Linnebo, O., & Horsten, L. F. M. (2016). Term Models for Abstraction Principles. *Journal of Philosophical Logic*, *45*(1), 1-23. https://doi.org/10.1007/s10992-015-9344-z

Term models for abstraction principles

Leon Horsten (University of Bristol)

Øystein Linnebo (University of Oslo)

Draft of November 26, 2014

Abstract

Kripke's notion of groundedness plays a central role in many responses to the semantic paradoxes. Can the notion of groundedness be brought to bear on the paradoxes that arise in connection with abstraction principles? We explore a version of grounded abstraction whereby term models are built up in a 'grounded' manner. The results are mixed. Our method solves a problem concerning circularity and yields a 'grounded' model for the predicative theory based on Frege's Basic Law V. However, the method is poorly behaved unless the background second-order logic is predicative.

1 Introduction

There has recently been a lot of interest in *abstraction principles*, which are principles of the form:

$$(\Sigma_0) \qquad \S\alpha = \S\beta \leftrightarrow \Phi(\alpha, \beta),$$

where the variables 'α' and 'β' range over items of a certain sort, '§' is an operator taking items of this sort to objects, and '$\Phi(\alpha, \beta)$' expresses an equivalence relation on items of this sort.[1] When the variables are of second order, we will follow Frege and speak of them as ranging over *concepts*. Two famous examples are Hume's Principle and Frege's Basic Law V:

$$(\mathrm{HP}_0) \qquad \#F = \#G \leftrightarrow F \approx G$$

$$(\mathrm{V}_0) \qquad \varepsilon F = \varepsilon G \leftrightarrow \forall x(Fx \leftrightarrow Gx)$$

(The subscript '0' will be explained shortly.) The right-hand side of Hume's Principle abbreviates the claim that the concepts F and G are equinumerous; that is, that there is a relation

[1] See for instance (Hale and Wright, 2001), (Fine, 2002), and (Burgess, 2005).

that one-to-one correlates the Fs and the Gs. This principle plays a key role in Fregean approaches to arithmetic. For Hume's Principle is consistent and suffices, when combined with second-order logic and some natural definitions, to derive all of second-order Dedekind-Peano Arithmetic. But Basic Law V, which is concerned with extensions and the equivalence relation of coextensionality, is inconsistent against the background of any second-order logic that contains at least Π^1_1-comprehension (Burgess, 2005, p. 47).

This raises the question of what forms of abstraction are acceptable. We would like to draw a well motivated line between acceptable and unacceptable forms of abstraction. This is often known as *the bad company problem*.[2] We wish to explore a response to this problem that has received less attention than it deserves.[3] Kripke's notion of groundedness is rightly celebrated as a response to the semantic paradoxes. Can an analogous notion of groundedness be articulated and used to provide an account of acceptable abstraction?

The purpose of this article is to explore a particular explication of the idea of grounded abstraction that naturally comes to mind when taking one's inspiration from Kripke. As is well known, Kripke starts with the set of (Gödel numbers of) sentences of the language of arithmetic supplemented with a truth predicate. He then builds up larger and larger extensions and anti-extensions of the truth-predicate in a grounded manner. Transposed to the case of abstraction, this suggests that we start with a set T of closed *abstraction terms*, that is, singular terms of the form $\S x.\phi(x)$ with no free variables. We then endeavor to build up an equivalence relation R on T in a 'grounded' manner such that the R-equivalence classes of T yield a model for the relevant abstraction principle and some fragment of second-order logic. Since any closed abstraction term is to denote its own R-equivalence class in the resulting model, this model can thus be regarded as a *term model* for the relevant abstraction principle.

The results of our investigation are mixed. The method that we develop addresses an important concern about circularity and yields a 'grounded' model for the predicative theory based on Frege's Basic Law V (in what we will shortly call its *axiomatic* version). Although this theory is already known to be consistent,[4] our construction establishes the stronger claim that the theory has a *natural* model—in a sense we make precise. Other results are limitative. Our explication of the idea of groundedness turns out to be poorly behaved unless

[2]See (Linnebo, 2009b) for an overview.

[3]However, there has recently been some enthusiasm for the approach: see (Horsten and Leitgeb, 2009), (Leitgeb, ming), and (Linnebo, 2009a).

[4]See (Heck, Jr., 1996).

the background second-order logic is predicative. This limitation is not noted in earlier studies of this approach to grounded abstraction, such as (Horsten and Leitgeb, 2009). This severely limits the strength of the abstractionist theories that can be justified in this way. In particular, anyone seeking an account of grounded abstraction that is strong enough to justify a substantial amount of classical mathematics will need an alternative explication of the idea of groundedness.[5]

A side benefit of our investigation has to do with Frege himself. As several commentators have observed, Frege appears to rely on something like term models in some important arguments.[6] The primary example is *Grundgesetze* I, §§29-31, where Frege attempts to prove that every well-formed expression of his language has a unique denotation, using ideas related to his famous 'context principle'. Since he took sentences to denote truth-values, this would ensure that his system in consistent. So by Russell's paradox, we know that the proof must be flawed. Can any aspects of Frege's argumentative strategy be salvaged? We shed light on this question by determining the potential for the term model approach. In particular, we prove that (Dummett, 1991)'s very negative assessment of its potential is only partially right.

Since we wish to consider term models, it will be convenient to consider abstraction principles of a *schematic* rather than an *axiomatic* form. The axiomatic form is given by (Σ_0) above. The schematic form is as follows:

$$(\Sigma) \qquad \S x.\phi(x) = \S x.\psi(x) \leftrightarrow \Phi[\phi/F, \psi/G],$$

where § is a variable-binding operator taking formulas to singular terms, and where the right-hand side is the formula that results from $\Phi(F, G)$ when any occurrences of Ft and Gt, for any first-order term t, are (simultaneously) replaced by corresponding occurrences of the formulas $\phi(t)$ and $\psi(t)$ respectively. For instance, the schematic form of Basic Law V is:

$$(\mathrm{V}) \qquad \varepsilon x.\phi(x) = \varepsilon x.\psi(x) \leftrightarrow \forall x(\phi(x) \leftrightarrow \psi(x)).$$

As a notational convention, we will add a subscript '0' to indicate that an abstraction principle is axiomatic rather than schematic. Clearly, a schematic principle (Σ) is at least as strong as the corresponding axiomatic one (Σ_0), as is seen by letting $\phi(x)$ and $\psi(x)$ be Fx and Gx

[5] See e.g. (Linnebo, 2013) and (Studd, 2014) for two attempts to do this.

[6] See e.g. (Dummett, 1991, esp. ch. 17), (Heck, Jr., 1997), (Heck, Jr., 2012, ch.s 3 and 5), and (Linnebo, 2004).

respectively. Conversely, (Σ_0) allows us to reproduce all instances of (Σ) whose formulas $\phi(x)$ and $\psi(x)$ are allowed to figure in the second-order comprehension axioms. However, when full impredicative comprehension isn't assumed, the axiomatic principle (Σ_0) will be weaker than the schematic one (Σ).[7]

The article is organized as follows. We begin by clarifying the idea of term models for abstraction principles (Section 2). The construction of such term models is threatened by two different forms of impredicativity (Section 3). Our method of groundedness nicely overcomes one of the threats (Section 4). We illustrate the value of this discovery by constructing a natural model for the axiomatic version of Basic Law V (but not the schematic version) and predicative second-order logic (Section 5). Unfortunately, the other threat from impredicativity remains stubborn and means that our method works only when the background second-order logic is predicative (Section 6).

2 Term models for abstraction principles

We now examine how term models for abstraction principles can be constructed. As mentioned, the idea is to identify an appropriate set T of abstraction terms and build up, in a 'grounded' manner, an equivalence relation R on T such that the set of R-equivalence classes of members of T gives rise to a term model of the relevant abstraction principle. The task of this section is to explain the general idea of such term models; in particular, to identity some constraints on our choice of first- and second-order domains.

