's jump in terms of two generators: a truth generator **T**, and a logic generator **M**. . . .

8.1 TRUTH

The philosophical significance of the truth generator **T** is that it expresses the view that if ϕ then it is true that ϕ because ϕ. In other words, it is true that ϕ in virtue of it being the case that ϕ. And the relevant notion of something holding in virtue of something else is precisely the concept of strict full immediate grounding $\lhd$ from the previous chapter. In symbols, I propose the following way of supplementing the formal notion of semantic groundedness with philosophical content.

$$\text{Read } \mathbf{T}\,\frac{\phi}{\mathsf{T}\ulcorner\phi\urcorner} \text{ as } \phi \lhd \mathsf{T}\ulcorner\phi\urcorner\text{, and}$$
$$\text{read } \mathbf{T}\,\frac{\neg\phi}{\mathsf{T}\dot{\neg}\ulcorner\phi\urcorner} \text{ as } \neg\phi \lhd \mathsf{T}\dot{\neg}\ulcorner\phi\urcorner.$$

Firstly, this understanding of Kripke's truth generator is natural. If we are asked, why is it true that ϕ? then to say, because ϕ, provides a full, and immediate answer.

Secondly, for ϕ an $\mathcal{L}_{\mathrm{ta}}$-sentence **TM**-grounded in true arithmetic, **M** the generator of some salient monotone logic, the relation between ϕ and $\mathsf{T}\ulcorner\phi\urcorner$ satisfies the formal principles of grounding, in particular Fine's *Pure Logic of Grounding*. In fact, this is just the special case of a general theorem about the formal properties of the relations of priority we obtain from the general theory of chapter 1.

Recall that for a generator Φ, we say that x is grounded in some gg ('$gg <_\Phi x$') if x have a Φ-priority tree whose leaves are gg. Further, recall that we say that y Φ-depends on x ('$x \prec_\Phi y$') if y has a Φ-priority tree one of whose leaves is x. Let us refer to these notions as *strict* priority and define derived notions of *weak* priority. For one, let us write $x \preceq_\Phi y$ if $x \prec_\Phi y$ or $x = y$. For another, let us write $xx \leqslant_\Phi y$ if $xx <_\Phi y$ or xx are exactly y. To see that this definition is reasonable, recall that I have stipulated the *root* of a tree not to be a *leaf*.

Theorem 1. *For every generator Φ, read $<$ as $<_\Phi$, $\prec$ as $\prec_\Phi$ etc. We have the following principles.*

Subsumption

$$(< / \leqslant)\ \frac{xx < y}{xx \leqslant y} \qquad \frac{\phi \prec \psi}{\phi \preceq \phi}\ (\prec / \preceq)$$

$$(< / \prec)\ \frac{xx, y < z}{\phi \prec \psi} \qquad \frac{\Gamma, \phi \leqslant \psi}{\phi \preceq \psi}\ (\leqslant / \preceq)$$

$$Cut(\leqslant / \leqslant)\ \frac{xx_0 \leqslant y_0 \qquad xx_1 \leqslant y_1 \qquad \qquad y_0, y_1, \ldots \leqslant z}{xx_0, xx_1, \ldots \leqslant z}$$

Transitivity

$$(\preceq / \preceq)\ \frac{\phi \preceq \psi \qquad \psi \preceq \chi}{\phi \preceq \chi} \qquad \frac{\phi \preceq \psi \qquad \psi \leqslant \chi}{\phi \prec \chi}\ (\prec / \preceq) \qquad (\preceq / \prec)\ \frac{\phi \preceq \psi \qquad \psi \prec \chi}{\phi \prec \chi}$$

$$Identity\ \frac{}{x \leqslant y} \qquad \frac{x \prec x}{\perp}\ Non\text{-}Circularity$$

$$Reverse\ Subsumption\ \frac{x_0, x_1, \ldots \leqslant z \qquad x_1 \prec z \qquad x_2 \prec z \qquad \ldots}{x_0, x_1, \ldots < z}$$

Proof. Firstly, the subsumption rules $(< / \leqslant)$ and $(\prec / \preceq)$ are immediate from the definition of $\leqslant_\Phi$ respectively $\preceq_\Phi$.[1] Similarly for the identity rule. Secondly, for Cut recall that $xx \leqslant_\Phi y$ if and only if y has a Φ-priority tree such that xx are the leafs, or its root y. Now, each of the premises of Cut is witnessed by some such tree. All we need to do to witness the conclusion is to attach the trees witnessing $xx_i \leqslant y_i$ to the node y_i of the tree witnessing $y_0, \ldots \leqslant z$. Doing so we construct a Φ-priority tree such that $xx_0, xx_1, \ldots$ are the leafs, or its root z – a tree that witnesses the conclusion $xx_0, xx_1, \ldots \leqslant z$.

By analogous, simpler constructions we show that the rules of transitivity, too, are validated by Φ. Thirdly, the non-circularity of $\prec_\Phi$ follows from the fact that that $\prec_\Phi$ is well-founded (p. 8). Finally, in order to show that the rule of reverse subsumption is validated, we need to show that z has a Φ-priority tree whose leaves are $x_0, x_1 \ldots$. We know for each i, z has a Φ priority tree one of whose leaves is x_i. But now, assume that z is the one and only of the $x_0, x_1, \ldots$ – then it is both the root of a tree and a leaf, hence not its root, contradiction. Hence, z is not the one and only of the $x_0, x_1, \ldots$ but has a Φ-priority tree whose leaves are exactly them. □

Thus, the priority trees of inductive definitions give rise to simple models of Fine's pure logic of grounding.

In particular, therefore, our reading of the truth generator **T** as immediate full grounding provides us a model for the other notions of grounding, too.

Recall that $\Sigma <_{\mathbf{tsk}} \phi$ iff ϕ has a **T-SK**,Σ-priority tree, and $\phi \prec_{\mathbf{tsk}} \psi$ iff ψ has a **T-SK**-priority tree one of whose leaves is ϕ. Let $I_{\mathbf{tsk}}$ be the set of **T-SK** grounded sentences.

1 Recall that xx, y are some things that y is one of.

8.2 LOGIC

In the previous section I have made a case for understanding the Kripkean truth generator **T** in terms of the notion of grounding from chapter 7. However, semantic groundedness is not matter of the truth generator **T** alone, but is the result of its interplay with how we derive complex sentences from literals. I now turn to this second component of my *fine* analysis of semantic groundedness, the logic generators **M** (§ 2.3). In this section, I argue that certain logic generators also are well understood as tracking connections of ground.

Here, the situation is more complicated for two reasons. On the one hand, I would like to provide a connection between the formal concept of semantics groundedness and metaphysical grounding that covers the various types of Kripkean fixed point constructions, say Weak Kleene logic as well as Cantini supervaluation (see §2 above). On the other hand, there is only little work on how metaphysical grounding interacts with logic, and even less of it arrives at definite verdicts.

However, it is worth observing that theorem 1 also holds for combined generators, in particular therefore for generators **T-M**.

Corollary 2. *For any monotone logic generator* $\boldsymbol{M}$*,* $\prec_{\boldsymbol{tm}}$*,* $\prec_{\boldsymbol{tm}}$*,* $\leqslant_{\boldsymbol{tm}}$ *and* $\preceq_{\boldsymbol{tm}}$ *satisfy the principles of the pure logic of ground.*

However, this fact does not suffice to render plausible a reading of the logic generators **M** in terms of grounding. The reason is that the pure logic of ground, as its name indicates, does not concern the interaction of grounding with logic. Given a monotone logic generator such as the Strong Kleene generator **SK** (§ 2.4) we may say that, for example, $\phi \prec_{\mathbf{sk}} \neg(\phi \vee \psi)$ (cf figure 8 , p. 25). Corollary 2 ensures that $\prec_{\mathbf{sk}}$ satisfies basic principles of grounding. However, it is silent as to whether this statements respects how grounding interacts with logic. Simply put, the pure logic of ground for $\prec_{sk}$ must *take* $\neg(\phi \vee \psi)$ *as an atomic fact*, and will have to treat each case of **SK**-generation as a brute fact. It is blind even towards the prima facie constraint that we would like $\phi \prec_{\mathbf{sk}} \neg(\phi \vee \psi)$ to be the case if and only if $\psi \prec_{\mathbf{sk}} \neg(\phi \vee \psi)$, too. What is needed are principles to adjudicate on how grounding interacts with logic.

As indicated above, research on grounding has not arrived yet at a consensus as to such an *impure* logic. As I have settled for Kit Fine's principles of a pure logic of ground, however, it is natural to equally connect with his proposal of an *impure* logic Fine [2012b]. Fine presents two distinct sets of axioms (§7).

> The first provides sufficient conditions for something to ground a logically complex truth of a specified form; the second provides necessary conditions.

My goal is to assess whether the logic generators of semantic groundedness are well understood in terms of metaphysical grounding. These logic generators, however, are given by introduction rules. Recall, however, in the process of generating grounded truths, the logic generator is used solely to close a given set of literals under logic. In other words, a logic generator is used to generate complex truths from simple ones. Therefore, it will suffice to compare the logic generators with those axioms of Fine's *impure logic of ground* that provide *sufficient* conditions.

Recall the notation of Fine's pure logic of ground (definition 39). In particular, recall that we write $\phi < \psi$ if the fact that ψ is fully, strictly grounded in the fact that ϕ.

Definition 40 (Impure Logic of Grounding – Introduction Rules, the Sentential Part)**.** Let $\mathcal{L}_g$ be a language of the pure logic of ground, i.e. extend some first-order language $\mathcal{L}$ by the grounding connectives. Say that a $\mathcal{L}_g$-sentence ζ is *upwards*-derived from a set of $\mathcal{L}_g$-sentences Γ in the *impure logic of ground* if there is a proof of ζ from Γ by the pure logic of ground (def. 39) and the following rules.

$$\wedge\ \frac{\phi \qquad \psi}{\phi, \psi < (\phi \wedge \psi)}$$

$$\vee - L\ \frac{\phi}{\phi < \phi \vee \psi} \qquad \neg\wedge\ \frac{\neg\phi \qquad \neg\psi}{\neg\phi, \neg\psi < \neg(\phi \wedge \psi)} \qquad \frac{\neg\phi \qquad \neg\psi}{\neg\phi, \neg\psi < \psi \vee \phi}\ \neg\vee$$

$$\vee - R\ \frac{\psi}{\psi < \phi \vee \psi} \qquad \frac{\phi}{\phi < \neg\neg\phi}\ \neg\neg$$

There is some complication pertaining to the rules for the universal quantifier. To accommodate it, Fine offers two distinct rules. I will return to it below. For the time being, however, I adopt the following rules which assume that we work in a non-free background logic and that the language $\mathcal{L}$ provides a name $\overline{o}$ for every object o of the domain. These assumptions are reasonable for my purpose, as I intend to apply the impure logic of ground to the language $\mathcal{L}_{ta}$ of semantic groundedness.

Definition 41 (Impure Logic of Grounding – Introduction Rules, the First-Order Part)**.**

$$\forall\ \frac{\phi(\overline{o}) \qquad \phi(\overline{p}) \qquad \ldots}{\phi(\overline{o}), \phi(\overline{p}), \ldots < \forall x\big(\psi(x)\big)}\ \text{for all } o \qquad \neg\exists\ \frac{\neg\phi(\overline{o}) \qquad \neg\phi(\overline{p}) \qquad \ldots}{\neg\phi(\overline{o}), \neg\phi(\overline{p}), \ldots < \neg\exists x\big(\psi(x)\big)}\ \text{for all } o$$

$$\exists\ \frac{\phi(\overline{o})}{\phi(\overline{o}) < \exists x\big(\psi(x)\big)}\ \text{for some } o \qquad \neg\forall\ \frac{\neg\psi(\overline{o})}{\neg\forall x\big(\psi(x)\big)}\ \text{for some } o$$

Now, it suffices to note that Fine's rules for the introduction of $<$ correspond precisely to the rules in terms of which I gave the Strong Kleene generator **SK** (p. 23). To see this, take any rule of Fine's system, and rewrite its conclusion of the form $\phi_0, \phi_1, \ldots < \psi$ like this:

$\frac{\phi_0 \quad \phi_1 \quad \ldots}{\psi}$. Doing so, you arrive at one of the **SK** rules from definition 12. For example, the rule $\neg\vee$ thus corresponds to the rule $SK_{\neg\vee}$.

$$\frac{\neg\phi \qquad \neg\psi}{\neg(\phi \vee \psi)}$$

Consequently, any reason to accept Fine's axioms as characterizing how metaphysical grounding commutes is a reason to believe that the logic generator **SK** respects this interplay of logic and grounding. More so, Fine's case for his impure logic amounts to a case that the generator **SK** is well viewed as tracking grounding in the precise sense that if is some sentence ϕ is generator through **C** from some sentences $\psi_0, \psi_1, \ldots$ then we may well view ϕ as *logically grounded* in them.

I conclude that Fine's work on the interplay of grounding and logic supports my proposed interpretation of Kripke's concept of semantic groundedness based on Strong Kleene logic. Not only does the truth generator **T** satisfy the formal principles of the *pure logic of ground*, we also have found that the Strong Kleene logic generator **SK** corresponds precisely to Fine's *impure logic of ground*.

...

8.3 DISCUSSION

It may come as a surprise that Strong Kleene logic gives rise to the rules of logical grounding, and not classical or, say, intuitionistic logic. In fact, this may be considered a *reductio* of my proposed connection between metaphysical grounding and semantic groundedness.

9 AN ITERATIVE CONCEPTION OF PROPER CLASSES

In this chapter I develop a conception of class that stands to Kripkean class theories of section 3.4 above, as the iterative conception of set stands to standard set theory.

9.1 TWO IDEAS OF COLLECTION

Recall the definitional idea of collection I outlined in section 3.1. Usually, people take Russell's paradox to show that the definitional idea of collection is flawed. I think this is too quick. For one, I'd ask for a fair comparison. Just as a naive notion of class there also is a naive notion of set. According to it, every plurality forms a set. In particular, therefore, the sets that don't contain themselves form a set. It contains itself just in case it doesn't. Contradiction.

Prima facie, therefore, the definitional idea is not worse off than the combinatorial. But, in the case of sets we have overcome our naïvety. We have replaced the naive notion by a mature conception of set. This is the iterative conception.

I propose to develop the definitional idea to a conception of class which saves the definitional idea from paradox, and is philosophically as substantial as the iterative conception of set. I propose an *iterative conception of class*.

9.2 ITERATIVE CONCEPTION OF CLASS

9.2.1 *Truths*

The definitional idea of collection motivates a change in perspective: we no longer attend to objects, and their combination to sets, but to facts.

Let me explain. First, of course, I need to make explicit what I mean by 'fact' in the present setting. A fact is a true proposition. Staying in line with Platonism I take propositions to be abstract entities. I assume them to be structured and individuated *finely*. To fix matters, I assume that every sentence of the language of set and class theory ϕ expresses exactly one proposition $[\phi]$ of set- or class theory, and every such proposition is expressed by exactly one such sentence.

9.2.2 *From (Definition) to Truths*

We wish to describe the world of classes. So we ask: "What classes are there"? According to the definitional idea, a class is given by its defining condition. This means that an object is a member of the class of the phis if the proposition that a satisfies under phi is true.

Of course, we need to be careful. Initially this thought led us to naive class comprehension, which is inconsistent. Comprehension must *not* hold for every proposition. We must somehow restrict the schema to the *safe* cases. This is my main goal, and this is where I think metaphysics can help us. For the time being I assume, that we do have a criterion of safety and can separate the safe from the paradoxical propositions.

And if we have this tool, and know that the propositions that a is a phi is true and is safe, then we may infer that a is contained in the class of the phis.

This class, however, is just another object, so first order logic allows us to infer that there is a class containing a. Our question was: "What classes are there"? Based on the safe truth that a is a phi we have now arrived at a partial answer. There is a class containing a. More generally, there is a class of all the objects a such that the proposition that a is a phi holds and is safe.

It is thus how the definitional idea of collection leads us from the usual, object-oriented viewpoint to a new, fact-oriented perspective. So, we now ask: "What propositions hold?"

~~~~

At this juncture, it is important to distinguish between two projects. On the one hand, our goal may be to develop a theory of class above all our science and mathematics. Let me call this the *comprehensive* project. If we engage in it, then we will start out from all propositions that hold according to science and mathematics. On the other hand, we confine ourselves with our two initial ideas, the combinatorial and the definitional view, and note that the combinatorial idea has been spelt out satisfactorily by standard set theory.

We may then focus on developing class theory atop of set theory alone. Our starting point now is the whole of set-theoretic truths. I will engage in this *focused* project.

Consequently, a first partial answer to the question "What propositions hold?" is: "All truths of set theory".

~

9.2.3 *Predicative Classes*

Well, as we've just seen we do have a good understanding of the world of sets. A first partial answer to the question "What propositions hold?" therefore is: "All truths of set theory".

For example, it is of course true that 4 is an ordinal. So we can speak of the class of the ordinals and say that 4 is contained in this class. More generally, for every set theoretic definition Phi we speak of the class of Phis. Thus, we get the predicative classes.

9.2.4 *Impredicative Classes*

But of course, we want more. Since, set theory already tells us all there is to say about predicative classes.
~~~~

Where set theory stops, and class theory begins, is with such classes that can*not* be set-theoretically defined. This question for impredicative classes, however, leads us to the question for true *class-theoretic* propositions.

And here, we're facing a problem. A class-theoretic proposition involves statements of the form: 'a is in the class of the Phis'. But Phi may itself be a class-theoretic condition. Thus, a class-theoretic truth may presuppose the truth or falsity of some class-theoretic proposition. How do we make sure that these presuppositions bottom out?

I will offer a solution to this problem. But first, we have to understand better the problem. What does it mean for a truth to presuppose some other truths?

To determine whether such a proposition is true we thus already have to be able to evaluate a class-theoretic proposition.

9.2.5 *Grounding 1*

Grounding to the rescue. The concept of grounding is an ideal candidate to address the problem of mutual presuppositions in a systematic manner.

Let's briefly take stock.

I started out from two distinct ideas of collection: Combination and Definition. If taken at face value, both ideas lead to paradox. But, if we pair the combinatorial idea with *ontological constituency*, then we arrive at a consistent and philosophically substantial conception of set. I want an analogous solution for classes.

It's not about objects, but about propositions. So ontological consistency will not provide a conception of class. We need some analogous notion for propositions. And this is just what grounding does for us.

So, my slogan is:

> Grounding is constituency for the definitional idea.

9.2.6 *Grounding 2*

Grounding has the right formal properties.

First, only true propositions stand in the relation of grounding. Second, grounding is unique, on its left as well as on its right side. Third, if you trace the grounds of a truth ϕ, you will never be led back to ϕ itself.

The absence of grounding circles is well formulated graph-theoretically. For this we picture the propositions as points in the plain, and connect two points just in case one point represents a proposition that is partially but immediately grounded in the proposition represented by the other. I call this a grounding graph.

NON-CIRCULARITY A grounding graph does not have any cycles.

In sum, grounding has the same formal properties as the set theorist's concept of ontological constituency.

9.2.7 *Proper Classes and Grounding*

Of course, formal principles by themselves don't tell us what a given class-theoretic proposition is grounded in. But, there is an easy answer.

An object a is contained in the class of the Phis *because* a is a Phi. This common place can be specified: the truth that a is contained in the class of Phis is *grounded in* the truth that a is a Phi.

It is this principle which allows for a platonist explication of the intuitive definitional idea. On this basis I will develop a hierarchy of class-theoretic truths.

And we can say more: Classes are concept-*extensions*. They are extensional objects. Why is the class of the phis identical to the class of the psis? *Because* everything is a phi just in case it is a psi. Again, invoking the notion of grounding this common place can be specified: The truth that the class of the phis identical to the class of the psis is *grounded in* the truth that everything is a phi just in case it is a psi.

9.2.8 *Impure Logic*

What about the grounds of a logically complex proposition? For example, does the truth that ϕ ground the truth that ϕ or ψ, for any ψ? This topic is discussed lively in present-day metaphysics, and has proved hard to come by.

Therefore, I prefer to remain neutral on this question. And I think I can be neutral. Since, my goal is to convey the idea that grounding induces a hierarchy on the class-theoretic truths. I don't need to determine in every detail how this grounding works.

So I assume that we have chosen some set of rules concerning grounding and the connectives and quantifiers, that is, an *impure* logic of grounding.

In all likelihood this will contain

$$\frac{\Delta \lhd \phi}{\Delta, \phi \lhd [\phi \vee \psi]}$$

and may also contain

$$\frac{[\Delta] \lhd [\phi]}{[\Delta], [\phi] \lhd [\neg\neg]}$$

But as I said, such principles are controversial. All I need is to make two very weak assumptions.

Secondly,

9.2.9 *Putting Grounding to Use*

Earlier I asked: What does it mean for one truth to *presuppose* another? Now I give the answer. I propose to understand this relation of presupposition in terms of grounding. More precisely, in terms of the concept of *partial*, *mediate* grounding.

Partial, mediate grounding '$A \prec B$' is the smallest relation on propositions such that

- $A \prec B$ if there are some Δ and A is among them and the Δ ground B,
- $A \prec B$ if there is some C such that $A \prec C$ and there are some Δ and C is among them and the Δ ground B.

We say that B presupposes A if A partially, mediately grounds B ($A \prec B$).

Note that this notion of presupposition allows for an alternative formulation of the non-circularity of grounding.

NON-CIRCULARITY There are no $\prec$-chains $A_0 \prec A_1 \prec \ldots \prec A_n$ such that $A_0 = A_n$.

On this basis I now turn to formulate my iterative conception of class.

9.2.10 *Iterative Conception of Proper Classes 1*

First, as noted earlier, the truths of set-theory are given and need not be grounded in class theory. I add this assumption to the general theory of grounding.

(BASIC TRUTHS) For every set-theoretic truth ϕ there are no class-theoretic propositions Γ such that $\Gamma \lhd \phi$.

We may also assume all *logically true* eta-proposition to be basic. But this I consider optional.

9.2.11 *Iterative Conception of Proper Classes 2*

I call a proposition *grounded* if it is mediately but fully grounded in basic truths, and restrict class comprehension to these grounded truths.

CC For every condition Phi and every object a, if either the proposition that a is a Phi, or the proposition that a is not a Phi is grounded, then the following holds: a is contained in the extension of Phi just in case a is a Phi.

This principle corresponds to the restricted schema of set-comprehension from the iterative conception of set, just that now it's about *classes* and *true propositions*.

Note that, in accordance with our proposition-perspective, we restrict which condition-object *pairs* may be inserted on the right-hand side of comprehension.

In effect, for the same concept, some instance may hold while another may not.

This is the schema of class comprehension of my iterative conception of class.

Conceptions are nice, but we want more. We want mathematical structures which model, in the scientist's sense, our philosophical conceptions.

For the iterative conception of set, there are many such models. Every initial segment of the cumulative hierarchy up to some limit rank models restricted set comprehension.

More precisely, we want a formal model for the comprehension principle just proposed.

10

PUZZLES OF GROUND

10.1 INTRODUCTION

In the previous chapter, I have applied the philosophical idea of metaphysical grounding (§**??** to the Kripkean concept of grounded truth (§**??**. I showed that the formal relation of **T-SK** priority, which orders the grounded truths, satisfies the principles of grounding. On this basis I proposed to understand the philosophical significance of Kripke's model constructions from the fact that his truth generator tracks the intuitive thought that it is true that ϕ because ϕ, and it is not true that ϕ because $\neg\phi$. In symbols:

$$\text{(GT)} \qquad \begin{array}{r} \phi < T\ulcorner\phi\urcorner \\ \neg\phi < \neg T\ulcorner\phi\urcorner \end{array}$$

However, my proposal faces a challenge. This very thought, which I proposed to be the philosophical content of semantic groundedness, brings us dangerously close to a certain family of paradoxes observed recently by Kit Fine 2010. In this chapter I will present these puzzles and discuss possible responses.

10.2 PUZZLES OF GROUND

In his 2010 "A Puzzle of Ground", Fine shows that some principles of groundedness are inconsistent (with principles of logic). This inconsistency, however, is no reason to give up the notion of grounding. Instead, Fine argues, we need to find a balance between the logical principles involved and the principles of grounding.

In a nutshell, the paradox is derived as follows.

> It's a fact that everything exists. This fact, call it f_0, is one thing that exists, so its existence contributes to making it the case that everything exists. So, everything exists *partly in virtue* of this fact's existing. Likewise, though, f_0 exists in virtue of everything existing. So everything exists partly in virtue of everything existing. This can't be. It contradicts the irreflexivity of grounding.

Several substantial assumptions about grounding as well as logical principles are involved in this reasoning. In his paper, Fine makes them explicit in terms of a formal theory of grounding. Fortunately, however, such a theory is already available to me, Fine's own *pure logic of ground* ('PLG') (§ 7.3). So let $\mathcal{L}_g$ be the language of first-order logic whose sole non-logical symbols are the grounding operators $<, \leqslant, \prec, \preceq$, governed by the rules of PLG and ILG.

In particular, recall that the impure logic of ground comprises the following rules.

$$\neg\wedge\ \frac{\neg\psi}{\neg\phi, \neg\psi < \neg(\phi \wedge \psi)} \qquad \neg\exists\ \frac{\neg\phi(\overline{o}) \qquad \neg\phi(\overline{p}) \qquad \ldots}{\neg\phi(\overline{o}), \neg\phi(\overline{p}), \ldots < \neg\exists x\big(\psi(x)\big)} \text{ for all o}$$

Now, let us add to this minimal theory of grounding firstly the following rule of truth introduction.

$$\neg\text{T-Intro}\ \frac{\neg\phi}{\neg T\ulcorner\phi\urcorner}$$

Note that this rule captures in the object-language one part of the truth generator **T** in terms of which I proposed to understand Kripke's model construction.

Secondly, let us add the principle that nothing is both true and false, in other words that our truth predicate is *consistent*.

(Cons) $$\neg\exists x(Tx \wedge T\dot{\neg}x)$$

In choosing this principle I deviate slightly from the way how Fine presents the puzzle of grounding and truth [**?**, p. 102f] Fine derives a contradiction from the principle that everything is either true or not. However, this principle is at odds with the Kripkean approach to truth. Although there are ways of making them compatible (cf. *closing off* partial models, p. 42), I prefer to avoid this complication and work with the assumption of consistency, which in turn goes well with semantic groundedness.

Finally, we add the principle GT which expresses neatly what I propose to be the philosophical content of semantic groundedness.

(GT)

$$\phi < T\ulcorner\phi\urcorner$$
$$\neg\phi < \neg T\ulcorner\phi\urcorner$$

This principle In his 2010, however, Fine shows that the resulting system of grounding and truth is inconsistent.

1. $\neg\exists x(Tx \wedge T\dot{\neg}x)$ (Cons), short: $\neg\epsilon$
2. $\neg(T\overline{o} \wedge T\dot{\neg}\overline{o})\ldots\neg(T\ulcorner\epsilon\urcorner \wedge T\dot{\neg}\ulcorner\epsilon\urcorner)\ldots < \neg\exists x(Tx \wedge T\dot{\neg}x)$ 1, $\neg\exists$
3. $\neg(T\ulcorner\epsilon\urcorner \wedge T\dot{\neg}\ulcorner\epsilon\urcorner) \prec \neg\epsilon$ 1, 2 Subsumption ($\substack{<\\ \prec}$)
4. $T\ulcorner\neg\epsilon\urcorner$ 1, T-Intro
5. $\neg T\ulcorner\epsilon\urcorner$ 1, $\neg$T-Intro
6. $\neg T\ulcorner\epsilon\urcorner \prec \neg(T\ulcorner\epsilon\urcorner \wedge T\dot{\neg}\ulcorner\epsilon\urcorner)$ 4,5,$\neg\wedge$, Subsumption ($\substack{<\\ \prec}$)
7. $\neg\epsilon \prec \neg T\ulcorner\epsilon\urcorner$ GT, Subsumption ($\substack{<\\ \prec}$)
8. $\neg\epsilon \prec \neg\epsilon$ 3, 6, 7, Transitivity of $\prec$
9. $\bot$ 7, Non-Circularity

10.3 SOLUTION ROUTES EXPLORED BY FINE

Having presented this and several other puzzles, Fine turns to explore possible solution routes (§§6,7). He argues that it would not help to retract to a notion of intransitive *immediate* ground, since the contradiction could still be derived using its ancestral. More generally, no notion of ground which allows for circularity can account for the intuition that if ϕ because ψ, then the fact that ϕ provides a satisfactory explanation as to why ψ, or at least any such explanation of the fact that ϕ can be extended to one of the fact that ψ (p. 105). Any notion of ground that answers to this intuition, however, will lead to the inconsistent reasoning.

Fine turns to objections against Factual Grounding. ϕ may be true without there to be a fact ψ, since

1. truth does not require correspondence to a fact, or
2. existence of $[\phi]$ may be independent of whether ϕ or not (thin notion of existence).

Fine reformulates the puzzle such that these moves do not help (p. 107). The usage of existence, namely, is not needed – equally well one may reason in terms of facts *obtaining*. This modification is straightforward for the particular version (the one based on the assumption 'Something exists' – now we use 'Some fact obtains'). In the case of the argument based on the universal statement, the universal quantifier is restricted to facts that obtain.

Next, Fine considers the thought that Factual Grounding fails for the reason that ϕ because of the fact that ϕ, and not the other way around as suggested by the axiom.[1] This route may be found plausible, for example, on a truth-maker account.

Nonetheless, Fine rejects it, as it leads to a vicious regress of grounds (p. 107).

Fine concludes from his discussion that the following principles are not to be called in question:

- Factual Grounding
- Propositional Grounding
- Truth introduction

A possible route is the restriction of

- Factual Existence
- Propositional Existence

1 Keep in mind that grounding is irreflexive.

which Fine identifies with Russell's predicativism, and rejects, too.

The remaining principles, however, are all justified from classical logic, Fine argues. This link is obvious for the logical principles

- Universal Middle
- Particular Middle
- Universal Existence
- Particular Existence

but needs some motivation in the case of the grounding principles.

- Universal Grounding
- Existential Grounding
- Disjunctive Grounding

Fine claims that these are implicit in the classical truth-conditions (p. 108). I do not see, however, the order of justification here. Does Fine suggest that the classical truth-conditions are based on groundedness considerations? In this case, classical logic surely could not be used to justify the grounding principles. Or is it rather that the theory of ground is just a way of spelling out classical semantics? In this case, the puzzle of ground seems just another truth paradox.

At any rate, Fine infers that classical logic is '... in tension with itself' (p. 108); whereby he must have in mind a broad understanding of logic that includes the metaphysical assumptions of factual and propositional grounding as well as a substantive fragment of truth theory.

Fine points out that one need not either abandon the logical principles in a wholesale manner or reject all of the grounding principles. A compromise is available. He suggest to weaken the principle of disjunctive grounding

$$\phi \rightarrow \phi \prec (\phi \vee \psi)$$

The underlying idea, Fine submits, is that a true disjunction is grounded in *one* of its disjuncts. The axiom, however, requires that if ϕ as well as ψ, *both* facts ground the fact that ϕ or ψ. It may therefore be replaced by the weaker principle

$$(\phi \vee \psi) \rightarrow [\phi \prec (\phi \vee \psi) \vee [\psi \prec (\phi \vee \psi)]$$

Similar considerations motivate *weak existential grounding*:

$$\exists x \phi(x) \rightarrow \exists x[\phi(y) \prec \exists \phi(x)]$$

This modest weakening of the system prevents the Particular Arguments from going through. These weakened principles of ground are compatible with the axioms of particular existence and particular middle. However, they still do not allow for the assumption of universal existence and middle, and no similar move is available to suppress the Universal Argument.

In sum, Fine lists four responses to the puzzle of ground, none of which is fully satisfactory but each has its own advantages and disadvantages.

- Predicativism: Endorse all logical and ground-theoretic assumptions but reject factual and propositional existence.
- Compromise impredicativism: Weaken disjunctive and existential grounding and reject universal existence and middle.
- Extremist, logic-sceptical impredicativism: Reject principles of classical logic but endorse ground-theoretic assumptions.
- Extremist, ground-sceptical impredicativism: Reject ground-theoretic assumptions but endorse principles of classical logic.

10.4 BEING TARSKIAN ABOUT GROUNDING

There is one option that Fine does not consider. We can fruitfully work with such principles of truth introduction and grounding as lead to Fine's puzzles. In fact, the way of dealing with truth and grounding I have in mind is precisely how I have done so in previous chapters: I spoke of grounding in the *meta-theory* where we construct our models of truth. So, there is an alternative to the solution routes of the previous section: let us be *Tarskian* about truth and grounding.

Some more detail is in order. I will show that we can consistently extend the language of truth by the expressive resources to speak about the grounding relations among propositions *expressed* in the *language of truth*.

Let $\mathcal{L}_{ta}$ be the language of truth from chapter 2. Note that first-order arithmetic whose induction scheme is extended to formulae with 'T' (the theory PAT), represents computably enumerable sets of $\mathcal{L}_{ta}$ sentences. Let Σ be such a set. Then there is an $\mathcal{L}_a$-formula $\sigma(x)$ in the language such that PAT $\vdash \sigma(\ulcorner\phi\urcorner)$ iff $\phi \in \Sigma$. Note further that the formula $\underset{\cdot}{\Sigma}$ itself is encoded in $\mathcal{L}_a$ by a term $\ulcorner\sigma\urcorner$. In the following, I will use these terms as $\mathcal{L}_a$-labels for c.e. sets of $\mathcal{L}_{ta}$-sentences.

Extend the language $\mathcal{L}_{ta}$ further by four *relation* symbols '$\tilde{<}$', '$\tilde{\leqslant}$', '$\tilde{\prec}$' and '$\tilde{\preceq}$'. The resulting language $\mathcal{L}_{gta}$ will be the metalanguage in which we can consistently speak about the grounding relations among facts expressed in $\mathcal{L}_{ta}$.