2.1 How to work *modulo* an equivalence

The idea of working *modulo* an equivalence relation R can be implemented in two different ways. One option is to let the first-order domain be the set of R-equivalence classes of T and interpret the identity predicate '$=$' as real identity on this set. On this implementation, each abstraction term in T will denote its own R-equivalence class. Another option is to let the first-order domain be the term set T itself and instead interpret the identity predicate '$=$' by means of R. On this option, each abstraction term in T will denote itself, while an identity based on terms from T will be deemed true iff these terms are R-equivalent. Although the two implementations are mathematically equivalent, many of our constructions are smoother

[7] The situation is analogous to that of the two theories ACA and ACA_0 of second-order arithmetic, from which our notation is inspired.

in the context of the second implementation, which we therefore adopt. (Of course, any of the resulting structures will have an associated quotient structure.) Henceforth, our term set will therefore be referred to as 'D_1', which frees up the letter 'T' for other uses.

In order to ensure that Leibniz's Law remains valid, we need to ensure that R satisfies the following closure condition:

($\star$) Let t, t', and T be closed abstraction terms, and let $T' = T[t'/t]$ be the result of substituting t' for each occurrence of t in T. Then, if $\langle t, t' \rangle \in R$, then $\langle T, T' \rangle \in R$.

For now, we simply assume as given an equivalence relation R satisfying ($\star$) and on this basis examine how to construct a structure for the relevant language. In Section 4, we will examine how such an R can be built up such that the resulting structure is a model for the desired abstraction principle.

2.2 The need to consider open abstraction terms

There is little point in letting the term set D_1 consist of anything other than closed abstraction terms, that is, abstraction terms with no free variables. For in the term models that we will examine, each term in D_1 is to denote itself. But we do not want an open abstraction term to denote itself: an open term should only be assigned a denotation relative to an assignment to its free variables.

However, if our construction is to result in a model for an abstraction principle, it will nevertheless have to involve open abstraction terms, even though these are not present in D_1. The reason is that our language permits quantification into abstraction terms. Indeed, such quantification plays a key role in standard uses of abstraction principles. Consider for instance the Fregean definition of the notion of a number:

(Def-N) $$N(x) \leftrightarrow \exists F(x = \#u.Fu)$$

This makes crucial use of second-order quantification into the open abstraction term $\#u.Fu$. Such quantification is also required in the definition of the relation P of immediate predecession, which plays a crucial role in so-called *Frege arithmetic* (that is, arithmetic based on Hume's Principle and second-order logic):

(Def-P) $$P(x, y) \leftrightarrow \exists F \exists w[Fw \wedge x = \#u(Fu \wedge u \neq w) \wedge y = \#u.Fu]$$

2.3 The interpretation of open abstraction terms

As explained, each closed abstraction term from D_1 is to denote itself. How are we to handle the open abstraction terms, which, as we have seen, also need to be considered? As a warm-up, consider the case where the only free variables are of first order. Assume, for instance, that the term t is $\S u.\phi(u, v_1, \ldots, v_m)$, with all free variables displayed. Let σ be a variable assignment that assigns to each variable v_i a closed term t_i. Relative to σ, the term t can be taken to denote the term $\S u.\phi(u, t_1, \ldots, t_n)$. So we need to ensure that this term is an element of D_1.

What about free variables of *second* order? We would like these too to be be eliminable once an assignment σ has been specified. We will now see that this requirement constrains our choice of a second-order domain. To keep things simple, we consider only the case where the second-order parameters are monadic. The general case is analogous.

A natural option is to let every element X of the second-order domain D_2 be represented by an object language formula ψ with a unique free variable of first order, with the understanding that X is to apply to $t \in D_1$ just in case ψ is satisfied when t is assigned to its unique free variable. This approach makes available a natural interpretation of the abstraction operator § relative to any variable assignment σ. To see this, consider an open abstraction term t of the form $\S u.\phi(u, v_1, \ldots, v_m, F_1, \ldots, F_n)$. The assignment σ will ascribe to each second-order variable F_i an object language formula ψ_i with one free variable. Relative to this assignment we can rewrite t by replacing each predication of the form $F_i s$, where s is any first-order term, with $\psi_i(s)$. When combined with the treatment of first-order parameters explained above, this will allow every open abstraction term t to be rewritten, relative to an assignment σ, as a closed abstraction term t'. Relative to σ, we can therefore let t denote the closed term t'. Of course, we need to ensure that t' is an element of D_1 and will return to this at the end of the next subsection.

An alternative option, one might have thought, is to let the second-order domain D_2 be the powerset of D_1. But on this approach, how are we to evaluate an open abstraction term t relative to an assignment? Consider the general case where t has free variables of both first and second order. Assume that t is of the form $\S u.\phi(u, v_1, \ldots, v_m, F_1, \ldots, F_n)$, with all free variables displayed. Then each variable F_i is assigned a set S_i from D_2. Now there is no systematic way to reduce the term t, relative to this assignment, to a term in D_1.[8]

[8]Note that the argument is independent of the particular choice of D_1. For any set D_1 of abstraction

Accordingly, on this approach we are unable to eliminate second-order parameters.[9]

We conclude that, in order to be able to interpret open abstraction terms, the second-order domain D_2 must consist of object language formulas with an appropriate number of free variables.[10]

2.4 The first- and second-order domains

A further constraint on D_2 is that we must be able to define the satisfaction of an atomic predication Vt—where V is a second-order variable and t is any first-order term—relative to the assignment of any formula from D_2 to V. The crafting of such a definition turns out to be rather delicate. Two different options will be explored in Sections 3.2 and 6. The former option requires a restriction to predicative second-order logic; the latter promises to avoid this restriction but is unsuccessful. For present purposes, it suffices simply to assume that a set of open formulas have been provided as our second-order domain D_2, and that a definition has been provided of what it is for an atomic predication of the mentioned sort to be satisfied by a member of D_2.

We also need to decide which abstraction terms are to populate the first-order domain D_1. If we are studying a schematic abstraction principle, then D_1 will consist of all the closed abstraction terms in the relevant language. Clearly, this choice of D_1 will allow the reductions involved in our interpretation of an open abstraction term relative to an assignment to go through. If instead we are studying an axiomatic abstraction principle, then it is appropriate to let D_1 consist of all closed terms of the form $\S u.\phi(u)$, where ϕ is in D_2. For the kinds of choices of D_2 that we will consider—namely, all predicative formulas or all formulas whatsoever—it is easy to verify that D_1 again allows the reductions involved in the

terms, the second-order parameters are chosen from the set $\wp(T)$, which means there will be no natural way of eliminating such parameters from a given abstraction term t to yield a term in D_1.

[9] We are here ignoring two options which strike us as desperate and unattractive. One is to exclude from our language all abstraction terms with free second-order variables. This would enable us to let the term set D_1 be the set of abstraction terms thus restricted, and let D_2 be the powerset of D_1. However, this would come at the cost of sacrificing some of the core applications of abstraction, which, as we observed in Section 2.2, require second-order quantification into the scope of abstraction operators.

A second desperate option is to operate with *two* (monadic) second-order domains: a narrower one, corresponding to the first of the above options, and a wider one, corresponding to the second option. Assume the language contains two sets of second-order variables. Then it is possible to let one set of second-order variables range only over the narrower domain and admit quantification with respect to such variables into the scope of abstraction operators, while not admitting this for the second set of second-order variables, which range over the wider domain. See (Burgess, 2005, p. 119) for discussion of a closely related idea. However, this option strikes us as contrived and not much better than first constructing a term model and then adding a layer of second-order quantification 'by hand'.

[10] This limitation is not noted in earlier studies of grounded abstraction, such as (Horsten and Leitgeb, 2009).

interpretation of open abstraction terms to go through.