Recall the Finean theory of grounding, both the structural (the *pure* logic) and the logical principles (the *impure* logic). It is formulated in a language with sentential connectives '$<$', '$\leqslant$', '$\prec$' and '$\preceq$', and given by rules that allows us, for example, to infer $\psi \preceq \phi$ from $\psi \prec \phi$

(definition 39 above), or $\phi < (\phi \vee \psi)$ from ϕ (definition 40) Now, these rules are easily translated into rules for the language $\mathcal{L}_{gta}$. First, restrict the formation rules for the grounding connectives to sentences of the language of truth $\mathcal{L}_{ta}$ and sets thereof. Then, using the coding machinery of PAT, translate an atomic formula $\Sigma < \phi$ as $\ulcorner\sigma\urcorner \tilde{\precsim} \phi$, $\Sigma \prec \phi$ as $\ulcorner\sigma\urcorner \precsim \phi$ etc.. Thus, the rules mentioned become:

$$\text{Sub } (\precsim/\tilde{\preceq}) \; \frac{\ulcorner\psi\urcorner \precsim \ulcorner\phi\urcorner}{\ulcorner\psi\urcorner \tilde{\preceq} \ulcorner\phi\urcorner} \qquad \frac{\phi}{\phi \tilde{\precsim} (\phi \vee \psi)} \vee L, \qquad \phi \in \mathcal{L}_{ta}$$

And analogously for the other rules of PLG and ILG. Note, however, that the resulting rules will ever only allow us to reason about grounding relations between facts *expressed in* $\mathcal{L}_{ta}$.

Let the *Tarskian* theory of truth and grounding TTG be the least set of $\mathcal{L}_{gta}$-sentences containing KF+Cons, closed under these rules as well as the rule[2]

$$\frac{T\ulcorner\phi\urcorner}{\ulcorner\phi\urcorner \tilde{\precsim} \underset{\cdot}{T} \ulcorner\phi\urcorner} \text{GT}$$

As the theory of truth KF+Cons proves every instance of $T\ulcorner\phi\urcorner \to \phi$ [Halbach, 2011a, 15.19], TTG thus comprises principles that look dangerously close to what Fine's puzzle shows to be incompatible. The only relevant difference between TTG and the system of section 10.2 above is that a boundary is drawn between truth and grounding. Formally, the grounding rules only apply to $\mathcal{L}_{ta}$-sentences, and the principles of truth do not apply to sentences containing the new relation symbols '$\tilde{\precsim}$' etc. This saves grounding from paradox, in the precise sense that TTG has a model, indeed a very natural one.

Recall that $\Sigma <_{\mathbf{tsk}} \phi$ iff ϕ has a **T-SK**,Σ-priority tree, and $\phi \prec_{\mathbf{tsk}} \psi$ iff ψ has a **T-SK**-priority tree one of whose leaves is ϕ.

Definition 42. Let $\mathbb{N}(I^+_{sk}, <_{\mathbf{tsk}}, \prec_{\mathbf{tsk}})$ be the $\mathcal{L}_{gta}$ model such that

1. $\mathbb{N}(I^+_{sk}, <_{\mathbf{tsk}}, \prec_{\mathbf{tsk}}) \vDash T\ulcorner\phi\urcorner$ iff $\ulcorner\phi\urcorner \in I^+_{sk}$,
2. $\mathbb{N}(I^+_{sk}, <_{\mathbf{tsk}}, \prec_{\mathbf{tsk}}) \vDash \ulcorner\sigma\urcorner \tilde{\precsim} \ulcorner\phi\urcorner$ iff $\Sigma <_{\mathbf{tsk}} \phi$,
3. $\mathbb{N}(I^+_{sk}, <_{\mathbf{tsk}}, \prec_{\mathbf{tsk}}) \vDash \ulcorner\sigma\urcorner \tilde{\leqq} \ulcorner\phi\urcorner$ iff $\Sigma = \{\phi\}$ or $\Sigma <_{\mathbf{tsk}} \phi$,
4. $\mathbb{N}(I^+_{sk}, <_{\mathbf{tsk}}, \prec_{\mathbf{tsk}}) \vDash \ulcorner\psi\urcorner \precsim \ulcorner\phi\urcorner$ iff $\psi \prec_{\mathbf{tsk}} \phi$ and
5. $\mathbb{N}(I^+_{sk}, <_{\mathbf{tsk}}, \prec_{\mathbf{tsk}}) \vDash \ulcorner\psi\urcorner \tilde{\preceq} \ulcorner\phi\urcorner$ iff $\psi = \phi$ or $\psi \prec_{\mathbf{tsk}} \phi$.

Proposition 18.

$$\mathbb{N}(I^+_{sk}, <_{tsk}, \prec_{tsk}) \vDash \text{TTG}$$

Proof. Immediately from corollary 2. $\square$

2 I do not see the need to extend the induction scheme of PAT to formulae with $\tilde{\precsim}$ etc, and will not do so.

11 A MODAL LOGIC OF GROUNDEDNESS

I propose to use philosophical concepts of priority, in particular ontological dependence and metaphysical grounding, to spell out the philosophical significance of the formal concepts of groundedness, for example the well-foundedness of sets (§1.4) or Kripke's semantic groundedness (§2). In the previous chapter I have discussed the challenge to my proposal which arises from Fine's *puzzles*. I concluded that we can hold on to all principles involved if we separate the object-language of, say, truth, from the meta-language of grounding.

However, this Tarskianism about grounding and truth faces a challenge. In this chapter, I will present the objection, clarify it and develop a response.

11.1 THE GHOST OF THE HIERARCHY

The challenge is this: if we have to ascend to a meta-language in order to express the philosophical content of grounded truth, then the notion of grounded truth is not available to us in our own language.

This is by no means a new observation. Formally, the relevant relation of partial mediate grounding is just **TSK**-dependence $\prec_{\text{tsk}}$, for the Kripke truth generator **T** and the Strong Kleene logic generator **SK**. Thus, the challenge is closely linked to Kripke's 1975 remark that (Kripke 1975:714)

> [...] the induction defining the minimal fixed point is carried out in a set-theoretic meta-language, not in the object language itself. [...] The ghost of the Tarski hierarchy is still with us.

Accordingly, I will speak of the *ghost challenge*.

Let me clarify the challenge. It goes in four steps. The first step is to argue that we want to be able to say when a sentence is grounded, and when it is not. That is, we want to express groundedness. I think this is plainly right, and follows naturally from the basic idea underlying the groundedness approach to truth. The unrestricted T-schema is inconsistent. We want to restrict it to the grounded sentences. For this, we need to be able to say when a sentence is grounded.

The second step is to point out that we cannot express groundedness in the language of truth as it is given to us. This is certainly true.

If we regiment groundedness by a least fixed point construction, we can give a rigorous proof. We observe that the least fixed point of a jump operator which turns truth in a model into a new model (be it based on Strong Kleene or on some other monotone evaluation scheme) is essentially Π^1_1, whereas every set definable in arithmetic plus truth is Σ^1_1.

Thirdly, we observe that in a standard meta theory, say set theory, we can say when a sentence is grounded and when it is not.

Now, it is concluded that we must ascend to this meta-theory. This final step I would like to resist. As it stands, namely, this argument is sound only if we assume, in addition, that groundedness can *only* be expressed in a meta-theory. In other words, the ghost challenge requires that there is no way of saying when a sentence is grounded and when it is not other than in the usual, meta-theoretic manner. This is a strong assumption and has not been sufficiently corroborated. In fact, my goal in this paper is to give reasons to doubt it. I will outline an alternative route to expressing groundedness, one that does not lead us up a hierarchy of theories.

Before I do so, however, it is worth distinguishing the challenge I intend to address from the problem known in the trade as *revenge*. The object theory cannot express the fact that in the least fixed point model, the liar sentence is not true, or not determinately true. I will not attempt to answer the revenge objection.

To see that these are distinct problems, assume, for the sake of the argument, that we accept the revenge challenge. We accept that if we look at our theory from the outside, there are facts pertaining to truth which we cannot express using the truth predicate of our theory.

Then, the challenge from groundedness being a meta-theoretic notion is still pressing. For, assume that it is true that in order to say which sentences are grounded and which are not, we have to ascend to a meta-theory. Now let's assume we wish to speak of the grounded truths of our one, all-encompassing universal theory, from which we cannot ascend to any essentially stronger meta-theory. Then, it seems we cannot make sure that our truth-theoretic statements do not lead to paradox, since we cannot ascertain whether or not a given sentence is grounded.

If groundedness is an essentially meta-theoretic notion, then we cannot carry out the desired restriction of the T-schema to its grounded instances without ascending to a meta-theory. The groundedness approach to truth would not be available to us in our most general theory.

In a nutshell, the revenge problem is about *how much* we can do with the grounded truth predicate, whereas the ghost challenge is about *whether* the groundedness approach can be carried out in our own language in the first place.

To be explicit: The challenge I will attempt to answer is not about truth, but about groundedness.

11.2 SIDESTEPPING THE GHOST

In the rest of this chapter, my goal is to develop a new way of expressing groundedness, a regimentation that is not meta-theoretic.

So here's my response to the challenge: There is a way of expressing groundedness other than ascending to the meta-theory. We can use *tense* instead. English already has the vocabulary for this, and so would the language of our universal theory. But even if this wasn't the case, and we would have to extend our language by temporal operators, this would still not mean to ascend to a meta-theory.

Let me use the following metaphor. Whereas adding meta-theory makes up a *vertical* extension of our theory, my proposal is that of a *horizontal* enrichment. We do not need to let the ghost chase us up the hierarchy, we better sidestep it, which is anyway more convenient.

So, the theorist of grounded truth can respond to the challenge from the ghost. At least, she can answer the challenge to the extent that ascending to a meta-theory is a *vertical* enrichment – in the next section I will show how to do it *horizontally*.

At this point, many readers will sense a worry which I may as well address now. It goes as follows. Truth is absolute. It does not make sense to speak of something becoming true. Or at least: our response to paradox should not rely on a contentious temporalism, or even worse, relativism.

The temporal vocabulary is not supposed to be taken literally. Here's an analogy: the realist about set theory makes use of temporal vocabulary when expressing her iterative conception of sets, in particular, to express the priority of a set over its elements. My use of tense is just analogous. It is not used to express matter-of-fact temporal relations, but the idea of groundedness.

Below, I will give a natural model.

11.3 A LOGIC FOR GROUNDEDNESS

I now turn to implementing the proposal. I will formulate a theory of truth based on a logic of well-ordered time. However, what follows is merely one way of carrying out my idea, and I don't think that my philosophical proposal stands or falls by its success.

As usual, let $\mathcal{L}_{\text{ta}}$ be the language of first order arithmetic extended by a unary relation symbol 'T'. For simplicity, I assume that the language does not contain a primitive symbol '$\rightarrow$' and that we define the material conditional in terms of negation and disjunction.

I wish to enrich this familiar machinery by resources to express the groundedness of truth without ascending to a meta-theory. My goal is to enable a theory of grounded truth to express the priority of ϕ over $T\ulcorner\phi\urcorner$. I want it to be able to state that for $T\ulcorner\phi\urcorner$ to be true at some point, ϕ must have been true earlier.

In order to achieve this, I will *modalize* the first-order setting of truth. More precisely, I will add the resources of *tense logic*. Tense logic is formulated using two primitive modal operators, one operator **G** looking *forwards* and reading "it will always be the case that ...", another operator **H** looking *backwards* and reading "it has always been the case that ...".

I extend $\mathcal{L}_{ta}$ further by these two operators $\mathbf{G}$ and $\mathbf{H}$ to the language $\mathcal{L}_{tam}$ ('*m*' for *modalized*). As usual, we define $\mathbf{P}\phi$ as $\neg\mathbf{H}\neg\phi$, and $\mathbf{F}$ analogously. Occasionally, I will use $\mathbf{A}\phi$ (read: 'it is always the case that' ϕ) as a meta-linguistic abbreviation for $\mathbf{H}\phi \wedge \phi \wedge \mathbf{G}\phi$, and $\mathbf{S}$ (read: 'it is sometimes the case that' ϕ) as short for $\mathbf{P}\phi \vee \phi \vee \mathbf{F}\phi$ [Garson, 1984, p. 292]. Although it comprises two modal operators $\mathbf{G}$ and $\mathbf{H}$, the language $\mathcal{L}_{tam}$ is interpreted in ordinary models (W, R, D, d) of first-order modal logic. $\mathbf{G}\phi$, on the one hand, holds at some point w from W iff it holds at every point R-accessible from w, that is, at every point v such that wRv. $\mathbf{H}\phi$, on the other hand, holds at w iff it holds at every point *conversely* accessible, that is, at every point v such that vRw. In effect, $\mathbf{G}$ "looks forward" and $\mathbf{H}$ "looks backward".

This already allows us to express the first component of my intuitive gloss on groundedness. The truth of $T\ulcorner\phi\urcorner$ presupposes the truth of ϕ, that is: $T\ulcorner\phi\urcorner$ only if ϕ earlier, that is: $T\ulcorner\phi\urcorner \to \mathbf{P}\phi$.

I will assume that time is well-ordered. Formally, I will restrict attention to models (W, R, D, d) such that R well-orders W.[1]

Quantified modal logic is hard, both technically and philosophically. Fortunately, I do not have to deal with its subtleties. All I want to say is that *truth* changes over time. Therefore, I can let the quantifiers be governed by standard, non-free first-order logic, and assume all terms to be *rigid designators* [Garson, 1984]. The result is a basic first-order logic of well-ordered time: "*woq*". I recall some basic definitions.

Definition 43 (Validity and Consequence in Quantified Modal Logic)**.** Let $\mathfrak{F}$ be any frame (W, R) and $\mathfrak{M}$ be any model of first-order modal logic based on $\mathfrak{F}$, we say that a sentence ϕ is *valid* in $\mathfrak{M}$ iff for every $w \in W$, $\mathfrak{M} \models \phi[w]$. We call ϕ *valid* in $\mathfrak{F}$ iff for every model $\mathfrak{M}$, ϕ is valid in $\mathfrak{M}$. Finally, let $\mathfrak{f}$ be a class of frames (e.g. the well-ordered frames *wo*) we say that ϕ is a *consequence over* $\mathfrak{f}$ of some set of sentences Σ (in symbols: $\Sigma \models_{\mathfrak{f}} \phi$ iff for every model $\mathfrak{M}$ based on $\mathfrak{F} = (W, R)$ and every $w \in W$, if $\mathfrak{M} \models \Sigma[w]$ then $\mathfrak{M} \models \phi[w]$.

Thus, I will write $\Sigma \models_{woq} \phi$ if for every *woq*-model (W, R, D, d) and every $w \in W$, if $(W, R, D, d) \models \Sigma$ then $(W, R, D, d) \models \phi$.

Note that *woq* validates the Barcan formulae for both operators.

The first-order logic of well-ordered time is a powerful tool. For two structures S and S' I write $S \simeq S'$ for the statement that there is an isomorphism between S and S'.

Theorem 2 (Scott, Garson)**.** *Let $\mathcal{L}_{am}$ be the language of first-order arithmetic plus tense operators '$\mathbf{G}$' and '$\mathbf{H}$'. Extend $\mathcal{L}_{am}$ by a unary predicate symbol 'N' to the language $\mathcal{L}_{amp}$. There are $\mathcal{L}_{amp}$ sentences Σ such that the following holds.*

1 Recall that a relation R on W is a well-order iff it is linear (transitive, anti-symmetric, comparable) and in every set of worlds $V \subseteq W$ there is an R-least world.

Let (W, R, D, d) *a constant-domain model of the first-order logic of well-ordered time* woq. *Then we have that for every point* $w \in W$*,*

$$(W, R, D, d) \models \Sigma[w] \Rightarrow (D, d) \simeq \mathbb{N}$$

That is, every point that satisfies Σ *is in fact a* standard *model of arithmetic.*

Proof. See Garson (1984), sections 3.2.2 and 3.2.3.
(Sketch) Let Σ comprise the following

N1 $\forall x \mathbf{P}(Nx \wedge \mathbf{H}\neg Nx \wedge \mathbf{G}\neg Nx)$

- "Every object of the domain has the property N at exactly one time"

N2 $\mathbf{A}\forall x\forall y\big((Nx \wedge Ny) \to x = y\big)$

- "No two things have the property N at the same time"

Thus, every model of N1 and N2 will have an injective function from the (possibly non-standard) domain of $\mathfrak{M}$ into the well-ordered set W.

Note that for every model $\mathfrak{M} = (W, R, D, d)$ such that P1$\wedge$P2 is true at some $w \in W$, we have that $\mathfrak{M} \models \mathbf{S}(P\overline{n} \wedge \mathbf{F}P\overline{m})[v]$, $v \in W$, just in case $P\overline{n}$ is true at an R-earlier point than $P\overline{n}$. Thus, N1 and N2 have allowed us to define, by $\mathbf{S}(Px \wedge \mathbf{F}Py)$, a restriction R^N of the well-ordered relation R to those points that some object of the domain is mapped to.

Σ contains another axiom. In order to avoid confusion with the operator $\mathbf{S}$, assume that the successor function is denoted in $\mathcal{L}_{\text{amp}}$ by '$'$'.