2.5 Putting everything together

Let's now put all the components together to define a Tarskian notion of satisfaction in the term model based on a set D_1 of abstraction terms and an equivalence relation R that satisfies the closure condition ($\star$).

- The first-order domain is the set of abstraction terms D_1.
- The second-order domain D_2 is a set of first-order formulas with an appropriate number of free (first-order) variables.
- We have explained how an abstraction term t is to be interpreted relative to a variable assignment σ by showing how t, relative to σ, can be reduced to a term in D_1.
- The identity predicate is interpreted by means of R.
- We have assumed as given a definition of satisfaction of atomic predications of the form $Vt_1, \ldots, t_n$ relative to a variable assignment σ.
- The truth-functional connectives and the first- and second-order quantifiers are interpreted in the standard Tarskian way.

We write $\mathcal{M}[R]$ for the resulting structure and $\mathcal{M}[R] \models_\sigma \phi$ for the notion of satisfaction of a formula ϕ relative to a variable assignment σ. As soon as the underspecified components have been fully described, we will verify that $\mathcal{M}[R]$ is indeed a model for a fragment of second-order logic. Notice that our notation leaves the term set D_1 implicit. The reason is that D_1 will mostly be fixed, whereas the equivalence relation R will vary.

2.6 Term models relative to an arbitrary base model

The model construction outlined above can also be carried out relative to an arbitrary base model. Consider a base model $\mathcal{M}_0$ of some first-order base language $\mathcal{L}_0$. Assume we wish to abstract on $\mathcal{M}_0$ with respect to an abstraction principle (Σ). Let $\mathcal{L}$ be the language that results from $\mathcal{L}_0$ by adding the abstraction operator § and, if (Σ) is a second-order principle, also second-order resources.

In this case, it makes sense to allow the abstraction terms to contain arbitrary parameters from the domain M_0 of $\mathcal{M}_0$. We do this by letting the first-order domain D_1 be the result of adding to M_0 all ordered pairs of the form $\langle \S u.\phi(u, \bar{v}), \bar{b}\rangle$, where the first coordinate is an abstraction term proper and the second coordinate is a string of parameters from M_0 to be assigned to the string $\bar{v}$ of non-designated free variables that occur in the abstraction term. Our goal is now to build up an equivalence relation R on D_1 (which had better not identify any of the elements of M_0). As before, we let the second-order domain D_2 consist of first-order formulas of $\mathcal{L}$ with a designated free variable; however, this time we allow the formula to contain additional parameters with values in M_0, to be coded by means of ordered pairs as above. It is easy to verify that the other definitions listed in Section 2.5 can be adapted to this modified setting.

3 Concerns about impredicativity

According to (Dummett, 1991, ch. 17), impredicativity is 'the serpent' that entered Frege's paradise. It is this serpent that undermined Frege's attempt to construct a term model for Basic Law V, which in turn was meant to ensure the reference of its abstraction terms via an invocation of the 'context principle'. In this section we show that impredicativity is indeed the source of the main threats to the construction of term models for abstraction principles. However, the picture we defend is both more nuanced than Dummett's and somewhat less bleak.

3.1 Two kinds of impredicativity

We begin by distinguishing between two different kinds of impredicativity, each of which will be shown to pose a threat to our desired construction. One kind pertains to the background second-order logic. Consider the second-order comprehension scheme:

(Comp) $$\exists R \forall x_1 \ldots \forall x_n [Rx_1 \ldots x_n \leftrightarrow \phi(x_1, \ldots, x_n)],$$

where ϕ may contain free variables of first and second order that are not displayed (but not, of course, R). As usual, an instance of this scheme is said to be *predicative* provided that the formula ϕ contains no bound second-order variables; otherwise, the instance is said to be *impredicative.*

Another kind of impredicativity pertains to the abstraction principles themselves. Say that an abstraction principle is *impredicative* if the singular terms on its left-hand side purport to denote objects that are included in the range of some quantifier occurring on its right-hand side; otherwise say that it is *predicative.* For instance, (HP) and (V) are impredicative because their right-hand sides quantify over all objects, including the ones referred to on the left-hand side. By contrast, the abstraction principle for directions is predicative:

(Dir) $$d(l_1) = d(l_2) \leftrightarrow l_1 \parallel l_2$$

For the variables l_1 and l_2 range only over lines, not over directions.

3.2 The problem of impredicative comprehension

The question of how to define the satisfaction of an atomic predication of the form $Vt_1, \ldots, t_n$ in a term model relative to an assignment σ came up in Section 2.4 but was deferred. We now develop one answer. (Another answer will be discussed in Section 6 and found wanting.) We focus on the monadic case. The polyadic case is analogous. So assume σ assigns to the monadic second-order variable V the formula $\psi(x)$.

By far the most natural approach is to reduce the question of whether

(1) $$\mathcal{M}[R] \models_\sigma Vt$$

to what we hope will be a simpler question of whether

(2) $$\mathcal{M}[R] \models_\sigma \psi(t).$$

However, if ψ contains bound second-order variables, then the latter question will be just as complex as the former, which means that the envisaged satisfaction clause for predications of the form Vt is not guaranteed to be well-defined.

If, on the other hand, we restrict the second-order domain to formulas that are predicative (in the usual sense of containing no bound second-order variables), then the envisaged reduction will succeed. To see this, let the second-order domain D_2 consist of all and only *first-order* formulas ψ of the object language with a unique free variable of first order. Then every question of the form (1) reduces to a question of the form (2), without the latter ever leading back to the former.

This ensures that all permissible questions receive answers. But are the answers the right ones? Assume t and t' are R-equivalent. We need to ensure that the two associated questions of the form (1) receive the same answer. The following proposition shows that this—and more—is ensured by our closure condition ($\star$) from p. 5.

Proposition 1 Let D_2 consist of all first-order formulas with one or more free variables, D_1 be a set of closed abstraction terms that is closed under the two reduction procedures involved in the interpretation of open terms, and R be an equivalence relation on D_1 that respects ($\star$). Then $\mathcal{M}[R]$ is a model of predicative second-order logic.

Proof. The only part of the claim that is not immediate from the definitions summarized in Section 2.5 is that $\mathcal{M}[R]$ satisfies Leibniz's Law and predicative second-order comprehension. For the former, let t and t' be R-equivalent members of D_1 and ϕ any formula. We need to show that for any assignment σ, we have:

$$\mathcal{M}[R] \models_\sigma \phi \leftrightarrow \phi[t'/t] \tag{3}$$

We prove this claim by induction on the syntactic complexity of ϕ. Assume first that ϕ is atomic. There are just two cases to consider. One is that ϕ is an identity. If t flanks the identity predicate, then (3) is immediate from the fact that R is an equivalence relation. If instead t occurs as a subterm of a complex term that flanks the identity predicate, then we additionally need to invoke ($*$). The other case is that ϕ is a predication of the form $V s_1, \ldots, s_n$, where V is a second-order variable. To handle this case, it suffices to show that (3) holds for any *first-order* formula ϕ. We show this by an inner induction on the syntactic complexity of ϕ. The only base cases are now the two kinds of identity contexts that we considered above. And the induction steps involve preservation of (3) under negation, disjunction and quantification, all of which are straightforward. This completes the inner induction. Returning to the outer induction, all that remains is to prove the induction steps. But these are the same as in the inner induction. This completes the outer induction as well.

Next, let ϕ be a predicative formula, possibly with parameters of first and second order. Relative to the variable assignment σ, these parameters can be replaced by abstraction terms or formulas, as described in Section 2.3. Relative to σ, the formula ϕ can thus be rewritten as a formula ψ in D_2, which establishes that comprehension on ϕ relative to σ is permitted. $\dashv$

By contrast, the impredicative second-order comprehension axioms are not guaranteed to hold. This restriction to theories with no more than predicative comprehension is a serious limitation. For instance, (HP) plus predicative second-order logic yields only Robinson arithmetic Q, rather than full second-order Peano-Dedekind arithmetic. And although (V) plus predicative second-order logic has the virtue of being consistent, it is known to be very weak.[11] This raises the question of whether a less restrictive approach to the second-order domain is possible. We will return to this question in Section 6, where we will defend a negative answer. Until then, we will rely on the natural approach outlined above and accept the predicativity restrictions to which this gives rise.

3.3 Are impredicative abstraction principles viciously circular?

We now turn to the equivalence relation R on the term set D_1, which has so far simply been assumed. The idea is that R should hold between two closed abstraction terms just in case they are to be regarded as co-denoting. Does an abstraction principle enable us to build up a suitable equivalence relation R?

Attempts to do so face a major obstacle, which is articulated by Dummett in a discussion of a closely related attempt of Frege's in *Grundgesetze* I, §§29-31.[12] The abstraction principle in question is (V).

> Frege [...] proceeds to lay down the condition for the truth of a statement of identity between value-ranges under the guise of fixing the reference of the abstraction operator. That will depend upon the truth of a universally quantified statement [...]. The truth-value of that statement will in turn depend upon the application of some complex predicate to every element of the domain, and hence, in effect, upon the truth-value of every result of inserting a value-range term in its argument-place. Since these statements are likely to involve further identity-statements between value-range terms of unbounded complexity, Frege's stipulations are not well founded: the truth-value of an identity-statement cannot be construed as depending only on the references of less complex terms or on the truth-values of less complex sentences. (Dummett, 1991, pp. 221-222)

A bit of explication won't hurt. The truth-value of an identity statement $\varepsilon x.\phi(x) = \varepsilon x.\psi(x)$

[11](Ganea, 2007) proves the theory to be exactly as strong as Q.

[12]A similar worry is expressed by (Heck, Jr., 1997, pp. 460–61) and (Fine, 2002, p. 88).

is supposed to be fixed by the truth-value of the quantified statement $\forall x(\phi(x) \leftrightarrow \psi(x))$. But the truth-value of the latter statement will depend on the application of the open formula (or 'complex predicate') $\phi(x) \leftrightarrow \psi(x)$ to every value-range term in the language, including ones of complexity greater than the value-range terms flanking the original identity statement.[13] Indeed, since the formulas $\phi(x)$ and $\psi(x)$ may contain the identity predicate, we may be led to identity statements as complex as, or even identical with, the one with which we started.[14] Frege's attempt to assign truth-values is therefore not well-founded.

What went wrong? It is important to realize that the problem is entirely independent of the impredicativity of the background second-order logic. The complications just canvassed arise *irrespective of whether impredicative second-order comprehension is allowed.* Rather, the problem is that the truth-condition of a quantified formula $\forall x\, \phi(x)$ obviously depends on the range of the quantifier. But in a term model, this range consists of equivalence classes of abstraction terms that are deemed to co-refer and will thus depend on the truth-conditions of identity statements between two abstraction terms. Yet the converse dependency seems to obtain as well. For the truth-conditions of such identity statements are to be determined via (Σ) and will thus depend on the truth-conditions of various quantified formulas. So there seems to be a circular relation of dependency: the truth-conditions of quantified formulas depend on those of identity statements, which in turn depend on the truth-conditions of quantified formulas. Let's call this *the circularity worry.*

In fact, the converse dependency just mentioned arises only for impredicative abstraction principles. It is only when the right-hand side of an abstraction principle quantifies over the objects referred to on its left-hand side that the truth-conditions for identity statements will depend on the truth-conditions for formulas that quantify over these very objects. (We are here assuming that the right-hand side of the abstraction principle does not contain the abstraction operator §. This assumption will henceforth be left implicit.) Unfortunately, the fact that predicative abstraction principles avoid the circularity worry provides little solace. For it is straightforward to construct models for predicative abstraction principles.[15] It is only for the mathematically and foundationally more interesting class of impredicative

[13] For instance, the 'complex predicate' has to be applied to the terms $\varepsilon u(u = \varepsilon x.\phi(x))$ and $\varepsilon u(u = \varepsilon x.\psi(x))$.

[14] For instance, this situation will arise if $\phi(x)$ is of the form $\exists u \exists v(u = v \wedge \theta(u, v, x))$.

[15] Consider for instance the predicative version of Basic Law V formulated in a two-sorted language with one sort for ordinary objects and another sort reserved for extensions, where the abstraction terms on the left-hand side belong to the latter sort, while the right-hand side belongs entirely to the former. To construct a model, start with any domain D of ordinary objects, let the second-order quantifiers range over the powerset of D, let this powerset also be the domain of extensions, and finally interpret the operator ε as the identity function.

second-order abstraction principles that it is hard to construct models.

In short, when an abstraction principle is predicative, the search for term models avoids the circularity worry but is not needed. And when an abstraction principle is impredicative, this search promises to be of real value but cannot avoid the circularity worry. Our only hope is thus to face the worry head on and show how the circular relation of dependency need not be vicious. Thankfully, we will now see how a method of groundedness enables us to overcome the circularity observed by Dummett and others. This is a partial vindication of the strategy employed by Frege in *Grundgesetze* I, §§29-31, despite the undeniable problems that Frege's argument encounters. In particular, our method delivers, in a perfectly grounded way, a term model of the sort sought in (Dummett, 1991, p. 221), thus disproving Dummett's claim that Frege's strategy is fundamentally flawed.

4 How groundedness overcomes the circularity worry

Previous investigations show that there are indeed cases where the circularity worry can be overcome. One example concerns Davidson's criterion of identity for events, which says that two events are identical just in case they share the same causal relations to any third event. The impredicativity of this criterion gives rise to a similar circularity worry, which (Horsten, 2010) shows not necessarily to be vicious. We also draw inspiration from (Horsten and Leitgeb, 2009).

Our strategy is to build up approximations to the desired equivalence relation in stages. Throughout, we will assume that D_1 is a set of closed abstraction terms that is closed under the two reduction procedures involved in the interpretation of open terms.

Definition 1 (Approximations) Let an *approximation* be an ordered pair $E = \langle E^+, E^- \rangle$ of sets of abstraction terms from D_1 such that:

(i) E^+ is an equivalence relation

(ii) E^+ is disjoint from E^-

One approximation E *is extended by* another F (in symbols: $E \sqsubseteq F$) iff $E^+ \subseteq F^+$ and $E^- \subseteq F^-$.

Intuitively, E^+ represents the pairs of abstraction terms that have been determined as equivalent, and E^-, the set of pairs of terms that have been determined as inequivalent.

Recall the closure condition that we need to impose on an equivalence relation R in order to ensure that $\mathcal{M}[R]$ respects Leibniz's Law:

($\star$) Let t, t', and T be closed abstraction terms, and let $T' = T[t'/t]$ be the result of substituting t' for each occurrence of t in T. Then, if $\langle t, t' \rangle \in R$, then $\langle T, T' \rangle \in R$.

Definition 2 (Admissible extensions) An dyadic relation R on D_1 is an *admissible extension* of an approximation E (in symbols:[16] $E \sqsubseteq R$) iff the following conditions are met.

(i) R is an equivalence relation

(ii) R respects ($\star$)

(iii) R respects all the positive and negative information encoded in E, in the sense that $E^+ \subseteq R$ and $E^- \cap R = \varnothing$

We next define an operation D that takes us from one approximation E to a better one. The definition uses our Tarskian notion of satisfaction from Section 2.5. Because the abstraction terms involved are closed, we can suppress the assignment σ.