N3 $\forall x\forall y\Big(y = x' \leftrightarrow x\underset{\cdot}{R}y \wedge \forall z\big((z \neq y \wedge x\underset{\cdot}{R}z) \to y\underset{\cdot}{R}z\big)\Big)$

- "y is the successor of x just in case y is the least z $\underset{\cdot}{R}$-greater than x"

Finally, add to Σ the axioms of Robinson arithmetic, in particular:

Q0 $\forall x(x' \neq 0)$

Q1 $\forall x\big(x = 0 \vee \exists y(x = y')\big)$

Now let $w \in W$ and assume that $(W, R, D, d) \models \Sigma[w]$. As said before, the fact that (W, R, D, d) satisfies N1 and N2 implies that the objects of the domain D are mapped injectively to points in W. These are well-ordered by R, and $x\underset{\cdot}{R}y$ expresses precisely this restriction R^N of R. N3 ensures that $\overline{n} = \overline{m}'$ holds at w just in case n is the R^N-successor of m.

Thus, Q0 ensures that '0' denotes the R^N-least point w_0 in W, and Q1 that R^N is an ω-sequence. Hence, the objects in the domain at w are (isomorphic to) the standard numbers, and $\mathfrak{M} = \mathfrak{N}$. □

Corollary 3. *The logic $\models_{\text{woq}}$ is not axiomatizable.*

Proof. (Idea, for details see Garson (1984:294).) Analogously to how we infer the incompleteness of second-order logic from Gödel's incompleteness theorem and the fact that second-order arithmetic is categorical.

The key observation is that for any sentence ϕ of first-order arithmetic,

$$\models_{woq} \Sigma \to \phi \Rightarrow \mathfrak{N} \models \phi$$

□

Hence, there is no proof procedure complete with respect to the first-order logic of well-ordered time *woq*. This stands in contrast with many other first-order logics of time, that have such complete axiomatizations. For example, the logic of linear time is complete. In the following, I will therefore develop a theory of grounded truth on the basis of linear time. As we will see, it will allow us to approximate syntactically what we have just found to be beyond the reach of axiomatization, the logic of well-ordered time.

Unlike *woq*, this logic of linear time *lq* is axiomatizable. There are complete proof procedures for *lq*, for example systems of labelled tableaux [Priest, 2008, 14.7.12]. However, as I will largely reason about rather than within *lq*, it will prove useful to work with a Hilbert-style axiomatic proof system. Thus, let $\Sigma \vdash_{\text{lq}} \phi$ if there is a proof of ϕ from Σ in a Hilbert-style axiomatization of the first-order logic of linear time *lq*

Since, of course, every well-ordering is a linear order but not vice versa, the logic of well-ordered time *woq* is properly stronger than *lq*. In particular, complete proof systems for *lq* are sound with respect to the logic of well-ordered time *woq*.

$$\Sigma \vdash_{lq} \Rightarrow \Sigma \models_{woq}$$

11.4 A MODAL LOGIC OF GROUNDED TRUTH

I now formulate a theory of truth MGT in the logic of linear time *lq*. My goal is to capture the notion of grounded truth over first-order arithmetic. Accordingly, MGT is based on first-order arithmetic. More precisely, it includes first order Peano Arithmetic ('PA'), whose induction schema we generalize to the extended language $\mathcal{L}_{\text{tam}}$. Further, our modal logic is intended to express the step-by-step construction of a type-free truth predicate over arithmetic. The base theory, however, is not subject to this development. It holds at every stage of the construction, hence *necessarily* in our chosen modal logic. Consequently, I put an '**A**' ("always") in front of every PA axiom. Let '**A**PA' denote the resulting $\mathcal{L}_{\text{tam}}$ theory. As a result, MGT proves $\mathbf{A}\phi$ for every PA theorem ϕ in the language of arithmetic.

Lemma 17. *Let ϕ be any $\mathcal{L}_a$-sentece.*

$$\text{PA} \vdash \phi \Rightarrow \mathbf{A}\text{PA} \vdash_{1q} \mathbf{A}\phi$$

Arithmetic is a convenient base. However, this choice of a base theory is not essential to what follows.

Now, I add axioms to govern the interaction of 'T' with our modal vocabulary.

Ground $\mathbf{S}(\forall x \neg Tx)$

The axiom (Ground) corresponds to how I formulated the second component of groundedness, well-foundedness constraint on presupposition: once, nothing was true.

These axioms couched in a first-order logic of linear time provide a general framework for theory of grounded truth. By itself, however, it leaves open whether anything at all becomes true at some point.

What needs to be added now are axioms of truth-introduction.

My goal in the present paper is specific. I want a theory to express the notion of groundedness as captures by Kripke's model construction. Thus, I do not want it just to say that more and more sentences become true, but that this happens according to the *Strong Kleene* jump. Hence, I need axioms that say how truth is introduced according this evaluation scheme. However, the logic of tense we have chosen as our framework is based on classical first-order logic. And it is generally desirable to stick to a classical setting. In sum, we need axioms that express truth introduction according to the Strong Kleene jump operator, and do so in classical logic.

Fortunately, such axioms are available, in the truth axioms of the system KF (for *Kripke-Feferman*). They express, in the object language of arithmetic plus 'T', that T is closed under the semantic clauses of a partial model $\mathbb{N}(X^+, X^-)$ [Halbach, 2011b, p. 204]. As a consequence, KF characterizes precisely the fixed points of the Strong Kleene jump operator (ibid., theorem 15.15). And, KF is a classical theory. So, I will use the KF truth axioms to characterize, in the object-language, how in the course of Kripke's model construction, more and more sentences become true.

For this, however, the KF axioms need to be modified in an important way. Since, as they stand, they describe an arbitrary fixed point and not the step-by-step *construction*. For example, one KF axiom is that a sentence ϕ is true if and only if it is true that ϕ is true. What we would like to say, however is that if ϕ is true then it *will be* true that it is true that ϕ is true, and it is true that ϕ is true only if ϕ has been true earlier.

But precisely for this purpose we have availed ourself of tense logic. So, I will reformulate the KF axioms using its operators.[2]

2 '$\underset{\cdot}{Sent}_{ta}$' arithmetizes the syntactic property of being an $\mathcal{L}_{ta}$-sentence. Note that this property is definable in arithmetic by a formula all of whose quantifiers are bounded. Hence, PA$\vdash \underset{\cdot}{Sent}_{ta}(\ulcorner\phi\urcorner)$ iff ϕ is an $\mathcal{L}_{ta}$-sentence.

TKF1 $\mathbf{A}\forall x\forall y\,((Tx\underset{\cdot}{=}y \to \mathbf{P}x = y) \wedge (x = y \to \mathbf{F}\,Tx\underset{\cdot}{=}y \wedge \mathbf{G}\,Tx\underset{\cdot}{=}y))$

TKF2 $\mathbf{A}\forall x\forall y\,\big((Tx\underset{\cdot}{\neq}y \to \mathbf{P}x \neq y\big) \wedge \big(x \neq y \to \mathbf{F}\,Tx\underset{\cdot}{\neq}y \wedge \mathbf{G}\,Tx\underset{\cdot}{\neq}y)\big)$

TKF12 $\mathbf{A}\forall x\Big(\big(T\underset{\cdot}{T}x \to \mathbf{P}(Tx \wedge \mathbf{H}\neg Tx)\big) \wedge \big(Tx \to \mathbf{F}\,T\underset{\cdot}{T}x \wedge \mathbf{G}\,T\underset{\cdot}{T}x\big)\Big)$

TKF13 $\mathbf{A}\forall x\Big(\big(T\underset{\cdot}{\neg}\underset{\cdot}{T}x \to (\mathbf{P}T\underset{\cdot}{\neg}x \vee \neg \mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(x))\,\big) \wedge \big((T\underset{\cdot}{\neg}x \vee \neg \mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(x)) \to$
$\mathbf{F}\,T\underset{\cdot}{\neg}\underset{\cdot}{T}x \wedge \mathbf{G}\,T\underset{\cdot}{\neg}\underset{\cdot}{T}x)\Big)$

Note the first conjunct of TKF12. It says that if it is true that it is true that ϕ, not only at some point in the past it was the case that it is true that ϕ, but in fact there was an earliest such point.

What about the connectives and quantifiers? In Kripke's construction, at every stage, truth is closed under Strong Kleene logic. This closure is expressed by the remaining KF axioms which govern the interaction of 'T' with the quantifiers and connectives other than '$\neg$'. Hence, I add these axioms as they are, merely putting an 'always' ('$\mathbf{A}$') in front.

TKF3 $\mathbf{A}\forall x\,(\mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(x) \to (T\underset{\cdot}{\neg}\underset{\cdot}{\neg}x \leftrightarrow Tx))$

TKF4 $\mathbf{A}\forall x\forall y\,(\mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(x\underset{\cdot}{\wedge}y) \to (Tx\underset{\cdot}{\wedge}y \leftrightarrow Tx \wedge Ty))$

TKF5 $\mathbf{A}\forall x\forall y\,(\mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(x\underset{\cdot}{\wedge}y) \to (T\underset{\cdot}{\neg}\,(x\underset{\cdot}{\wedge}y) \leftrightarrow T\underset{\cdot}{\neg}x \vee T\underset{\cdot}{\neg}y))$

TKF6 $\mathbf{A}\forall x\forall y\,(\mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(x\underset{\cdot}{\vee}y) \to (Tx\underset{\cdot}{\vee}y \leftrightarrow Tx \vee Ty))$

TKF7 $\mathbf{A}\forall x\forall y\,(\mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(x\underset{\cdot}{\vee}y) \to (T\underset{\cdot}{\neg}\,(x\underset{\cdot}{\vee}y) \leftrightarrow T\underset{\cdot}{\neg}x \wedge T\underset{\cdot}{\neg}y))$

TKF8 $\mathbf{A}\forall y\forall x\,(\mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(\underset{\cdot}{\forall}yx) \to (T\underset{\cdot}{\forall}yx \leftrightarrow \forall z(\mathit{ClTm}(z) \to Tx(z/y))))$

TKF9 $\mathbf{A}\forall y\forall x\,(\mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(\underset{\cdot}{\forall}yx) \to (T\underset{\cdot}{\neg}\underset{\cdot}{\forall}yx \leftrightarrow \exists z(\mathit{ClTm}(z) \wedge T\underset{\cdot}{\neg}x(z/y))))$

TKF10 $\mathbf{A}\forall y\forall x\,(\mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(\underset{\cdot}{\exists}yx) \to (T\underset{\cdot}{\exists}yx \leftrightarrow \exists z(\mathit{ClTm}(z) \wedge Tx(z/y))))$

TKF11 $\mathbf{A}\forall y\forall x\,(\mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(\underset{\cdot}{\exists}yx) \to (T\underset{\cdot}{\neg}\underset{\cdot}{\exists}yx \leftrightarrow \forall z(\mathit{ClTm}(z) \to T\underset{\cdot}{\neg}x(z/y))))$

The result is my *modal logic of grounded truth* MGT: always arithmetic, the axiom of ground, and KF turned into axioms of step-by-step truth introduction.

Remark 2. Note that I understand KF as not including the consistency of truth

$$\forall x(\mathit{Se\underset{\cdot}{n}t}_{\mathrm{ta}}(x) \to \neg(Tx \wedge T\underset{\cdot}{\neg}x))$$

Often, this formula is added to the KF axioms, since it ensures a number of pleasing results. I do not have to do though, since in the present, tensed context, consistency can be shown to follow. See theorem 4 below.

Note that the first conjuncts of axioms TKF1,2,12 and 13 allow us, if we have established a certain atomic sentence, to introduce respectively iterate the truth predicate *later*. For example, TKF1 allows us to infer, from $x = y$, that at some point later in time, $T x \dot{=} y$. More generally, we can show that whenever a sentence is ascribed truth at some point, then it will remain so henceforth.

Lemma 18.

$$\mathrm{MGT} \vdash_{lq} \mathbf{A}\forall x\big(\mathrm{S\dot{e}nt}_{ta}(x) \rightarrow (Tx \rightarrow \mathbf{G}Tx)\big)$$

Proof. (*sketch*) Induction on positive complexity (see p. 53) within MGT. As the PA induction has been extended to the language with truth predicate, firstly the positive complexity of an $\mathcal{L}_{ta}$ formula ϕ is represented in PA by a functional term $PC^{\bullet}$ such that $\mathrm{PA} \vdash PC^{\bullet}(\ulcorner\phi\urcorner) = \overline{n}$ iff the positive complexity of ϕ is n. Secondly, PA proves the following induction principle.[3]

$$\begin{aligned}\forall x\Big(&Se\dot{n}t_{ta}(x) \wedge \forall y\big(Se\dot{n}t_{ta}(y) \wedge PC^{\bullet}(y) \leqslant PC^{\bullet}(x) \wedge (Ty \rightarrow \mathbf{G}Ty)\big)\\ &\rightarrow (Tx \rightarrow \mathbf{G}Tx)\Big)\\ &\rightarrow \forall x\big(Se\dot{n}t_{ta}(x) \rightarrow (Tx \rightarrow \mathbf{G}Tx)\big)\end{aligned}$$

The base cases are then taken care of by the axioms TKF1,-2,-12 and -13; the induction step by the compositionality axioms TKF4-11 and the induction hypothesis. □

Corollary 4.

$$\mathrm{MGT} \vdash_{lq} \mathbf{A}\forall x\big(\mathrm{S\dot{e}nt}_{ta}(x) \rightarrow (\neg Tx \rightarrow \mathbf{H}\neg Tx)\big)$$

As I will show in the next section, the theory MGT has natural models. They provide a strong case for MGT as a theory of *grounded* truth. However, already from a proof-theoretic point of view, MGT has several desirable properties, as we will see in the remainder of the present section.

The axioms TKF1 through TKF13 are well viewed as a modalization of KF. It is natural to ask how the system MGT relates to the standard, non-modal theory KF. To answer this question, we translate the language of truth $\mathcal{L}_{ta}$ into the language $\mathcal{L}_{tam}$.

Definition 44. We define a mapping $(\cdot)^{\star}$ from the $\mathcal{L}_{ta}$-formulae into the $\mathcal{L}_{tam}$-formulae. Let a, b $\mathcal{L}_{ta}$ terms and ϕ, ψ be $\mathcal{L}_{ta}$ formulae.

$$\begin{aligned}(a = b)^{\star} &= a = b\\ (Ta)^{\star} &= \mathbf{S}\,Ta\\ (\neg\phi)^{\star} &= \neg(\phi)^{\star}\\ (\phi \vee \psi)^{\star} &= (\phi)^{\star} \vee (\psi)^{\star}\\ (\forall x\phi)^{\star} &= \forall x(\phi)^{\star}\end{aligned}$$

3 See also [Halbach, 1996, pp. 40f]

Lemma 19. *For every set of $\mathcal{L}_{ta}$-formula Γ and every $\mathcal{L}_{ta}$-formula ϕ, if there is a proof of ϕ from Γ in first order logic then there is a proof of $(\phi)^\star$ from $(\Gamma)^\star$ in* loq.

$$\Gamma \vdash \phi \Rightarrow (\Gamma)^\star \vdash_{lq}$$

Proof. By an induction on the length of proof l, exploiting the fact that our mapping $(\phi)^\star$ translates connectives and quantifiers homophonically.

If $l = 1$, then $\phi \in \Gamma$ or ϕ an axiom of first order logic. If $\phi \in \Gamma$, then we also have: $(\phi)^\star \in (\Gamma)^\star$, hence $(\Gamma)^\star \vdash_{woq} (\phi)^\star$. If ϕ is a logical axiom, then so is its translation $(\phi)^\star$, since our function $(\cdot)^\star$ translates the connectives and quantifiers homophonically.

Consider proof of length $n+1$, and assume that for proofs of length $\leqslant n$, if $\Delta \vdash \psi$ then $(\Delta)^\star \vdash_{lq} (\psi)^\star$. Then $\Gamma = \Delta \cup \{\psi\}$ and ϕ is obtained from ψ by one application of Generalization (Gen), or $\Gamma = \Delta \cup \{\psi, \psi \to \phi\}$ and ϕ is obtained from ψ by one application of Modus Ponens (MP).

(MP): $(\Gamma)^\star = (\Delta)^\star\{(\psi)^\star, (\psi \to \phi)^\star\}$ and $(\Gamma)^\star \vdash_{lq} (\phi)^\star$ since *lq* extends classical first order logic and in particular is closed under Modus Ponens.

(Gen): $(\Gamma)^\star = (\Delta)^\star\{(\psi)^\star\}$, and $(\Gamma)^\star \vdash_{lq} (\phi)^\star$ since the translation function leaves quantifiers untouched and *lq* comprises ordinary first order logic. □

Lemma 20. *Let ϕ be a KF axiom. We have that*

$$\mathrm{MGT} \vdash_{lq} (\phi)^\star$$

Proof. By completeness, it suffices to show that if ϕ is a KF axiom then

$$\mathrm{MGT} \vDash_{lq} (\phi)^\star$$

That is, for every linearly ordered, constant domain model $\mathfrak{M} = (W, R, D, d)$ and every point $w \in W$, if $\mathfrak{M} \vDash [w]\mathrm{MGT}$ then $\mathfrak{M} \vDash [w](\phi)^\star$.

(KF1) We wish to show that $\mathrm{MGT} \vDash_{woq} (\forall x \forall y(T x \dot{=} y \leftrightarrow x = y))^\star = \forall x \forall y(\mathbf{S}\, T x \dot{=} y \leftrightarrow x = y)$. ($\leftarrow$) Assume $x = y$. By TKF1, first conjunct, we have that $\mathbf{F}, T x \dot{=} y$. Hence, by definition, $\mathbf{S}\, T x \dot{=} y$, as desired.