Definition 3 Let $D^+(E)$ be the set of pairs $\langle \S x.\phi_1, \S x.\phi_2 \rangle$ of elements of D_1 such that:

$$\text{for every admissible extension } R \sqsupseteq E \text{, we have } \mathcal{M}[R] \models \Phi[\phi_1/F, \phi_2/G].$$

Let $D^-(E)$ be the set of ordered pairs $\langle \S x.\phi_1(x), \S x.\phi_2(x) \rangle$ of abstraction terms from D_1 such that $\mathcal{M}[R] \models \neg\Phi[\phi_1/F, \phi_2/G]$ for any admissible extension R of E. Let $D(E) = \langle D^+(E), D^-(E) \rangle$.

That is, we let $D^+(E)$ (alternatively: $D^-(E)$) consist of the ordered pairs of abstraction terms such that, for *any choice* of admissible extension R of E, these terms are identified (alternatively: distinguished) by the criterion of identity Φ in the term model based on D_1 and R.

Lemma 1 (a) If E is an approximation, then so is $D(E)$.

(b) D is monotone with respect to the ordering $\sqsubseteq$.

[16] Note that the symbol '$\sqsubseteq$' is being used ambiguously for the relation between one approximation and another that extends it, and for the relation between an approximation and an admissible refinement. In practice, this will cause no confusion.

Proof. For (a), it is straightforward to show that $D^+(E)$ is an equivalence relation and that $D^+(E)$ is disjoint from $D^-(E)$. For (b), consider an approximation $E = \langle E^+, E^- \rangle$ that is extended by another approximation $F = \langle F^+, F^- \rangle$. Then an equivalence relation R is an admissible extension of F only if it is an admissible extension of E. It follows that $D(E)$ is extended by $D(F)$, i.e. that D is monotone. $\dashv$

We can now describe the method of groundedness. We start with the empty approximation E_0 given by $E_0^+ = E_0^- = \varnothing$. Lemma 1 allows us to iterate applications of the operation D by defining $E_{\alpha+1} = \langle D^+(E_\alpha), D^-(E_\alpha) \rangle$ and $E_\lambda = \langle \bigcup_{\gamma<\lambda} E_\gamma^+, \bigcup_{\gamma<\lambda} E_\gamma^- \rangle$ for limit ordinals λ. As usual, the monotonicity of D and cardinality considerations ensure that we eventually reach a fixed point $E_\beta = E_{\beta+1}$. We say that E_β^+ and E_β^- consist of all the *grounded* facts about identity and distinctness. These facts have been established in a natural and conclusive way, despite the circularity worry.

Here it is easy to lose sight of an important fact. Although each of the admissible extensions that we consider in the course of the above procedure satisfies ($\star$), it is not the case that each of the E_α^+ does. Consider the case of Basic Law V, and let t and t' be the terms $\varepsilon u(u = u)$ and $\varepsilon u(u = u \vee u = u)$, respectively. Then $\mathcal{M}[E_1^+]$ thinks $t = t'$ but $\varepsilon u(u = t) \neq \varepsilon u(u = t')$, in violation of ($\star$). Thankfully, such pathologies are avoided when α is a fixed point.

Proposition 2 Assume that β is a fixed point. Then E_β^+ satisfies the closure condition ($\star$).

Proof. Assume $\langle t, t' \rangle \in E_\beta^+$. Let $T \in D_1$ and $T' = T[t'/t]$. We wish to show $\langle T, T' \rangle \in E_\beta^+$. Clearly, there is a formula ϕ such that $T = \S u.\phi$. We thus also have $T' = \S u.\phi'$, where ϕ' abbreviates $\phi[t'/t]$. Since β is a fixed point, it suffices to show that, for every admissible extension R of E_β, we have:

$$\mathcal{M}[R] \models \Phi[\phi/F, \phi'/G] \tag{4}$$

Since Φ is an equivalence relation, $\mathcal{M}[R]$ obviously thinks that ϕ is Φ-equivalent with itself. Hence (4) follows by an application of Leibniz's Law, which is available by Proposition 1. $\dashv$

To what extent does the method of groundedness succeed in resolving the circularity problem? That is, under what conditions does our least fixed point E_β give rise to a term model for the abstraction principle (Σ) in the context of predicative second-order logic (which is what we have restricted ourselves to, at least for the time being)? The answer turns out to be largely positive. The first step towards seeing this is the following proposition:

Proposition 3 Assume E_β is a fixed point such that $E^+_\beta \cup E^-_\beta = D_1 \times D_1$. Then $\mathcal{M}[E^+_\beta]$ is a model of each instance of (Σ) where the identity sign on the left-hand side is flanked by terms from D_1.

Proof. Assume first that $\S x.\phi = \S x.\psi$ is true in the model $\mathcal{M}[E^+_\beta]$. Then the pair $\langle \S x.\phi, \S x.\psi \rangle$ made it into E^+_α at some $\alpha < \beta$. This means that, for any admissible extension R of E_α, the identity criterion $\Phi[\phi/F, \psi/G]$ is true in the model $\mathcal{M}[R]$. By Proposition 2, one such admissible extension is E^+_β itself, whence it follows that $\Phi[\phi/F, \psi/G]$ is true in $\mathcal{M}[E^+_\beta]$, as desired. Assume next that $\S x.\phi = \S x.\psi$ is false in the model $\mathcal{M}[E^+_\beta]$. Then the pair $\langle \S x.\phi, \S x.\psi \rangle$ made it into E^-_α at some $\alpha < \beta$, and analogous reasoning establishes that $\Phi[\phi/F, \psi/G]$ is false in $\mathcal{M}[E^+_\beta]$. $\dashv$

If, on the other hand, $E^+_\beta \cup E^-_\beta \subset T \times T$, then we get only a 'partial model' for the mentioned class of instances of (Σ); that is, a model where some identities are neither deemed true (by the ordered pair of the terms flanking the identity sign being in E^+_β) nor false (by this pair being in E^-_β).

The next step is to determine under what conditions the assumption of Proposition 3 is true. A particularly interesting—and demanding—test case is *predicative Frege Theory*, that is, the theory based on Basic Law V and predicative second-order logic. We show in the next section that our method of groundedness gives rise to a natural model for the axiomatic version of predicative Frege Theory. So here the circularity worry has been addressed in a complete and satisfactory way.

5 A natural model for predicative Frege Theory

We now show that our method of groundedness does indeed yield a natural model for the axiomatic version of predicative Frege Theory, that is, the predicative theory whose sole non-logical axiom is (V_0), rather than the axiom scheme (V).[17] Readers who are willing to take our claim on faith may therefore skim this section or even skip ahead to the next section.

[17] As already stated, we believe this is equivalent to the example discussed in (Dummett, 1991, p. 221). It might be objected that Dummett is concerned with *first-order* Frege theory. But there is no real difference, as can be seen by considering the term sets defined in Section 5.2.

5.1 The languages and theories

Although our target is the theory based on (V_0), it will be expedient to do some of the constructions with a variable-binding operator—as in $\varepsilon x.\phi$—rather than an operator taking second-order variables to singular terms— as in εF. The constructions involving the former operator will then be put to use to prove results about the latter.

Let $\mathcal{L}_0$ be the language of monadic second-order logic with identity whose sole non-logical expression is an operator ε taking monadic second-order variables to first-order terms. Let $\mathcal{L}$ be the same language except that the non-logical expression ε is a variable-binding operator. Notice in particular that, unlike earlier parts of the article, we are here restricting ourselves to monadic second-order logic.

Let PV_0 be the $\mathcal{L}_0$-theory with predicative second-order comprehension and the sole non-logical axiom (V_0). Let PV be the $\mathcal{L}$-theory with predicative second-order comprehension and the axiom scheme (V).

5.2 The existence of a natural model for PV_0

We define a set D_2 of first-order formulas of $\mathcal{L}$ and a set D_1 of abstraction terms from $\mathcal{L}$ by simultaneous recursion as follows. Let D_2 contain every first-order formula ϕ from $\mathcal{L}$ with x as its only free variable and all of whose abstraction terms are in D_1. Let D_1 contain the abstraction term $\varepsilon x.\phi$ whenever ϕ is in D_2. Notice that every abstraction term in D_1 is closed, since the single free variable of ϕ is bound by the operator εx. This will become important below.

Our strategy is to let D_2 serve as the second-order domain, and D_1, as the term set on which we build up an equivalence relation E by means of our groundedness procedure. Our first main result is that this strategy works.

Theorem 1 The groundedness procedure applied to the term set D_1 and the second-order domain D_2 yields of model for PV_0.

The proof of the theorem relies on a lemma, which we state now and prove later.

Lemma 2 Each identity question $t_1 = t_2$ involving terms from D_1 is settled after finitely many steps. That is, there is an n such that $\langle t_1, t_2 \rangle$ is either in E_n^+ or in E_n^-.

Proof of Theorem 1. By Lemma 2, the groundedness procedure reaches a fixed point after ω many steps. So by Proposition 2, E^+_ω satisfies $(*)$. So by Proposition 1, $\mathcal{M}[E^+_\omega]$ is a model of predicative second-order logic.

We interpret εV, relative to a variable assignment σ, as the term $\varepsilon u.\phi$, where $\sigma(V) = \phi$. What remains is now only to show is that (V_0) is true in our model. Consider an instance

$$\varepsilon F = \varepsilon G \leftrightarrow \forall x(Fx \leftrightarrow Gx),$$

and let σ be an assignment. Assume $\sigma(F) = \phi$ and $\sigma(G) = \psi$. We must show that

$$\mathcal{M}[E^+_\omega] \models \varepsilon x.\phi = \varepsilon x.\psi \leftrightarrow \forall x(\phi \leftrightarrow \psi).$$

This follows from Proposition 3 and the fact that Lemma 2 ensures $E^+_\omega \cup E^-_\omega = D_1 \times D_1$. $\dashv$

5.3 Proving Lemma 2

We now state and prove some further lemmas, which will be useful in our proof of Lemma 2.

Lemma 3 For each natural number n, PV proves the standard first-order formalization of the claim that there are at least n distinct objects.

Proof. Consider the sequence of abstraction terms $\{t_i\}_{i<\omega}$ where $t_0 = \varepsilon u(u \neq u)$ and $t_{n+1} = \varepsilon u(u = t_n)$ for each natural number n. A proof by induction shows that PV proves $t_i \neq t_j$ whenever $i \neq j$. The lemma now follows straightforwardly. $\dashv$

For short, say that a formula is QF if it is quantifier-free. Next we establish a quantifier-elimination result, which says that first-order formulas that contain only closed abstraction terms are equivalent to QF formulas.

Lemma 4 Let ϕ be a first-order formula of $\mathcal{L}$ all of whose abstraction terms are closed. Let $t_1, \ldots t_n$ be a list of all the abstraction terms that occur in ϕ and all the variables with at least one occurrence in ϕ outside the scope of the operator ε. Then there is a QF formula ψ in $\mathcal{L}$ such that:

(a) PV $\vdash \phi \leftrightarrow \psi$,

(b) ψ is a disjunction of conjunctions, each conjunction containing, for each identity statement containing the t_i, either it or its negation, but no other conjuncts.

Proof. By the Löwenheim-Behmann theorem[18] there is a formula θ, provably equivalent (in pure first-order logic) to ϕ, which fits the above description of ψ except that it may also contain some additional conjuncts concerning the number of u distinct from all the t_i. These additional conjuncts say either that there are exactly k such u for some $k < n$, or that there are at least n such u. But by Lemma 3, PV proves the existence of infinitely many objects. So any one of the additional conjuncts can either be proved, in which case the additional conjunct itself can be deleted, or disproved, in which case the whole conjunction of which the additional conjunct is part can be deleted. This yields the desired ψ. $\dashv$

We observe that Lemma 4 would fail if we allowed quantification into the scope of the operator ε (which of course we would if we considered the schematic version of Basic Law V). To see this, consider the formula $s(x)$ that says that x is a singleton: $\exists y(x = \varepsilon u(u = y))$. All QF formulas have finite or cofinite extensions, whereas there are simple models where $s(x)$ has an extension that is neither finite nor cofinite.

Lemma 5 Let $\varepsilon x.\alpha$ be a first-order abstraction term from D_1. Then there is a QF formula β from D_2 such that $\varepsilon x.\alpha = \varepsilon x.\beta$ is settled positively after finitely many steps.

Proof. We begin by establishing the useful fact that, after m steps, the procedure recognizes the existence of at least m distinct objects. Consider the sequence $\{t_i\}_{i<\omega}$ of abstraction terms from Lemma 3. A proof by induction shows that, for all $i < j \leq m$, $t_i \neq t_j$ is established by step m.

Let $\varepsilon x.\phi_1, \ldots, \varepsilon x.\phi_n$ be a list of all the abstraction terms that occur in $\varepsilon x.\alpha$, with $\phi_n = \alpha$. We may assume that $\varepsilon x.\phi_i$ occurs as a proper part of $\varepsilon x.\phi_j$ only if $i < j$. Assume $\varepsilon x.\phi_i$ is the first term on the list that is not QF. Apply Lemma 4 to ϕ_i to obtain an equivalent QF formula ψ_i from D_2. After a suitably large finite number m_i of steps, the equivalence $\phi_i \leftrightarrow \psi_i$ follows by pure first-order logic and the number of objects established to be distinct by step m_i. Thus, by step $m_i + 1$, the identity $\varepsilon x.\phi_i = \varepsilon x.\psi_i$ will have been established. Substitute the latter term for the former in the remaining terms $\varepsilon x.\phi_{i+1}, \ldots, \varepsilon x.\phi_n$ and repeat the process, starting with 'Assume $\varepsilon x.\phi_i$', until the entire list consists of QF terms. The final item on the list that results in this way gives us the desired $\varepsilon x.\beta$ such that $\varepsilon x.\alpha = \varepsilon x.\beta$. $\dashv$

We now define a notion of ε*-rank*, which measures the depth of its nesting of ε-operators.[19]

[18] See (Burgess, 2005, p. 63).

[19] Our notion of ε-rank is closely related to the notion of rank defined in (Wright, 1998, p. 364). However, our definition avoids the problems afflicting the latter ((Dummett, 1998, pp. 383-4) and (Fine, 2002, pp. 97-8,

Definition 4 Let the ε-rank of any variable be 0. Let the ε-rank of an abstraction term $\varepsilon x.\phi$ be $n + 1$, where n is the highest ε-rank of any term occurring in ϕ.

We now return to the proof of Lemma 2, which we postponed above and on which our main Theorem 1 depends.

Proof of Lemma 2. Recall that the lemma says that each identity question $t_1 = t_2$ involving terms from D_1 is settled after finitely many steps. By Lemma 5, we may assume t_1 and t_2 to be QF.

We proceed by induction on ε-rank. For the base case, assume the terms have ε-rank 1. By Basic Law V, the identity question $t_1 = t_2$ reduces to the new question $\forall x(\phi_1 \leftrightarrow \phi_2)$. But since ϕ_1 and ϕ_2 contain no ε-operators and are QF, this new question is settled already at the outset of our procedure. So the question $t_1 = t_2$ is settled by stage 1. For the induction step, assume all identity questions involving terms from D_1 of ε-rank $\leq n$ is settled by our procedure after finitely many steps. Assume t_1 and t_2 have ε-rank $\leq n+1$. By Basic Law V, the identity question $t_1 = t_2$ reduces to the new question $\forall x(\phi_1 \leftrightarrow \phi_2)$, which involves only ε-terms of ε-rank $\leq n$. The induction hypothesis then ensures us that there is a natural number N such that *all* identity questions involving the ε-terms that occur in this new question are settled by stage N. Thus the new question too is settled by stage N. So the question $t_1 = t_2$ is settled by stage $N + 1$. $\dashv$

Note that the proof relies essentially on the quantifier elimination result from Lemma 4, which in turn depends on the restriction to closed abstraction terms. Thus, it is unlikely that anything like this strategy can be extended to provide a model for the second-order theory that consists of predicative comprehension and the *schematic* version of Basic Law V, because this theory crucially involves first-order quantification into the scope of the operator ε.