($\rightarrow$) Assume $\mathbf{S}\, T x \dot{=} y$. There is a point such that $T x \dot{=} y$. Consequently, $x = y$ at some preceding point. But then, by RT, $\mathbf{A} x = y$ holds at that point, which makes $x = y$ hold at every point, including, by the linearity of R, the one we started out from, as desired.

Analogously, we show that $\mathrm{MGT} \vDash_{lq} (\mathrm{KF2})^\star$.

(KF12) We wish to show that $\mathrm{MGT} \vDash_{lq} \forall x(\mathbf{S}\, T \dot{T} x \leftrightarrow \mathbf{S}\, T x)$. ($\rightarrow$) Assume $\mathbf{S}\, T \dot{T} x$. Hence, at some point, $T \dot{T} x$. By TKF12, $\mathbf{P} T x$ and by definition, $\mathbf{S}\, T x$. ($\leftarrow$) Assume $\mathbf{S}\, T x$ and go to the point α where $T x$. By TKF12, therefore, $T \dot{T} x$ holds at at some later point β, and we conclude that α witnesses $\mathbf{S}\, T \dot{T} x$.

(KF13) Our goal is to show that MGT$\vDash_{lq}$ $\forall x(\mathbf{S}\,T\dot{\neg}\dot{T}x \leftrightarrow \mathbf{S}\,T\dot{\neg}(x) \vee \neg \mathit{Sent}_{ta}(x))$. ($\rightarrow$) As before, we go to a point α that witnesses $\mathbf{S}\,T\dot{\neg}\dot{T}x$, where we use TKF13 to infer $\mathbf{P}T\dot{\neg}x$. Moving on to this statement's witness $\beta < \alpha$, we find that here, $T\dot{\neg}x$ holds. We conclude that $\mathbf{S}\,T\dot{\neg}x$ holds at our starting point, as desired. ($\leftarrow$) If $\neg \mathit{Sent}_{ta}(x))$ or $\mathbf{S}\,T\dot{\neg}x$, we proceed as before, applying TKF13's first conjunct, and conclude $\mathbf{S}\,T\dot{\neg}\dot{T}x$.

(KF3-11) The compositionality axioms are all treated similarly, stripping off $\mathbf{S}$ and at that point, applying the relevant MGT axiom to obtain a witness for the desired claim. For example, to show the right-to-left direction MGT$\vDash_{lq}$ (KF3)* we go to the witness α of $\mathbf{S}\,Tx$. There, we make use of the fact that $T\dot{\neg}\dot{\neg}x \leftrightarrow Tx$ holds at this point, too, to conclude that the point witnesses the desired claim $\mathbf{S}\,T\dot{\neg}\dot{\neg}x$. □

Theorem 3. *MGT interprets KF.*

$$\mathrm{KF} \vdash \phi \Rightarrow \mathrm{MGT} \vdash_{lq} (\phi)^\star$$

Proof. From lemma 20 we know that MGT proves the translations of all KF axioms. Since lemma 19 ensures that derivation in KF is preserved, too, the claim is verified by a simple induction on the length of proof. □

This is a pleasing and useful result. Much is known about the interpretability strength of KF [Halbach, 2011b, §15.3]. Theorem 3 allows us to exploit these facts to relate MGT to Tarski's theory of truth and ramified analysis. Let '$\mathrm{RT}_{<\alpha}$' denote the theory of Tarskian truth over arithmetic iterated up to the ordinal α, and let '$\mathrm{RA}_{<\alpha}$' denote the theory of predicative second-order arithmetic, iterated up to the ordinal α. Recall that ϵ_0 is the limit of the sequence $\omega, \omega^\omega, \omega^{\omega^\omega}, \ldots$.

Corollary 5. *MGT interprets $RA_{<\epsilon_0}$ and $RT_{<\epsilon_0}$.*

In the precise sense of theorem 3, nothing is lost by couching KF in the logic of linear time. In fact, much is gained. MGT is strictly stronger than KF. KF, on the one hand, does not prove the consistency of truth, more precisely KF$\nvDash$ $\forall x(\mathit{Sent}_{ta}(x) \rightarrow \neg(Tx \wedge T\dot{\neg}x))$. To see this, recall firstly that by Feferman's classical result, for every Strong Kleene fixed point S, $\mathbb{N}(S)$ is a model of KF, and secondly that there are fixed points that contain both the liar sentence and its negation.

MGT, on the other hand, proves $\mathbf{A}\forall x(\mathit{Sent}_{ta}(x) \rightarrow \neg(Tx \wedge T\dot{\neg}x))$. To see this, we first need to introduce some terminology.

Definition 45 (T-complexity)**.** Equations $x = y$ have T-complexity 0. The T-complexity of $T\ulcorner\psi\urcorner$ is one greater than the T-complexity of ψ. $\neg\phi$ and $\exists x\phi$ inherit the T-rank of ϕ. The T-complexity of $\phi \wedge \psi$ and $\phi \vee \psi$ is the T-complexity of ϕ or ψ, whichever greater.

Theorem 4. *The modal logic of grounded truth proves the necessary consistency of truth. For every $\mathcal{L}_{ta}$-sentence ϕ,*

$$\mathrm{MGT} \vdash_{lq} \mathbf{A}\forall x(\mathrm{Sent}_{ta}(x) \rightarrow \neg(Tx \wedge T\dot{\neg}x)$$

Proof. For simplicity, I present the proof of the following schema which, however, is emulated within MGT similarly to how we proved lemma 11.4.

$$\neg(T\ulcorner\phi\urcorner \wedge T\dot{\neg}\ulcorner\phi\urcorner)$$

Let ϕ be any $\mathcal{L}_{ta}$-sentence. By completeness, it suffices to show that MGT $\models_{lq} \mathbf{A}\neg(T\ulcorner\phi\urcorner \wedge T\ulcorner\neg\phi\urcorner)$. So let $\mathfrak{M}$ be any *lq* model (W, R, D, d), let w be a point in W and assume that $\mathfrak{M} \models$ MGT. We show that $\mathfrak{M} \models \mathbf{A}\neg(T\ulcorner\phi\urcorner \wedge T\ulcorner\neg\phi\urcorner)[w]$ and do so by an induction on the T-complexity of ϕ (recall definition 45). At its base, assuming that ϕ is arithmetical, we run an induction on the *syntactic* complexity. So let ϕ be an atomic formula of arithmetic, i.e. an equation $a = b$. (we are now at the base of the inner of two nested inductions). Assume, for contradiction, that $T\ulcorner a = b\urcorner \wedge T\ulcorner a \neq b\urcorner$ holds at w. Then, by TKF1 and TKF2, $\mathbf{P}x = y \wedge \mathbf{P}x \neq y$. Hence, at some point uRv, $x = y$ and at some point $u'Rv$, $x \neq y$. But by lemma 17, $\mathbf{A}x = y \wedge \mathbf{A}x \neq y$, contradiction. hence $\neg(T\ulcorner a = b\urcorner \wedge T\ulcorner a \neq b\urcorner)$ at v, as desired.

For complex arithmetical sentences ϕ, the claim that $\mathbf{A}\neg(T\ulcorner\phi\urcorner \wedge T\dot{\neg}\ulcorner\phi\urcorner)$ follows from the axioms TKF3-11 and the induction hypothesis. For example, $T\ulcorner\neg\psi\urcorner \wedge T\ulcorner\neg\neg\psi\urcorner$ becomes $T\ulcorner\neg\psi\urcorner \wedge T\ulcorner\psi\urcorner$ by TKF3, which directly contradicts our induction hypothesis.

Now assume ϕ to be of T-complexity $n+1$, and assume that for all sentences ψ of lower complexity, $\mathfrak{M} \models \neg(T\ulcorner\psi\urcorner \wedge T\ulcorner\neg\psi\urcorner)[w]$. Again, we conduct an induction on the syntactic complexity of ϕ. If ϕ is atomic, we know that $\phi = T\ulcorner\psi\urcorner$ for some sentence ψ of T-complexity n. Assume, for contradiction, that $T\ulcorner T\ulcorner\psi\urcorner\urcorner \wedge T\neg T\ulcorner\psi\urcorner$ holds at w. By the right-hand conjuncts of TKF12 and TKF13, we know that $\mathbf{P}T\ulcorner\psi\urcorner \wedge \mathbf{P}T\ulcorner\neg\psi\urcorner$ is true at w. Hence, for some v, uRw, $T\ulcorner\psi\urcorner$ is true at v and $T\ulcorner\neg\psi\urcorner$ is true at u. Since R is a linear ordering, we can assume without loss of generality that uRv. By lemma 18 we have that $T\ulcorner\neg\psi\urcorner$ must hold at all points R-later than u, in particular at v. Hence, $\mathfrak{M} \models T\ulcorner\psi\urcorner \wedge T\ulcorner\neg\psi\urcorner[v]$. Again, since vRw lemma 18 allows us to infer that this conjunction $T\ulcorner\psi\urcorner \wedge T\ulcorner\neg\psi\urcorner$ holds at at the point w, contrary to our induction hypothesis.

The induction step, at which we assume ϕ to be complex, is taken care of, as before, by the axioms TKF3-11 and the induction hypothesis that for every constituent ψ of ϕ, $\mathfrak{M} \models \neg(T\ulcorner\psi\urcorner \wedge T\ulcorner\neg\psi\urcorner)[w]$. For example, let ϕ be $\exists x\psi$, and assume, for contradiction, that $\mathfrak{M} \models \neg(T\ulcorner\exists x\psi\urcorner \wedge T\ulcorner\neg\exists x\psi\urcorner)[w]$. The axioms TKF10 and TKF11 allow us to infer that $\exists y(ClTm(y) \wedge T\ulcorner\psi(\dot{y}/x)\urcorner) \wedge \forall y(ClTm(y) \rightarrow T\ulcorner\neg\psi(\dot{y}/x)\urcorner)$ must hold at w. Let y_0 witness the first conjunct, and specialize the second conjunct to it. We get that at w, it must be that $\mathfrak{M} \models T\ulcorner\psi(\dot{y}_0/x)\urcorner \wedge T\ulcorner\neg\psi(\dot{y}_0/x)\urcorner[w]$. This, however, contradicts our induction hypothesis.

This completes the proof that for every *lq*-model $\mathfrak{M}$ and every point $w \in W$, if $\mathfrak{M} \models$MGT$[w]$ then $(W, R, D, d) \models \mathbf{A}\forall x(\underset{\cdot}{Sent}_{ta}(x) \rightarrow \neg(Tx \wedge T\dot{\neg}x))[w]$. $\square$

Corollary 6. *MGT proves that only sentences are true.*

$$\mathrm{MGT} \vdash_{lq} \mathbf{A}\forall x(Tx \rightarrow \underset{\cdot}{\mathrm{Sent}}_{ta}(x))$$

Proof. We emulate the corresponding proof for KF+Cons [Halbach, 2011a, p. 212]. □

Theorem 4 is the first piece of evidence that going modal pays off for the theorist of grounded truth. In addition, it allows her to show that her truth predicate obeys the first principle of groundedness (see p. **??** above): for ϕ to be true, it must have been the case that ϕ earlier. This explains why I did not have to add it as an axiom, as I did with the second principle, in the form of (Ground).

Corollary 7. *For every $\mathcal{L}_{ta}$-sentence ϕ,*

$$\mathrm{MGT} \vdash_{lq} \mathbf{A}(T\ulcorner\phi\urcorner \rightarrow \mathbf{P}\phi)$$

Proof. By completeness, it suffices to show for every *lq* model $\mathfrak{M} = (W, R, D, d)$ and every point $w \in W$,

$$\mathfrak{I} \models \mathrm{MGT}[w] \Rightarrow \mathfrak{I} \models \mathbf{A}(T\ulcorner\phi\urcorner \rightarrow \mathbf{P}\phi)[w]$$

So let ϕ be any $\mathcal{L}_{ta}$-sentence, w some point in W and assume that $\mathfrak{I} \models \mathrm{MGT}[w]$. Further, let v be w or any point to the left or right of w (that is, let $v \in W$). In order to show that $\mathfrak{I} \models T\ulcorner\phi\urcorner \rightarrow \phi[v]$, we reason by induction on the positive complexity of ϕ. If ϕ atomic then the claim follows directly from TKF1,-2, -12 and TKF13. The interesting case is that of showing $\mathfrak{I} \models T\ulcorner\neg Ta\urcorner \rightarrow \mathbf{P}\neg Ta[v]$. So assume that $\mathfrak{I} \models T\ulcorner\neg Ta\urcorner[v]$. Then, since we assume TKF13 to hold at v, we know that for some uRv, $(W, R, D, d) \models T\underset{\cdot}{\neg} s[u]$. At this point, it is theorem 4 and the fact that $\mathfrak{I} \models \neg(Ta \wedge T\underset{\cdot}{\neg} s)[u]$ that allows us to proceed and conclude that $\neg Ta$ holds at u, hence $\mathbf{P}\neg Ta$ holds at v.

At the induction step, where we assume $\mathfrak{I} \models T\ulcorner\psi\urcorner \rightarrow \psi[v]$ to hold for every ψ of lower complexity than ϕ, the claim follows from a combination of TKF3-11 and the induction hypothesis. For example, assuming $\mathfrak{I} \models T\ulcorner\neg(\psi \wedge \zeta)\urcorner[v]$, we infer from $\mathfrak{I} \models (\mathrm{TKF5})[v]$ that $T\ulcorner\neg\psi\urcorner$ or $T\ulcorner\neg\zeta\urcorner$ holds at some uRv. Either way, however, our induction hypothesis and logic then licences the inference (at u) of $\neg(\psi \wedge \phi)$, as desired. □

Corollary 8. *For every $\mathcal{L}_{ta}$-sentence ϕ,*

$$\mathrm{MGT} \vdash_{lq} \mathbf{A}(T\ulcorner\phi\urcorner \rightarrow \phi)$$

Proof. For every ϕ, assuming $T\ulcorner\phi\urcorner$ we get $\mathbf{P}\phi$ from corollary 7. Then, we show ϕ on the basis of lemma 18 or lemma 17, for sentences containing 'T' or not, respectively. □

Corollary 9. *Let τ be a* truth-teller*, such that PA$\vdash \tau \leftrightarrow T\ulcorner\tau\urcorner$. Then*

$$\mathrm{MGT} \vdash_{lq} \mathbf{A}\neg T\ulcorner\tau\urcorner$$

Proof. By completeness of *lq*, it suffices to show that for every *lq*-model $\mathfrak{M} = (W, R, D, d)$ and every point $w \in W$, if $\mathfrak{M} \models \mathrm{MGT}[w]$ then $\mathfrak{M} \models \mathbf{A}\neg T\ulcorner\tau\urcorner[w]$.

So let $\mathfrak{M}$ be an *lq*-model, $w \in W$ and assume that if $\mathfrak{M} \models \mathrm{MGT}[w]$. For contradiction, assume that $\mathfrak{M} \models \mathbf{S}T\ulcorner\tau\urcorner[w]$. Then at w or at some point v to the left or right of w (we know R to be linear), $\mathfrak{M} \models T\ulcorner\tau\urcorner[v]$. Since $\mathfrak{M} \models \mathrm{TKF12}[w]$, and $v = w$ or to the left or right of w, $\mathfrak{M} \models \mathrm{F}T\underset{\cdot}{T}\ulcorner\tau\urcorner[v]$. Hence at some u to the right of v, $T\underset{\cdot}{T}\ulcorner\tau\urcorner$. Since $\mathfrak{M} \models \mathrm{TKF12}$ $[w]$ and we know u to be to the left or right of, if not identical to w, $\mathfrak{M} \models \mathbf{P}(T\ulcorner\tau\urcorner \wedge \mathbf{H}\neg T\ulcorner\tau\urcorner)[u]$. Consequently, at some t left of u, $T\ulcorner\tau\urcorner$ as well as $\mathbf{H}\neg T\ulcorner\tau\urcorner$. By corollary 7, for some further s left of t (hence somewhere to the left or right of, or identical with w), $\mathfrak{M} \models \tau[s]$. But because PA holds at s and by our assumption about this sentence τ, $\mathfrak{M} \models T\ulcorner\tau\urcorner[s]$. However, because $\mathbf{H}\neg T\ulcorner\tau\urcorner$ holds at t to the right of s, we also know that $\mathfrak{M} \models \neg T\ulcorner\tau\urcorner[s]$, contradiction. □

Corollary 9 indicates that we are on the right track towards a theory of grounded truth.

Question 1. How does MGT relate to Burgess' theory KFB, which is intended as an axiomatization of the least fixed point, and likewise proves truth to be consistent and a truth-teller not to be true?

One thing is clear, though. We want MGT to do better than KFB. As recently observed by Volker Halbach, KFB holds in other fixed points than the least one. Thus, KFB is not capable of singling out exactly the grounded truths.

The results of this section provide some evidence that MGT, unlike KF, is a theory of grounded truth. The main challenge, however, is to show that our theory can express sufficiently much of the original, semantic notion of groundedness. In the next section, I will make first steps into this direction.

11.5 MGT AND THE STAGES OF KRIPKE'S CONSTRUCTION

I now turn to the semantics of MGT. My goal in this section is to argue that MGT is a theory of *grounded* truth in a very robust sense. The main result of this section is theorem 5. It shows that MGT has a natural model: the stages of Kripke's construction (see §2.2). Moreover, MGT characterizes this particular model exactly.

How can this be? After all, the Kripke stages are well-ordered. MGT, however, is based on a logic of linearly ordered time. As we saw in section 11.3 (corollary 3), the logic of well-ordered time is not axiomatizable. Therefore, the theory MGT cannot distinguish between a Kripke-like but ill-founded sequence of models of truth, and our goal, the real order of Kripke stages.

As much as this is true, however, it is also irrelevant for the question whether MGT characterizes the Kripke construction. Let me give an analogy. KF is generally considered an adequate axiomatization of the Strong Kleene fixed points [McGee [1991]; Halbach [2011a]]. However, it is based on a merely first-order theory of arithmetic, whereas a Kripke fixed point is an expansion of the *standard* model. And of course, first-order Peano Arithmetic cannot single out the standard model.

Nonetheless KF is considered adequate. The reason is that *assuming arithmetic to be standard*, we can show that KF singles out the fixed points [Halbach, 2011a, theorem 15.15].

$$\mathbb{N}(X) \vDash \mathrm{KF} \Leftrightarrow X = \mathcal{J}_{sk}(X) \tag{34}$$

My goal is to show that MGT characterizes the stages of Kripke's construction just as well. *Assuming time to be well-ordered*, we can show that MGT characterizes the stages. More precisely, I will show that *well-ordered* models of MGT are isomorphic to the stages of Kripke's construction.

It may be thought that now I make too many assumption for the result to have much significance. Since, clearly, I still have to assume the number to be standard, as in the non-modal case. Thus, the theory characterizes groundedness only within the doubly narrow range of well-ordered, standard models.

However, things are not as they seem. Recall theorem 2 (p. 143). There is a set of principles N1-N3 in the language of tensed first-order arithmetic, such that any well-ordered model $\mathfrak{M}$ validates first-order arithmetic plus N1-N3 only if $\mathfrak{M}$ interprets the arithmetical vocabulary in the standard model $\mathbb{N}$. I will show that MGT proves, in linear time, such a set of principles (lemma 23). Hence, every well-ordered model of MGT respects the theory of standard numbers. Consequently, the analogy between the adequacy of KF of my result 5 is robust. In fact, just as for KF we *only* assume standardness, I *only* have to assume time to be well-ordered. Making this assumption, we will get the standard numbers for free.

To show that MGT proves principles that characterize the natural numbers, some stage-setting is needed. Firstly, observe that the syntactic relation "the formula ϕ is the result of applying x iterations of 'T' to ψ" is represented in PA by an $\mathcal{L}_{ta}$-formula that I will denote $T^{x}\ulcorner\phi\urcorner$, such that PA$\vdash \ulcorner\phi\urcorner = T^{x}\ulcorner\psi\urcorner$ iff, roughly, $\phi =$"$\underbrace{\dot{T}\dot{T}\ldots\dot{T}}_{x\text{-many}}\ulcorner\psi\urcorner$".

Using the $\dot{x}$ function from chapter 3 (p. 39) we enable the theory to quantify into the argument place x.

Lemma 21.

$$\mathsf{MGT} \vDash_{lq} \mathbf{A}\forall u\forall y\forall z\forall x(x = y + z \to TT^{\dot{y}}u \to TT^{\dot{x}}u)$$

Proof. We reason within MGT by (first-order) induction on z. The base case where $x = y$ is a truth of logic. The induction step follows from the first conjunct of axiom TKF12 and lemma 18, using the induction hypothesis. □

Secondly, let us define a special infinite sequence of $\mathcal{L}_{ta}$-formulae.

Definition 46. We define a sequence of formulae $\theta(x)$.

$$\theta(x) := TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \neg TT^{x+1}\ulcorner\overline{0} = \overline{0}\urcorner$$

For example, $\theta(\overline{0})$ is the sentence $T\ulcorner\overline{0} = \overline{0}\urcorner \wedge \neg T\underset{\cdot}{T}\ulcorner\overline{0} = \overline{0}\urcorner$.[4] The open formula $\theta(x)$ of the language of tensed truth $\mathcal{L}_{tam}$ will play the role of the predicate 'N' in terms of which we had formulated the principles N1-N3 of theorem 2. In order to show this, however, one further lemma is needed.

Lemma 22. *For every* lq*-model* $\mathfrak{M} = (W, R, D, d)$ *and every point* $w \in W$*, if* $\mathfrak{M} \vDash$*MGT*$[w]$ *then there is a point at which nothing is true but at every point accessible from it, something is true; more precisely, there is a* $v \in W$ *such that* $d(v)(\text{'}T\text{'}) = \varnothing$ *and for every* $u \in W$ *such that* vRu, $d(u)(\text{'}T\text{'}) \neq \varnothing$.

Furthermore, this point v *is the* least *point in the linear order* R*: there is no point* u *which sees* v.

Proof. Let $\mathfrak{M}$ be any linearly ordered constant-domain model, and w any point in it. Assume that $\mathfrak{M} \vDash$MGT$[w]$. Then $\mathfrak{M}$ validates the axiom of Ground: $\mathfrak{M} \vDash \mathbf{S}\forall x \neg Tx[w]$, i.e. $\mathfrak{M} \vDash \mathbf{P}\forall x \neg Tx \vee \forall x \neg Tx \vee \mathbf{F}\forall x \neg Tx[w]$. In each of these cases, there is some v such that $\mathfrak{M} \vDash \forall x \neg Tx[v]$.

Now, by first order logic and because $\mathfrak{M}$ validates MGT, in particular TKF1, $\mathfrak{M} \vDash \mathbf{A}(\overline{0} = \overline{0}) \wedge \mathbf{A}(\overline{0} = \overline{0} \rightarrow \mathbf{G}T\ulcorner\overline{0} = \overline{0})\urcorner[w]$.

By the linearity of R we know that wRv, vRw or $v = w$. In any case, $\mathfrak{M} \vDash \overline{0} = \overline{0} \wedge \overline{0} = \overline{0} \rightarrow \mathbf{G}T\ulcorner\overline{0} = \overline{0}\urcorner[v]$. Hence, $\mathfrak{M} \vDash \mathbf{G}T\ulcorner\overline{0} = \overline{0}\urcorner[v]$ and at every point u accessible from v, $\mathfrak{M} \vDash T\ulcorner\overline{0} = \overline{0}\urcorner[u]$, hence $d(u)(\text{'}T\text{'}) \neq \varnothing$.

It remains to show that there is no point u which sees v. For contradiction, assume that there is such a point u. Then, for the same reason as before, $\mathfrak{M} \vDash \mathbf{G}T\ulcorner\overline{0} = \overline{0}\urcorner[u]$. But since uRv, this means that $\mathfrak{M} \vDash T\ulcorner\overline{0} = \overline{0}\urcorner[v]$, contradiction. □

Finally, we are now in a position to show that MGT proves principles of the kind which we know to require a standard interpretation of the natural numbers.

Lemma 23. *Let* $\theta(x)$ *be the formula as defined in 46. We have that in the logic of linear time, MGT proves the following principles.*

θ.1 $\forall x \mathbf{S}\big(\theta(x) \wedge \mathbf{H}\neg\theta(x) \wedge \mathbf{G}\neg\theta(x)\big)$

4 Note that we cannot define $\theta(x)$ as $T^{x}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \neg T^{x+1}\ulcorner\overline{0} = \overline{0}\urcorner$ since $T^{x}\ulcorner\overline{0} = \overline{0}\urcorner$ is not a sentence, but a term.

θ.2 $\mathbf{A}\forall x \forall y\big((\theta(x) \wedge \theta(y)) \to x = y\big)$

θ.3 $\mathbf{A}\forall x \forall y\Big(x = y+1 \leftrightarrow \mathbf{S}\big(\theta(y) \wedge \mathbf{F}\theta(x)\big) \wedge \forall z\big(x \neq z \wedge \mathbf{S}(\theta(y) \wedge \mathbf{F}\theta(z)) \to \mathbf{S}(\theta(x) \wedge \mathbf{F}\theta(z))\big)\Big)$

Proof. (θ.1) In order not to assume standard numbers from the outside we reason within MGT, by induction on x. For this, we need to show that

$$\begin{aligned}\mathrm{MGT} \vdash_{lq} & \mathbf{S}\big(\theta(\overline{0}) \wedge \mathbf{H}\neg\theta(\overline{0}) \wedge \mathbf{G}\neg\theta(\overline{0})\big) \\ & \wedge \forall x\Big(\forall y\big(x < y \to \\ & \big(\mathbf{S}\big(\theta(\overline{y}) \wedge \mathbf{H}\neg\theta(\overline{y}) \wedge \mathbf{G}\neg\theta(\overline{y})\big) \to \mathbf{S}\big(\theta(\overline{x}) \wedge \mathbf{H}\neg\theta(\overline{x}) \wedge \mathbf{G}\neg\theta(\overline{x})\big)\big)\Big)\end{aligned}$$

For the first conjunct, by the completeness of LQ it suffices to show that for every *lq* model $\mathfrak{M}$, and every point w, if $\mathfrak{M} \models$MGT$[w]$ then there is some v to the left or right of w such that

$$\mathfrak{M} \models \mathsf{T}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \neg\mathsf{T}\underset{\cdot}{\mathsf{T}}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \mathbf{H}\neg\theta(\overline{0}) \wedge \mathbf{G}\neg\theta(\overline{0})[v]$$

So let w by any point in W and assume $\mathfrak{M} \models [w]$. From lemma 22, we know that there is an R-least point w_0 such that $\mathfrak{M} \models \neg\mathsf{T}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \overline{0} = \overline{0}[w_0]$. By the second conjunct and axiom TKF1, there is a point v, $w_0 R v$, such that $\mathfrak{M} \models \mathsf{T}\ulcorner\overline{0} = \overline{0}\urcorner[v]$. By the second conjunct of axiom TKF12, then, we know that there is a point u, vRu, such that $\mathfrak{M} \models \mathsf{T}\underset{\cdot}{\mathsf{T}}\ulcorner\overline{0} = \overline{0}\urcorner[u]$. Now we make use of the first conjunct of TKF12, and infer that there must be a point u', $u'Ru$, at which $\mathsf{T}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \mathbf{H}\neg\mathsf{T}\ulcorner\overline{0} = \overline{0}\urcorner$ is true. Let this be our witness. Firstly, by the first conjunct and axiom TKF12 we have that $\mathfrak{M} \models \mathbf{G}\mathsf{T}\underset{\cdot}{\mathsf{T}}\ulcorner\overline{0} = \overline{0}\urcorner[u']$, hence $\mathbf{G}\neg\theta(\overline{0})$. Secondly, by the second conjunct we know that at no point to the left of u' $\mathsf{T}\ulcorner\overline{0} = \overline{0}\urcorner$ will be true, hence $\mathfrak{M} \models \mathbf{H}\neg\theta(\overline{0})[u']$. Finally assume, for contradiction, that $\mathfrak{M} \not\models \neg\mathsf{T}\underset{\cdot}{\mathsf{T}}\ulcorner\overline{0} = \overline{0}\urcorner[u']$. Then $\mathsf{T}\underset{\cdot}{\mathsf{T}}\ulcorner\overline{0} = \overline{0}\urcorner$ must be true at this point, hence $\mathbf{P}\mathsf{T}\ulcorner\overline{0} = \overline{0}\urcorner$. But we already know that $\mathbf{H}\neg\mathsf{T}\ulcorner\overline{0} = \overline{0}\urcorner$ is true at u', contradiction. Therefore $\mathfrak{M} \models \theta(\overline{0})[u']$, as desired.

At the induction step, it again suffices to show that for every *lq* model $\mathfrak{M}$ and point w, for every $x \in D$, if

$$\mathfrak{M} \models \mathsf{MGT} \wedge \forall y\big(x < y \to \big(\mathbf{S}\big(\theta(y) \wedge \mathbf{H}\neg\theta(y) \wedge \mathbf{G}\neg\theta(y)\big)\big)[w]$$

then

$$\mathfrak{M} \models \mathbf{S}\big(\theta(x) \wedge \mathbf{H}\neg\theta(x) \wedge \mathbf{G}\neg\theta(x)\big)[w]$$

So let $y = x - 1$, such that for some point v to the left of right of w, $\mathsf{T}\mathsf{T}^{y}\ulcorner\overline{0} = \overline{0}\urcorner$ is true at v. Twice making use of the second conjunct of axiom TKF12, we have that for some u seen by v, $\mathfrak{M} \models \mathsf{T}\underset{\cdot}{\mathsf{T}}\mathsf{T}^{y+1}\ulcorner\overline{0} = \overline{0}\urcorner[u]$. Then, by the axiom's first conjunct we know that at some u', $u'Ru$, it is true that $\mathsf{T}\mathsf{T}^{y+1}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \mathbf{H}\neg\mathsf{T}\mathsf{T}^{y+1}\ulcorner\overline{0} = \overline{0}\urcorner$. Recall that

$x = y + 1$. Analogously to before, we therefore let this u' be our witness.

On the one hand, by the second conjunct we know that $\mathbf{H}\neg TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner$, hence $\mathbf{H}\neg\theta(x)$. From the first conjunct and axiom TKF12 we know that $\mathfrak{M} \models \mathbf{G}T\underset{\cdot}{T}T^{x}\ulcorner\overline{0} = \overline{0}\urcorner[u']$, hence $\mathbf{G}\neg\theta(x)$ is true at u', too.

On the other hand, assume for contradiction that $\mathfrak{M} \models TT^{x+1}\ulcorner\overline{0} = \overline{0}\urcorner[u']$. Then at u', it must be true that $\mathbf{P}TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \mathbf{H}\neg TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner$, contradiction.

$(\theta.2)$ follows from lemma 21 since, if by contraposition and without loss of generality we assume that $x < y$ then $\theta(y)$ entails $TT^{x+1}\ulcorner\overline{0} = \overline{0}\urcorner$, hence $\neg\theta(x)$.

$(\theta.3.)$ By completeness it suffices to show that for every *lq* model $\mathfrak{M} = (W, R, D, d)$ and every point $w \in W$, if $\mathfrak{M} \models$MGT$[w]$ then for every $v \in W$ and every $x, y \in D$, $x = y + 1$ is true at v iff

i. for some u, $\mathfrak{M} \models \theta(y) \wedge \mathbf{F}\theta(x)[u]$

ii. $\mathfrak{M} \models \forall z\big(x \neq z \wedge \mathbf{S}(\theta(y) \wedge \mathbf{F}\theta(z)) \to \mathbf{S}(\theta(x) \wedge \mathbf{F}\theta(z))\big)[v]$

For the left-to-right direction, assume $\mathfrak{M} \models x = y + 1[v]$. For (i) note that by $(\theta.1)$, there are u and u' such that

$$\mathfrak{M} \models TT^{y}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \neg TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner[u] \tag{35}$$

$$\mathfrak{M} \models TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \neg TT^{x+1}\ulcorner\overline{0} = \overline{0}\urcorner[u'] \tag{36}$$

To show that uRu', by the linearity of R it suffices to note that u and u' cannot be identical, and that since by lemma 18, $\mathfrak{M} \models \mathbf{G}TT^{x}[u']$ assuming $u'Ru$ leads equally to contradiction.

For (ii), let $z \neq x$ and assume that for some u, u', uRu', (35) and

$$\mathfrak{M} \models TT^{z}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \neg TT^{z+1}\ulcorner\overline{0} = \overline{0}\urcorner[u'] \tag{37}$$

By $(\theta.1)$ we know that there is a u'' such that

$$\begin{aligned} \mathfrak{M} \models\, & TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \neg TT^{x+1}\ulcorner\overline{0} = \overline{0}\urcorner \\ & \wedge \mathbf{H}\neg(TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \neg TT^{x+1}\ulcorner\overline{0} = \overline{0}\urcorner) \\ & \wedge \mathbf{G}\neg(TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner \wedge \neg TT^{x+1}\ulcorner\overline{0} = \overline{0}\urcorner)[u''] \end{aligned} \tag{38}$$

By lemma 21 we can infer from this that uRu''. It remains to show that $u''Ru'$. On the one hand, we note that if $u' = u''$ and $z > x$ then, since $TT^{z}\ulcorner\overline{0} = \overline{0}\urcorner$ is true at u'', lemma 21 implies that $TT^{x+1}\ulcorner\overline{0} = \overline{0}\urcorner$ must also be true there, thus contradicting (38). For $z < x$ the dual argument leads to a contradiction. On the other hand, it likewise cannot be the case that $u'Ru''$ since if $z > x$ then lemma 21 requires $TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner$ to be true at u'. By (38), however, $\mathfrak{M} \models \neg TT^{x}\ulcorner\overline{0} = \overline{0}\urcorner[u']$, contradiction. Again, by the dual argument we also rule out the case for $z < x$.

For the left-to-right direction, we firstly note that by lemma 18, for every x, y, $\theta(x) \wedge \mathbf{F}\theta(y)$ is true at u only if $\neg TT^{y}$ is true at u. Now, assume (i) and (ii) and, for contradiction, that $x \neq y+1$ is true at v. Let us write $x\Theta y$ iff $\mathfrak{M} \models \mathbf{S}\big(\theta(x) \wedge \mathbf{F}\theta(y)\big)[v]$. From (i) we get that $y\Theta x$. I claim that $x\Theta y+1$.