5.4 A more explicit description of the model

The previous theorem tells us that the groundedness procedure yields a model but provides little information about what this model is like. Our next theorem shows that more precise information about the model can be extracted.

fn. 27)) because of the very limited nature of the term set D_1 on which our notion is defined: every member of our D_1 is a *closed* abstraction term. So although our definition and the theorems which it underpins can be regarded as a partial vindication of Wright's project, it is important to realize the severity of the restrictions we impose.

Theorem 2 Consider an abstraction term t from D_1, and let $t_1, \ldots, t_n$ be the abstraction terms that occur in t (and thus are also in D_1). Then we can find a formula ψ such that:

- the model from Theorem 1 satisfies $t = \varepsilon x \psi$,
- ψ is either $x = t_1 \vee \ldots \vee x = t_n$ (with $x \neq x$ as a limit case, where $n = 0$) or $x \neq t_1 \wedge \ldots \wedge x \neq t_n$ (with $x = x$ as a limit case, where $n = 0$).

Proof. Assume $t = \varepsilon x.\phi$. By Lemma 4, ϕ can be rewritten as a disjunction of conjunctions, each conjunction containing, for each identity statement containing the t_i, either it or its negation, but no other conjuncts. Let ψ be this new formula. Next, we use the fact that all identity statements involving terms from D_1 are settled to eliminate all conjuncts from ψ not containing x: true conjuncts can be omitted, whereas for every false conjunct we can instead omit the disjunct in which this conjunct occurs. Let ψ' be the resulting formula.

Let θ be one of the disjuncts of ψ'. If θ contains as conjuncts two or more distinct identity statements of the form $x = t$, then θ imposes an impossible condition on x and the whole disjunct θ can simply be dropped from ψ'. We may thus assume that each disjunct is either a single identity statement of the form $x = t$ or $x = x$, or a conjunction of negated identity statements of these two forms. It is not hard to see that a disjunction of such formulas can be written in the form described in the theorem. $\dashv$

Corollary 1 Consider the following construction:

- Let T_1 be the set of the two terms $\varepsilon u(u = u)$ and $\varepsilon u(u \neq u)$.
- Let T_{n+1} be the set of all terms of the form $\varepsilon u(u = t_1 \vee \ldots \vee u = t_k)$ and $\varepsilon u(u \neq t_1 \wedge \ldots \wedge u \neq t_k)$, where each t_i is an elements of $\bigcup_{i \leq n} T_i$ and $t_1, \ldots, t_k$ is a strictly increasing sequence, in the lexicographic ordering.
- Then let $T = \bigcup_n T_n$.

The model for PV that results from the grounding procedure is isomorphic to the model where T serves as both first- and second-order domain, and where ε is interpreted as the identity mapping.[20]

[20](Dummett, 1991, p. 219) sketches a closely related model, which he claims is a model for first-order Frege theory. We suspect Dummett intended the same model as ours but misdescribed it slightly by defining D_{n+1} as 'the union of D_n with the set of all its finite and cofinite subsets': what we want are sets that are cofinite relative to the final model, not relative to D_n. If so, then our natural term model, whose existence Dummett

Proof. This is immediate from Theorem 2 and the observation that each of the abstraction terms t_i is of lower ε-rank than t. The requirement that $t_1, \ldots, t_k$ be a strictly increasing sequence is imposed in order to choose a single value-range term from each equivalence class of coreferring terms. $\dashv$

In unpublished work, Albert Visser has developed an alternative notion of 'naturalness' of models for abstraction principles. His analysis ties the idea of naturalness to initiality in a category of models and mappings. Somewhat surprisingly, Visser's candidate for a natural model of PV_0 is isomorphic to ours. We find it very satisfying that two independently motivated analyses of 'naturalness' should in this way give the same verdict.

5.5 Some open questions

Some consistent ways of going beyond the theory PV_0 are known. First, there is the theory PV, which is based on the schematic version of Basic Law V, rather than the axiomatic one, which (Heck, Jr., 1996) proves to be consistent. Then, there are the theories Δ^1_1-PV_0 and Δ^1_1-PV based on Δ^1_1-comprehension and the axiomatic and schematic versions of Basic Law V, respectively, and proved to be consistent in (Wehmeier, 1999) and (Ferreira and Wehmeier, 2002), respectively. We do not know whether there are natural models of any of these theories. (We are inclined to expect not. If so, this would establish the philosophically interesting point that not every consistent abstraction-based theory has a *natural* model—that is, a model generated by our groundedness procedure—and thus show our Theorem 1 to go well beyond the known fact that PV_0 is consistent.) Nor do we know whether our groundedness procedure yields a model for the schematic version of any other interesting abstraction principles.

6 Beyond predicative comprehension?

Two kinds of impredicativity were distinguished in Section 3 and shown to pose a challenge to the construction of term models for abstraction principles. The challenge posed by the impredicativity inherent in the abstraction principle itself has received a satisfying answer

denies on pp. 220–22, turns out to be isomorphic to the very model that he himself invokes as a purely technical trick in order to prove a consistency result. See also (Forster, 2008) for a discussion of this sort of construction and its relation to Church-Oswald models of set theory. Finally, the corollary also shows that the model that arises naturally from our application of the groundedness procedure is very similar to the model employed to prove the consistency of Basic Law V with Δ^1_1-comprehension in (Wehmeier, 1999).

in terms of our method of groundedness. As mentioned, however, this method only works when the second-order logic is predicative. This brings us to the second challenge, which concerns impredicative second-order logic. The problem is that the definition of satisfaction of an atomic predication of the form Vt that we described in Section 3.2 is only available for predicative second-order logic. Can we do better?

An attempt to do so, for the special case of Basic Law V, is found in (Kriener, 2014).[21] The restriction to predicative comprehension is a result of defining the satisfaction of Vt in a model in terms of the satisfaction of $\psi(t)$ in that model, where $\psi(x)$ is the formula assigned to V by the relevant variable assignment. What if we simply sidestep this reduction and handle the former satisfaction directly? In order to do so, we need an *application relation* $A \subseteq D_2 \times D_1$ which specifies when a formula from the second-order domain D_2 is to count as applying to a term from D_1. We can then define Vt to be satisfied by σ in the resulting model iff $\langle \sigma(V), t \rangle \in A$. Of course, for the resulting model to count as grounded, the application relation A would have to be constructed along with the equivalence relation R in a grounded manner. (Kriener, 2014) shows how this can be done.

As with our own strategy,[22] the alternative strategy can be carried out over an arbitrary base model. For simplicity, however, we restrict ourselves to sketching how the alternative strategy is implemented in pure term models.[23] Consider the language of pure second-order logic with identity ($=$) and a class abstraction symbol (ε). Models for this language are then constructed as follows. As before, we start with a domain D_1 that contains as elements all abstraction terms of the form $\varepsilon x.\phi$. In every model, an abstraction term denotes itself. Models also include an interpretation E for $=$ and the mentioned application relation A. In fact, A and E uniquely determine the model (because D_1 is held fixed), so a model can be denoted as $\mathcal{M}[A, E]$. Then we stipulate that:

$$\mathcal{M}[A, E] \models \varepsilon x.\phi = \varepsilon x.\psi \quad \text{iff} \quad \langle \varepsilon x.\phi, \varepsilon x.\psi \rangle \in E$$

$$\mathcal{M}[A, E] \models_\sigma Vt \quad \text{iff} \quad \langle \sigma(V), t \rangle \in A$$

The strategy now proceeds much as in the present article. As with our strategy, a method

[21]Strictly speaking, (Kriener, 2014) operates in a somewhat different setting, namely a first-order language with a two-place predicate 'η', where the intended reading of '$x \eta y$' is 'x is a member of the extension y'. However, it is straightforward to adapt his strategy to our setting. This is what we do in what follows.