Now, assume that $x\Theta y+1$ is witnessed by u and $x > y+1$. By our first observation above we then have that at u, it is true that $\neg \mathsf{T}\mathsf{T}^{y+1}\ulcorner\overline{0} = \overline{0}\urcorner$. However, since $\mathsf{T}\mathsf{T}^{x}\ulcorner\overline{0} = \overline{0}\urcorner$ is true at u and $x > y+1$, lemma 21 requires that $\mathfrak{M} \models \mathsf{T}\mathsf{T}^{y+1}[u]$ after all, contradiction.

If $x < y+1$ then by our observation above and the assumption that $y\Theta x$ $\mathfrak{M} \models \neg\mathsf{T}\mathsf{T}^{x}\ulcorner\overline{0} = \overline{0}\urcorner[u]$, contradicting, once more, lemma 21 according to which at u, $\mathsf{T}\mathsf{T}^{y}\ulcorner\overline{0} = \overline{0}\urcorner \wedge (\mathsf{T}\mathsf{T}^{y}\ulcorner\overline{0} = \overline{0}\urcorner \rightarrow \mathsf{T}\mathsf{T}^{x}\ulcorner\overline{0} = \overline{0}\urcorner)$.

It remains to show my claim that $x\Theta y+1$, i.e. that at some point it is true that $\mathfrak{M} \models \theta(y) \wedge \mathbf{F}\theta(y+1)$. By $\theta.1$ we know that there is a u such that $\mathfrak{M} \models \mathsf{T}\mathsf{T}^{y}\ulcorner\overline{0} = \overline{0}\urcorner[u]$. Making twice use of the second conjunct of axiom TKF12, we know that there is a u', such that $\mathsf{T}\dot{\mathsf{T}}\mathsf{T}^{y+1}\ulcorner\overline{0} = \overline{0}\urcorner$ is true at u'. By the axiom's first conjunct we then know that $\mathfrak{M} \models \mathsf{T}\mathsf{T}^{y+1} \wedge \mathbf{H}\neg\mathsf{T}\mathsf{T}^{y+1}[u'']$. By the reasoning as used in the proof of $\theta.1$ we show that in fact, $\theta(y+1)$ is true at this u'', which thus witnesses the truth of $\mathbf{F}\theta(y+1)$ at u, such that we can conclude that $x\Theta y+1$. □

Lemma 24. *For every* woq-*model* $\mathfrak{M} = (W, R, D, d)$, *if*

$$\forall w \in W\,(W, R, D, d) \models \mathrm{MGT}[w]$$

then $\mathfrak{M}$ *interprets the arithmetical vocabulary standard*

Proof. From theorem 2 and the previous lemma, which shows that MGT provides us with precisely such a set of principles that characterizes, over a well-ordered frame, the natural numbers. □

Recall Kripke's construction of an $\mathcal{L}_{ta}$ model, based on Strong Kleene logic.

Definition 47 (Kripke's Construction)**.** Let $\models_{SK}$ be a Strong Kleene satisfaction relation as defined in [Halbach, 2011b, 15.10]. Consider the following operator on pairs of sets of sentence-codes. Let (X^+, X^-) be any such pair.

$$\mathcal{J}_{sk}(X^+, X^-) := \mathcal{J}^+_{sk}(X^+, X^-), \mathcal{J}^-_{sk}(X^+, X^-)$$

where

$$\mathcal{J}^+_{sk}(X^+, X^-) := \{\ulcorner\phi\urcorner : \mathbb{N}(X^+, X^-) \models_{SK} \phi\}$$
$$\mathcal{J}^-_{sk}(X^+, X^-) := \{\ulcorner\phi\urcorner : \mathbb{N}(X^+, X^-) \models_{SK} \neg\phi\}$$

This operator $\mathcal{J}_{sk}$ induces a sequence $(I^{\alpha}_{SK})_\alpha = (I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK})_\alpha$.

$$\begin{aligned}
(I^{+,0}_{SK}, I^{-,0}_{SK}) &:= (\varnothing, \varnothing)\\
(I^{+,\alpha+1}_{SK}, I^{-,\alpha+1}_{SK}) &:= \mathcal{J}_{sk}(I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK})\\
(I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK})_\alpha &:= \bigcup_{\beta<\alpha} (I^{+,\beta}_{SK}, I^{-,\beta}_{SK})_\beta, \text{ for } \alpha \text{ limit.}
\end{aligned}$$

... and a least fixed point I_{SK}, just as monotone operators do. In particular, this least fixed point is reached at the first non-recursive ordinal ω_1^{CK}, which is limit.

For the purpose of axiomatizing Kripke's theory of truth, it is common to work with the *closed off* fixed point model $\mathbb{N}(I^+_{SK})$. However, we may also consider closing off each stage of Kripke's construction. Thus, we arrive at a well-ordering of (classical) models $\mathbb{N}(I^{+,\alpha}_{SK})$, which gives naturally rise to a model for the modal logic *lq*.

Definition 48 (KS – *Kripke's Stages*). Let ω_1^{CK} be the constructive ordinals. Let d_k map each constructive ordinal $< \omega_1^{CK}$ to a model of the language $\mathcal{L}_{tam}$ such that firstly, at every point, the arithmetical vocabulary is interpreted in the standard way on the set of natural numbers ω, and secondly, $d_k(\alpha)(\text{'T'}) = \mathbb{N}(I^{+,\alpha}_{SK})$.

Let KS be the model $(\omega_1^{CK}, <, \omega, d_k)$.

Recall that I write $S \simeq S'$ if the structure S is isomorphic to S'.

Theorem 5. *For every* woq-*model* $\mathfrak{M} \models (W, R, D, d)$*,*

$$\forall w \in W\, \mathfrak{M} \models \mathrm{MGT}[w] \textit{ if and only if } (W, R, D, d) \simeq \mathrm{KS}$$

Proof. ($\Rightarrow$) By lemma 24 we know that $\mathfrak{M}$ interprets the arithmetical vocabulary standard, such that we can, for simplicity, identify every point of W with a model $\mathbb{N}(X)$.

Having noted this, we reason by induction on the well-ordering R. We show that the R-least model is $\mathbb{N}(\varnothing) = \mathbb{N}(I^{+,0})_{SK}$. Then, assuming that some point w is the stage $\mathbb{N}(I^{+,\alpha}_{SK})$, we show that the R-next point v is the successor stage $\mathbb{N}(I^{+,\alpha+1}_{SK})$. Finally, we show that the R-limit of an initial segment of the points, which we know are the models $\mathfrak{N}\left(I^{+,\gamma}_{SK}\right)$ for $\gamma < \beta$, is the union model $\mathbb{N}(\bigcup\limits_{\gamma<\beta} I^{+,\gamma}_{SK})$.

So, let w_0 be the R-least model $\mathbb{N}(X)$ in W. Since $\mathfrak{M} \models$(Ground)$[w_0]$, w_0 or some point R-earlier than w_0 must be a model $\mathbb{N}(\varnothing)$. But there is no such point – after all, w_0 is the R-least point. Therefore, it must be that $w_0 = \mathbb{N}(\varnothing)$.

Now, assume that $w = \mathbb{N}(I^{+,\alpha}_{SK})$. We need to show that w's R-successor v is $\mathbb{N}(I^{+,\alpha+1}_{SK})$. We know that $v = \mathbb{N}(X)$. It remains to show, therefore, that $X = \{\ulcorner\phi\urcorner : \mathbb{N}(I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK}) \models_{SK} \phi\}$.

($\subseteq$) Let $\ulcorner\phi\urcorner \in X$, we want to show that $\mathbb{N}(I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK}) \models_{SK} \phi$. We know that $v \models T\ulcorner\phi\urcorner$ and reason by induction on the *positive complexity* of ϕ. If $\phi =$'$a = b$' then $v \models \mathbf{P}a = b$, since we assume axiom TKF1 to hold at v. Hence, at some point uRv, $a = b$. Note that since $a = b$ does not involve a partial predicate, if $a = b$ holds in some classical model $\mathbb{N}(X^+)$ then it also holds in the partial model $\mathbb{N}(X^+, X^-)$. Therefore, if $u = \mathbb{N}(I^{+,\alpha}_{SK})$ we are already done. If not, we make use of lemma 17 and conclude that $a = b$ must hold at every point; in particular, therefore, $\mathbb{N}(I^{+,\alpha}_{SK}) \models a = b$. Hence $\mathbb{N}(I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK}) \models_{SK} a = b$, as desired.

For $\phi =$'$a \neq b$' we reason just analogously, using TKF2 instead of TKF1.

If $\phi =$'Tb' such that $\mathfrak{M} \models T\ulcorner Tb \urcorner[v]$ then we know by the second conjunct of TKF12 that $\mathfrak{M} \models \mathbf{P}Tb[v]$. Hence, for some uRv, $u \models Tb$. If $u = w = \mathbb{N}(I^{+,\alpha}_{SK})$ then we are done. If uRw, then lemma 2 allows us to infer that $w \models Tb$ after all, and in fact $\mathbb{N}(I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK}) \models_{SK} Tb$, as desired.

If $\phi =$'$\neg Tb$' then we use the second conjunct of TKF13. This time, we get that $\mathfrak{M} \models \mathbf{P}T\underset{\cdot}{\neg} b \vee \neg \underset{\cdot}{Sent}_{ta}(s)[v]$. Assume that b denotes a sentence code $\ulcorner\psi\urcorner$ in v. Then we know, as before, that $\mathbb{N}(I^{+,\alpha}_{SK}) \models_{SK} T\ulcorner\neg\psi\urcorner$. If b does not denote a sentence code, then we know that its denotation is in the anti-extension $I^{-,\alpha}_{SK}$, indeed has been so from the first point onwards, and we conclude that $\mathbb{N}(I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK}) \models_{SK} \neg Tb$, as desired.

Now, consider $\phi =$'$\psi \wedge \zeta$' such that $v = \mathbb{N}(X) \models T\ulcorner\psi \wedge \zeta\urcorner$. Since $\mathfrak{M} \models$TKF4$[v]$, we know that $v \models T\ulcorner\psi\urcorner \wedge T\ulcorner\zeta\urcorner$. Hence, $\ulcorner\psi\urcorner, \ulcorner\zeta\urcorner \in X$. By our induction hypothesis, we know that $\mathbb{N}(I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK}) \models_{SK} \psi \wedge \zeta$, as desired. For $\phi =$'$\neg(\psi \wedge \zeta)$' we proceed analogously, exploiting the fact that TKF5 holds in the model.

Disjunction and the quantifiers are taken care of analogously. Recall that $\rightarrow$ is defined in terms of $\neg$ and $\vee$.

($\supseteq$) Let $\mathbb{N}(I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK}) \models_{SK} \phi$, we want to show that $\ulcorner\phi\urcorner \in X$, that is, $\mathfrak{M} \models T\ulcorner\phi\urcorner[v]$. Since partial truth in a model is contained by classical truth in it, we have that $\mathbb{N}(I^{+,\alpha}_{SK}) \models \phi$. We reason by induction on the positive complexity of ϕ. If $\phi =$'$a = b$' then $\mathbb{N}(I^{+,\alpha}_{SK}) \models \mathbf{G}T\ulcorner a = b\urcorner$, since we assume TKF1 to hold at $v = \mathbb{N}(I^{+,\alpha}_{SK})$. We assume v to be the R-successor of $w = \mathbb{N}(I^{+,\alpha}_{SK})$, hence $\mathfrak{M} \models T\ulcorner a = b\urcorner[w]$, as desired. Analogously for negated equations and sentences of the form Tb or $\neg Tb$.

If $\phi =$'$\psi \wedge \zeta$' we know from our induction hypothesis that $\mathbb{N}(I^{+,\alpha+1}_{SK}) \models T\ulcorner\psi\urcorner \wedge T\ulcorner\zeta\urcorner$. From the fact that $\mathbb{N}(I^{+,\alpha+1}_{SK}) \models$TKF4 we further know that $\mathbb{N}(I^{+,\alpha+1}_{SK}) \models T\ulcorner\psi \wedge \zeta\urcorner$, as desired. Analogously for negated conjunctions and the other connectives and quantifiers. Consequently, $X = \{\ulcorner\phi\urcorner : \mathbb{N}(I^{+,\alpha}_{SK}, I^{-,\alpha}_{SK}) \models_{SK} \phi\}$, and $v = \mathbb{N}(I^{+,\alpha+1}_{SK})$, as desired.

Finally, assume that $w_0, \ldots, v$ are the points $\mathfrak{N}\left(I^{+,\alpha}_{SK}\right)$ for $\alpha < \beta$. We show that the R-limit of the points is the model $\mathbb{N}(\bigcup_{\gamma<\beta} I^{+,\gamma}_{SK})$. We know that $v = \mathbb{N}(X)$ and show that $X = \bigcup_{\gamma<\beta} I^{+,\gamma}_{SK}$. The reasoning is similar to before and I confine myself to an outline. We reason by induction on the positive complexity of ϕ. For ($\subseteq$), we observe that for atomic sentences, the axioms TKF1, -2, -12 and -13 ensure that if $\ulcorner\phi\urcorner \in X$ then it is true at some preceding point, hence in their union $\bigcup_{\gamma<\beta} I^{+,\gamma}_{SK}$. The same can be inferred for complex sentences, from the induction hypothesis and the compositionality axioms TKF3-11.

($\Leftarrow$) Let w be any $\mathbb{N}(I_{SK}^{+,\alpha})$. Of course, KS$\models$ PA[w]. The truth of the axiom (Ground) is witnessed by the point $\mathbb{N}(\varnothing)$. To see that the modal axioms TKF1, -2 and -12 are sound, firstly note that generally, $\mathbb{N}(I_{SK}^{+,\alpha}) \models T\ulcorner\phi\urcorner$ only if for some $\beta < \alpha$, $\mathbb{N}(I_{SK}^{+,\beta}) \models \phi$. This validates each of these axiom's second conjunct.

I now turn to the first conjuncts. Let α be any stage. If atomic sentences $a = b$ and Ta hold in the classical model $\mathbb{N}(I_{SK}^{+,\alpha})$, then so they do in the partial model $\mathbb{N}(I_{SK}^{+,\alpha}, I_{SK}^{-,\alpha})$. Hence $\mathbb{N}(I_{SK}^{+,\alpha+1}) \models T\ulcorner a = b\urcorner \wedge T\ulcorner Tc\urcorner$. Further, since the sequence $(I_{SK}^{+,\alpha})_\alpha$ increases, $T\ulcorner a = b\urcorner \wedge T\ulcorner Tc\urcorner$ will hold at every later stage $\mathbb{N}(I_{SK}^{+,\beta})$, $\beta > \alpha$. This, however, is what the first conjuncts of TKF1 and TKF12 say, which are thereby shown to hold at every point in the model KS. TKF2 is taken care of analogously, noting that for arithmetical sentence of the form $a \neq b$, classical and partial truth at a stage coincide.

TKF13 requires a subtler treatment, since it concerns *negated* truth, whose behaviour on classical models differs from that on partial ones. Note, however, that the complication is not due to my modal setting, but pertains to the fact that we axiomatize Kripkean truth classically, and hence applies already to KF itself. As a result, the considerations which show the soundness of KF in any Strong Kleene fixed point can guide our investigation into the soundness of MGT. Generally, in order to see the soundness of TKF13, recall that $\neg\phi$ is in some extension $I_{SK}^{+,\alpha}$ just in case ϕ is in the corresponding *anti*-extension $I_{SK}^{-,\alpha}$, such that $T\ulcorner\neg\phi\urcorner$ again is in the successor extension $I_{SK}^{+,\alpha+1}$; and that for every term a not standing for a sentence (code), $\neg Ta$ is in the extension of 'T' right from the beginning.

In order to show that the compositionality axioms TKF3-11 hold in the model $\mathbb{N}(I_{SK}^{+,\alpha})$ it suffices to note that the truths in a Strong Kleene model are closed under double-negation, conjunction, disjunction and quantification; and further, that a negated conjunction (disjunction) is classically true just in case both of (one of) the negated conjuncts (disjuncts) are.[5]

This completes the proof of theorem 5. $\square$

Recall that I_{SK}^{+} is the extension of the Strong Kleene fixed point – the set of *grounded* truths. Recall further that we write $\Sigma \models_{woq} \phi$ if ϕ is a consequence of Σ over the well-ordered frames.

Corollary 10. *For every $\mathcal{L}_{ta}$-sentence ϕ,*

$$\ulcorner\phi\urcorner \in I_{SK}^{+} \textit{ iff } \mathrm{MGT} \models_{woq} \mathbf{S}\, T\ulcorner\phi\urcorner$$

Proof. We show that $\ulcorner\phi\urcorner \in I_{SK}^{+}$ just in case: for every *woq*-model $\mathfrak{M} = (W, R, D, d)$, and for every point $w \in W$ if $\mathfrak{M} \models \mathrm{MGT}[w]$ then $\mathfrak{M} \models \mathbf{S}\, T\ulcorner\phi\urcorner[w]$.