[22]See section 2.6.

[23]For a detailed description of the implementation of this strategy, see section 6 of (Kriener, 2014).

is specified for successively generating better models for the language. Again, the key ingredient is a jump operator that generates an 'improved' model from a given one. We need a notion of an identity interpretation *extending* ($\sqsupseteq$) another identity interpretation, and an application interpretation *extending* another application interpretation. And again, 'extending' entails extending as a superset, but further 'admissibility' conditions may be imposed. Then we stipulate a jump operation for the identity relation and a jump operation for the application relation. The 'jump' of identity is generated by basic law V:

$$\mathcal{E}(A, E) = \{\langle \varepsilon x.\phi, \varepsilon x.\psi \rangle : \forall \langle B, F \rangle \sqsupseteq \langle A, E \rangle \Rightarrow \mathcal{M}(B, F) \models \forall x(\phi(x) \leftrightarrow \psi(x))\},$$

And the jump of application is given by:

$$\mathcal{A}(A, E) = \{\langle \phi, \varepsilon x.\psi \rangle : \forall \langle B, F \rangle \sqsupseteq \langle A, E \rangle \Rightarrow \mathcal{M}(B, F) \models \phi(\varepsilon x.\psi)\}.$$

At limit stages, we take unions. It can then be shown that this is a monotone process, which of course always leads to a fixed point. Given suitable admissibility conditions, fixed point models can be shown to have desirable properties, such as satisfying the axiom of extensionality for classes.

Unfortunately, the method is beset by severe problems (as is acknowledged in the article): comprehension fails for some extremely simple—indeed predicative—conditions. An example is $x = t$ for an arbitrary abstraction term t. This in turn undermines the definition of singleton classes. It is an open question, however, whether the method fares better with abstraction principles other than Basic Law V.

The resulting situation seems to us to be the following. Our definition of satisfaction of an atomic predication Vt is immensely natural. Although it requires a restriction to predicative second-order logic, this restriction has the advantage of being systematic and well understood. The alternative definition of satisfaction of an atomic predication Vt is less natural and requires no less severe restrictions to the comprehension scheme of the background second-order logic. In fact, these restrictions are in some respects *more severe* than those associated with predicative second-order logic, such as in the mentioned example of singleton concepts and their extensions. More worrisome yet, these restrictions are poorly understood and have an air of *ad hocery*.

Although a conclusive assessment will have to await further investigation, the evidence currently available suggests that the challenge arising from the impredicativity of the second-order logic is far more robust than that arising from the impredicativity of the abstraction principles. Our tentative conclusion is thus that Dummett's diagnosis of 'the serpent' in Frege's paradise is at least half right: the Fregean project of constructing term models for abstraction principles is indeed incompatible with impredicative second-order logic.

7 Conclusion

This paper has examined the extent to which it is possible to construct term models for abstraction principles. We have disentangled two apparent limitations to which the construction is subject and on the basis of which (Dummett, 1991, ch. 17) rejected a closely related strategy of Frege's as fundamentally flawed. The first apparent limitation has been upheld: term models are available only when the background second-order logic is predicative. The only ways to escape this conclusion have been found unacceptable, namely to disallow quantification into abstraction terms, which the standard language permits and many applications require; or, alternatively, to rely on a different but poorly understood explication of the idea of groundedness, which requires us to foresake even singleton classes.

The second apparent limitation suggests that term models cannot be constructed for any abstraction principle that is impredicative. We have shown how this apparent limitation can be circumvented by means of an elegant method of groundedness. The power of this method was illustrated by constructing a grounded term model for the axiomatic version of Basic Law V with predicative second-order logic. The construction of this natural model goes beyond the known fact that the theory is consistent (Heck, Jr., 1996): for not every model is a natural model.[24]

References

Burgess, J. P. (2005). *Fixing Frege.* Princeton University Press, Princeton, NJ.

[24]We are grateful for valuable comments from Salvatore Florio, Jönne Kriener, Hannes Leitgeb, Jon Litland, Sam Roberts, Stewart Shapiro, Sean Walsh, Philip Welch, and two anonymous referees, as well as from the audiences of a Bristol workshop on ontological dependence and the mathematical logic seminar at the University of Oslo. Both authors were supported by AHRC-project Foundations of Structuralism and the second author also by an ERC Starting Grant (241098).

Dummett, M. (1991). *Frege: Philosophy of Mathematics.* Harvard University Press, Cambridge, MA.

Dummett, M. (1998). Neo-Fregeans in Bad Company? In Schirn, M., editor, *Philosophy of Mathematics Today.* Clarendon, Oxford.

Ferreira, F. and Wehmeier, K. (2002). On the Consistency of the Δ^1_1-CA Fragment of Frege's *Grundgesetze*. *Journal of Philosophical Logic*, 31:301–311.

Fine, K. (2002). *The Limits of Abstraction.* Oxford University Press, Oxford.

Forster, T. (2008). The iterative conception of set. *Review of Symbolic Logic*, 1(1):97–110.

Ganea, M. (2007). Burgess' pv is robinson's q. *Journal of Symbolic Logic*, 72(2):619–624.

Hale, B. and Wright, C. (2001). *Reason's Proper Study.* Clarendon, Oxford.

Heck, Jr., R. G. (1996). The Consistency of Predicative Fragments of Frege's *Grundgesetze der Arithmetik*. *History and Philosophy of Logic*, 17:209–220.

Heck, Jr., R. G. (1997). *Grundgesetze der Arithmetik* I §§29-32. *Notre Dame Journal of Formal Logic*, 38(3):437–474.

Heck, Jr., R. G. (2012). *Reading Frege's Grundgesetze.* Oxford University Press.

Horsten, L. (2010). Impredicative identity criteria. *Philosophy and Phenomenological Research*, 80(2):411–439.

Horsten, L. and Leitgeb, H. (2009). How abstraction works. In Hieke, A. and Leitgeb, H., editors, *Reduction and Elimination in Philosophy and the Sciences*, pages 217–226. Ontos Verlag, Frankfurt.

Kriener, J. (2014). The groundedness approach to class theory. *Inquiry*, 57(2):244–273.

Leitgeb, H. (Forthcoming). Abstraction grounded. In Ebert, P. and Rossberg, M., editors, *Abstractionism in Mathematics:* Status Belli.

Linnebo, Ø. (2004). Frege's Proof of Referentiality. *Notre Dame Journal of Formal Logic*, 45(2):73–98.

Linnebo, Ø. (2009a). Bad Company Tamed. *Synthese*, 170(3):371–391.

Linnebo, Ø. (2009b). Introduction [to a special issue on the bad company problem]. *Synthese*, 170(3):321–329.

Linnebo, Ø. (2013). How to harness Basic Law V. Forthcoming in Massimiliano Carrara and Friederike Moltmann (eds.), *Plurality and Unity: Philosophy, Logic, and Semantics* (Oxford: Oxford University Press).

Studd, J. (2014). Abstraction Reconceived. forthcoming in the British Journal of Philosophy of Science.

Wehmeier, K. (1999). Consistent Fragments of *Grundgesetze* and the Existence of Non-Logical Objects. *Synthese*, 121:309–28.

Wright, C. (1998). The Harmless Impredicativity of $N^{=}$ (Hume's Principle). In Schirn, M., editor, *Philosophy of Mathematics Today*. Clarendon, Oxford.