5 To see this, Halbach's lemma 15.6 and surrounding remarks are instructive [Halbach, 2011b, p. 205].

($\Leftarrow$) Assume that for every *woq*-model $\mathfrak{M} = (W, R, D, d)$ and every point $w \in W$, $\mathfrak{M} \models \mathbf{S}\,\mathrm{T}\ulcorner\phi\urcorner[w]$ if $\mathfrak{M} \models \mathrm{MGT}[w]$. We know, from the right-to-left direction of theorem 5, that the sequence KS of stages of Kripke's construction models MGT: for every $\alpha < \omega_1^{CK}$, KS$\models$ $\mathrm{MGT}[\alpha]$. By our assumption, therefore, for some point β, KS$\models$ $\mathbf{S}\,\mathrm{T}\ulcorner\phi\urcorner[\beta]$. Hence, $\phi \in d_k(\beta) = \mathbb{N}(I_{SK}^{+,\beta})$, for some β. Consequently, $\ulcorner\phi\urcorner \in I_{SK}^{+}$, as desired.

($\Rightarrow$) Now assume $\ulcorner\phi\urcorner \in I_{SK}^{+}$, such that $\ulcorner\phi\urcorner \in I_{SK}^{+,\alpha}$, for some α. From the left-to-right direction of 5 we know that for every *woq*-model (W, R, D, d) MGT is true at every point in W only if the model is KS. Hence, trivially, in every model that models MGT there is a point, namely $\mathbb{N}(I_{SK}^{+,\alpha})$, such that $\mathrm{T}\ulcorner\phi\urcorner$ holds at this point. Hence, for every *woq*-model and every point, if MGT holds there, so does $\mathbf{S}\,\mathrm{T}\ulcorner\phi\urcorner$, and we conclude that MGT$\models_{woq}$ $\mathbf{S}\,\mathrm{T}\ulcorner\phi\urcorner$, as desired. $\square$

11.6 DISCUSSION

Objection: *You do not allow the tense operators to occur within the scope of 'T'. Your modal logic is a meta-theory in disguise. Therefore, you have failed to respond to the ghost challenge.* Response: We need to distinguish between two projects. My present goal is to allow an existent, extensional theory of truth grounded in arithmetic to express groundedness. Another goal is to develop a grounded theory of truth in the modalized language $\mathcal{L}_{tam}$. This I did not attempt to do.

Objection: *The logic $\models_{woq}$ of well-ordered time is not axiomatizable. Hence, clearly, your key result 10 is not available to us in our own language. Therefore, the ghost challenge still stands.*

Again, we must distinguish two projects. On the one hand, one may attempt to give a formal system by which to *compute* whether or not a given sentence is grounded.

However, in view of the fact that the set of grounded sentences is not computable, in fact Π_1^1-complete, this is a hopeless project. I certainly did not attempt to do this.

On the other hand, we may confine ourselves to providing a means to express groundedness without semantic ascent. ...

11.7 CONCLUSION

The project of motivating a theory of type-free truth from the notion of groundedness faces the challenge that groundedness is a meta-theoretic notion. I offered a response to this challenge. We can express the idea of groundedness in our own language using intensional means, more precisely tense.

I presented one way of implementing this response and formulated a theory of truth based on the logic of well-ordered time. This axiomatic theory relates naturally to Kripke's semantic construction. I take this to be evidence for my proposal.

Part I

APPENDIX

BIBLIOGRAPHY

Peter Aczel. An introduction to inductive definitions. In Jon Barwise, editor, *Handbook of mathematical logic*, pages 739–782. Amsterdam, 1977. (Cited on page 4.)

Peter Aczel. Frege structures and the notions of proposition, truth and set. In Jon Barwise, Howard Jerome Keisler, and Kenneth Kunen, editors, *The Kleene Symposium*, pages 31–59. North-Holland, 1980. (Cited on page 36.)

Paul Audi. A Clarification and Defense of the Notion of Grounding. manuscript presented at Because II conference, Berlin August 2010, 2010. (Cited on pages 112 and 115.)

R. Batchelor. Grounds and consequences. *Grazer Philosophische Studien*, 80(1):65–77, 2010.

Hanoch Ben-Yami. Plural Quantification Logic: A Critical Appraisal. *The Review of Symbolic Logic*, 2(1), March 2009. (Cited on page 2.)

Stephen Blamey. Partial Logic. In *Handbook of Philosophical Logic*, volume 5, pages 262 – 353. D.Reidel, second edition edition, 2002. (Cited on page 18.)

George Boolos. The Iterative Conception of a Set. *The Journal of Philosophy*, 68:215–232, 1971. (Cited on pages 80 and 92.)

Ross T. Brady. The consistency of the axioms of abstraction and extensionality in a three-valued logic. *Notre Dame Journal of Formal Logic*, 12(4):447–453, 1971. (Cited on pages 38 and 56.)

John P. Burgess. The Truth is Never Simple. *The Journal of Symbolic Logic*, 51(3):663 – 681, 1986. (Cited on page 27.)

John P. Burgess. E Pluribus Unum: Plural Logic and Set Theory. *Philosophia Mathematica*, 12:193 – 221, 2004. (Cited on page 2.)

John P. Burgess. Friedman and the Axiomatization of Kripke' Theory of Truth. Unpublished manuscript, delivered at a conference in honour of the 60th birthday of Harvey Friedman at the Ohio State University, 14-17 May 2009, 2009. (Cited on pages 42 and 43.)

Andrea Cantini. A Theory of Truth Formally Equivalent to id1. *Journal of Symbolic Logic*, 55:244 – 258, 1990. (Cited on page 27.)

Andrea Cantini. *Logical Frameworks for Truth and Abstraction: An Axiomatic Study*, volume 135 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, 1996. (Cited on pages 36 and 44.)

Georg Cantor. Foundations of a general theory of manifolds: a mathematico-philosophical investigation into the theory of the infinite. In William Ewald, editor, *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, volume 2, pages 878 – 919. Clarendon Press, Oxford, 1883. (Cited on page 88.)

Georg Cantor. *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Springer, 1932. edited by Ernst Zermelo. (Cited on page 11.)

J.H. Conway. *On Numbers and Games*. Academic Press, Natick, MA, 2nd edition, 2001. (Cited on page 82.)

Fabrice Correia. *Existential dependence and cognate notions*. Philosophia, Munich, 2005. (Cited on page 112.)

Fabrice Correia. Grounding and Truth-functions. *Logique et Analyse*, 116, 2011.

W.B. Ewald. *From Kant to Hilbert: Readings in the Foundations of Mathematics*. Clarendon Press, 1996. ISBN 0198532717. (Cited on page 11.)

Solomon Feferman. Non-extensional type-free theories of partial operations and classifications, i. In *ISILC Proof Theory Symposion*, pages 73–118. Springer, 1975a. (Cited on page 36.)

Solomon Feferman. A language and axioms for explicit mathematics. In J. N. Crossley, editor, *Algebra and logic*, volume 450 of *Lecture Notes in Mathematics*, pages 87–139. Springer, 1975b. (Cited on pages 36, 37, and 41.)

Solomon Feferman. Constructive theories of functions and classes. In *Logic Colloqium '78*, volume 97 of *Studies in Logic and the Foundations of Mathematics*, pages 159–224. Elsevier, 1979. (Cited on page 37.)

Solomon Feferman. Reflecting on Incompleteness. *The Journal of Symbolic Logic*, 56:1–49, 1991. (Cited on page 41.)

Solomon Feferman. Axioms for determinateness and truth. *Review of Symbolic Logic*, 1(1):204–217, 2008. (Cited on page 25.)

Jose Ferreiros. *Labyrinth of Thought: A History of Set Theory and Its Role in Modern Mathematics*. Birkhäuser Verlag, Basel - Boston - Berlin, 2007 edition, 1999. (Cited on page 13.)

Kit Fine. The question of realism. *Philosophers' Imprint*, 1(1):1–30, 2001.

Kit Fine. Class and Membership. *The Journal of Philosophy*, 102(11), November 2005. (Cited on pages 64, 68, 74, and 75.)

Kit Fine. Some Puzzles of Ground. *Notre Dame Journal of Formal Logic*, 51(1):97–118, 2010. (Cited on pages 132 and 133.)

Kit Fine. The pure logic of ground. *The Review of Symbolic Logic*, 5(1): 1–25, 2012a. (Cited on page 115.)

Kit Fine. Guide to ground. In Fabrice Correia and Benjamin Schnieder, editors, *Metaphysical grounding: Understanding the structure of reality*, pages 37–80. Cambridge University Press, Cambridge, New York, Melbourne, 2012b. (Cited on page 120.)

Thomas Forster. The iterative conception of set. *Review of Symbolic Logic*, 1(1):97–110, 2008. (Cited on pages 2, 78, 79, 80, 81, 82, and 83.)

A.A. Fraenkel, Y. Bar-Hillel, and A. Levy. *Foundations of set theory*, volume 67. North Holland, 1973. (Cited on page 37.)

Harvey Friedman and Michael Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21, 1987. (Cited on pages 85 and 87.)

Kentaro Fujimoto. Relative truth definability of axiomatic truth theories. *Bulletin of Symbolic Logic*, 16(3):305–344, 2010. (Cited on page 25.)

Kentaro Fujimoto. Classes and truths in set theory. *Annals of Pure and Applied Logic*, 163:1484–1523, 2012. (Cited on pages 44 and 50.)

J. W. Garson. Quantification in modal logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, 1st edition, volume2: Extensions of Classical Logic*. Reidel, Dordrecht, 1984. (Cited on page 143.)

P. C. Gilmore. The consistency of partial set theory without extensionality. In T. Jech, editor, *Axiomatic Set Theory II*, Proceedings of Symposia in Pure Mathematics. American Mathematical Society, Providence, Rhode Island, 1974. (Cited on page 45.)

Kurt Gödel. What is cantor's continuum problem? In Solomon Feferman, John W. Dawson Jr., Warren Goldfarb, Charles Parsons, and Robert M. Solovay, editors, *Collected Works II*, pages 154 – 187. Oxford University Press, Oxford, 1947. (Cited on page 13.)

Volker Halbach. *Axiomatische Wahrheitstheorien*. Akademie Verlag, 1996. (Cited on page 148.)

Volker Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, Cambridge, 2011a. (Cited on pages 137, 152, and 154.)

Volker Halbach. *Axiomatic Theories of Truth*. Cambridge University Press, Cambridge, 2011b. (Cited on pages 22, 24, 38, 42, 43, 53, 87, 146, 150, 158, and 161.)

A. P. Hazen. Relations in lewis's framework without atoms. *Analysis*, 57(4):243–8211, 1997. (Cited on page 4.)

Allen Hazen. Relations in lewis's framework without atoms: A correction. *Analysis*, 60(4):351–8211, 2000. (Cited on page 4.)

Richard G. Heck. The Consistency of Predicative Fragments of Frege's Grundgesetze der Arithmetik. *History and Philosophy of Logic*, 17: 209–220, 1996. (Cited on page 52.)

Hans G. Herzberger. Paradoxes of Grounding in Semantics. *Journal of Philosophy*, 67:145 – 167, 1970. (Cited on pages 2 and 16.)

R Hinnion. Le paradoxe de russell dans des versions positives de la théorie naïve des ensembles. *CR Acad. Sci. Paris, Sér. I Math*, 304: 307–310, 1987. (Cited on page 45.)

Roland Hinnion and Thierry Libert. Positive abstraction and extensionality. *Journal of Symbolic Logic*, 68(3):828–836, 2003. (Cited on page 45.)

Luca Incurvati. How to be a minimalist about sets. *Philosophical Studies*, pages 1–19, 2012. (Cited on page 92.)

Ingasi Jané. Idealist and Realist Elements in Cantor's Approach to Set Theory. *Philosophia Mathematica*, 18:193 – 226, 2010. (Cited on page 11.)

Gerhard Jäger, Reinhard Kahle, and Thomas Studer. Universes in explicit mathematics. *Annals of Pure and Applied Logic*, 109:141–162, 2001. (Cited on page 37.)

Jeffrey Ketland. Identity and Indiscernibility. *The Review of Symbolic Logic*, 4(02):171–185, 2011. (Cited on page 56.)

Saul Kripke. Outline of a Theory of Truth. *The Journal of Philosophy*, 72(19):690 – 716, November 1975. Seventy-Second Annual Meeting Americal Philosophical Association. (Cited on pages 2, 16, 17, 18, and 39.)

Hannes Leitgeb. What Truth Depends On. *Journal of Philosophical Logic*, 35:155–192, 2005. (Cited on pages 2, 31, and 32.)

Hannes Leitgeb. Towards a logic of type-free modality and truth. In C. Dimitracopoulos, L. Newelski, D. Normann, and J. Steel, editors, *Logic Colloquium '05*, Lecture Notes in Logic, 28. cup, 2006. (Cited on page 34.)

David Lewis. *Parts of Classes*. Basil Blackwell, 1991. (Cited on page 4.)

Penelope Maddy. Proper Classes. *Journal of Symbolic Logic*, 48(1):113–139, Mar 1983. (Cited on pages 36, 38, 49, and 50.)

Penelope Maddy. Believing the axioms i. *The Journal of Symbolic Logic*, 53(2):481–511, June 1988. (Cited on page 37.)

Penelope Maddy. A theory of sets and classes. In Gila Sher and Richard Tieszen, editors, *Between Logic and Intuition: Essays in Honor of Charles Parsons*, pages 299–316. Cambridge University Press, 2000. (Cited on pages 36, 38, and 49.)

Tim Maudlin. *Truth and Paradox: Solving the Riddles*. Clarendon Press, Oxford, 2004. (Cited on pages 2 and 16.)

Timothy McCarthy. Ungroundedness in Classical Languages. *Journal of Philosophical Logic*, 17(1):61–74, 1988. (Cited on page 2.)

Vann McGee. *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Hackett, Indianapolis, Cambridge, 1991. (Cited on pages 16 and 154.)

Toby Meadows. Truth, dependence and supervaluation: Living with the ghost. *Journal of Philosophical Logic*, pages –, forthcoming.

Urs Oswald. *Fragmente von N̈ew Foundationsünd Typentheorie*. PhD thesis, ETH Zürich, 1976. (Cited on page 78.)

Charles Parsons. Sets and classes. *Noûs*, 8(1):pp. 1–12, 1974. (Cited on page 37.)

Charles Parsons. What is the Iterative Conception of Set? In Robert E. Butts and Jaakko Hintikka, editors, *Logic, Foundations of Mathematics and Computability Theory*, volume 9 of *University of Western Ontario Series in Philosophy of Science*, pages 335–367. Reidel, Dordrecht, 1977. (Cited on pages 36, 92, and 93.)

Charles Parsons. *Sets and Modality*, chapter 11. Cornell University Press, Ithaca, New York, 1983. (Cited on pages 93, 94, and 95.)

Michael Potter. *Set Theory and its Philosophy: A Critical Introduction*. Oxford University Press, Oxford, 2004. (Cited on pages 81, 92, and 93.)

Graham Priest. *An Introduction to Non-Classical Logic: From* If *to* Is. Cambridge University Press, Cambridge, second edition, 2008. (Cited on page 145.)

Agustin Rayo and Gabriel Uzquiano. *Absolute Generality*. Clarendon Press, Oxford, 2006. (Cited on page 81.)

Gideon Rosen. Metaphysical Dependence: Grounding and Reduction. In Bob Hale and Aviv Hoffman, editors, *Modality: Metaphysics, Logic and Epistemology*. Oxford University Press, Oxford, 2010. (Cited on pages 107, 110, and 112.)

Stewart Shapiro. *Foundations without Foundationalism: a case for second-order logic*. Oxford Logic Guides. Clarendon Press, 1991. (Cited on page 5.)

Joseph R. Shoenfield. Axioms of set theory. In Jon Barwise, editor, *Handbook of Mathematical Logic*, volume 90 of *of Studies in Logic and the Foundations of Mathematics*. North-Holland, Amsterdam, 1977. (Cited on page 92.)

Gabriel Uzquiano. Plural Quantification and Classes. *Philosophia Mathematica*, 11:67–81, 2003. (Cited on page 37.)

Floris T. van Vugt and Denis Bonnay. What makes a sentence be about the world? Towards a unified account of groundedness. manuscript June 11, 2011, 2011.

Albert Visser. Semantics and the Liar Paradox. In Dov Gabbay and Franz Günther, editors, *Handbook of Philosophical Logic*. Kluwer, 2nd edition, 2004. (Cited on pages 16 and 22.)

Hao Wang. Large sets. In Robert E. Butts and Jaakko Hintikka, editors, *Logic, Foundations of Mathematics and Computability Theory*, volume 9 of *University of Western Ontario Series in Philosophy of Science*, pages 309 – 335. Reidel, Dordrecht, 1977. (Cited on page 92.)

Hermann Weyl. *Philosophy of Mathematics and Natural Science*. Princeton University Press, Princeton, 1949. (Cited on page 93.)

Steve Yablo. Grounding, Dependence and Paradox. *Journal of Philosophical Logic*, 11(1):117–137, 1982. (Cited on pages 2, 5, and 19.)

Øystein Linnebo. Sets, Properties and Unrestricted Quantification. In Agustin Rayo and Gabriel Uzquiano, editors, *Absolute Generality*, pages 149 – 178. Clarendon Press, Oxford, 2006. (Cited on page 37.)