The Aristotelian Society
133rd session – 2011/2012

7 November 2011 | 16.15 – 18.00

# Reference by Abstraction

Øystein Linnebo
Birkbeck, University of London

The Woburn Suite
Senate House (South Block)
University of London
Malet Street
London WC1E 7HU
United Kingdom

** non-citable draft paper **

## ABSTRACT

Frege suggests that criteria of identity should play a central role in an account of reference, which will be capable of explaining reference to abstract objects. Inspired by Frege's suggestion, this paper develops a simplified but precise model of how we may come to refer to abstract letter types. We start with an interpreted language concerned with letter tokens. We then add vocabulary suitable for talking about letter types and adopt precise rules which ensure that the extended language is used precisely as if it was concerned with letter types. Since the rules are logically impeccable and invoke only unproblematic concrete objects, this extension is legitimate. There are nevertheless reasons to interpret the extended language as not only apparently but genuinely referring to letter types. But the reductionist character of the rules is used to argue that the abstract referents are metaphysically "lightweight".

## BIOGRAPHY

Øystein Linnebo is Professor of Philosophy at Birkbeck, University of London, where he arrived in January 2010, having held positions at the Universities of Bristol, Oxford, and Oslo. He obtained his PhD in Philosophy from Harvard University in 2002 and an MA in Mathematics from the University of Oslo in 1995.

Linnebo's research interests lie in philosophical logic (especially plural and higher-order logic, logical paradoxes, absolute generality); philosophy of mathematics (especially structuralism and Fregean approaches); metaphysics (especially abstract objects, criteria of identity, modality); early analytic philosophy (especially Frege); as well as parts of philosophy of language and philosophy of science. He is currently writing a book developing and defending a broadly Fregean approach to ontology and philosophy of mathematics

In 2010-2013 Linnebo is leading a European Research Council-funded project entitled "Plurals, Predicates, and Paradox: Towards a Type-Free Account".

# Reference by abstraction

Øystein Linnebo
Birkbeck, University of London

## 1 Introduction

In *Grundlagen* §62 Frege proposed that criteria of identity have an important role to play in an account of reference.

> If for us the symbol $a$ is to denote an object, then we must have a criterion which determines in every case whether $b$ is the same as $a$, even if it is not always within our power to apply this criterion. (Frege, 1953, §62)

The basic idea seems to me clear enough. Reference to an object has something to do with an ability to distinguish the referent from other objects and to re-identify it in different situations. To borrow an example from Quine, suppose you are standing at the bank of a river, watching the water floating by. What is required for you to refer to the river, as opposed to a particular segment of it, or the totality of its water, or the current temporal part of this water? According to Quine, you must at least implicitly be operating with some criterion of identity that informs you when two sightings of water count as sightings of the same referent. Or, to switch to another example, due to Frege, reference to a direction involves an ability to tell whether two directions—presented to you as the directions of two given lines—are identical. So in both cases, your act of referring is guided by a criterion of identity.

As Frege is keenly aware, one of the attractions of this idea is its great generality. The notion of a criterion of identity appears to be applicable to all kinds of objects, abstract as well as concrete. So if an account of reference can be built up around the notion of a criterion of identity, then this account is likely to enjoy the same generality. This would be a major advantage over competing approaches, such as ones that tie reference closely to causal connections. Since natural language suggests there is such a thing as reference to abstract objects (for instance to directions), it would be good to have an approach that doesn't close down the possibility of such reference at the very outset.

Despite the enthusiasm generated by Frege's idea, there has, in my opinion, been no satisfactory development of it.[1] The aim of this article is to improve on this situation by developing the Fregean idea in some detail. Although my inspiration is Frege, I do not claim that the resulting account is true to Frege's own intentions. My aim is systematic, not exegetical. Moreover, I make various simplifying assumptions—for which I offer no apology. What is most urgently needed at this stage is a properly worked out *model* of the phenomenon that Frege calls to our attention, not an account that fully matches our immensely rich and varied practices. If explanatorily successful, a scientific model can later be tweaked and extended to obtain a better fit with the data.

My plan is to consider a community of language users who refer to and quantify over various kinds of concrete objects, including letter tokens (henceforth 'inscriptions'). I then describe a situation in which the community extends their language by adding some new vocabulary and begins to use their extended language precisely *as if* they are referring to and quantifying over letter types (henceforth 'letters'). For instance, they begin to regard it as correct to utter 'this letter = that letter', pointing to two inscriptions, just in case the inscriptions count as equivalent by some agreed standard. (Here I make the simplifying assumption that there is an appropriate standard that all the speakers have internalized.) Our question is then whether, after extending their language, the members of the community are *actually* referring to letters (and thus to abstract objects)? On the one hand, everything is precisely as if they are. But on the other, all that has happened is that the speakers have come to regard it as correct to utter various sentences of the extended language just in case certain conditions on inscriptions (and thus concrete objects) are satisfied.

I argue that there are in fact reasons to prefer an interpretation of the extended language on which the speakers refer to letters. These abstract referents are obtained by a form of abstraction on inscriptions. However, the reductionistic character of the principles governing the correct use of the extended language is undeniable. I attempt to use this fact to articulate a sense in which the abstract referents are 'metaphysically lightweight'.

The resulting account of reference differs substantially from extant developments of the Fregean idea with which I began. The *cognoscenti* may appreciate an overview. My account avoids some controversial claims associated with the influential neo-Fregean approach of Bob

[1] I am, however, greatly indebted to the important discussions of this Fregean idea, and related ones, in (Dummett, 1981), (Wright, 1983), (Hale, 1987), (Dummett, 1991), and (Hale and Wright, 2001a).

Hale and Crispin Wright. I make no claim about the priority of syntactic categories to ontological ones.[2] (I would, however, deny that ontological categories can be made sense of independently of semantic notions.[3]) Nor do I invoke a deflationist conception of reference which equates the referentiality of a singular term $t$ with the truth of the object language sentence $\exists x(x = t)$.[4] And I avoid claims about the 'recarving' of content or about speakers' ability to translate sentences of the extended language into the base language.[5] Perhaps the most distinctive new aspect of my account is the view that, while there really is such a thing as reference to abstract objects, these objects are 'metaphysically lightweight'. This aspect brings with it another novelty, namely a restriction to abstraction principles that are predicative in the sense that the abstract objects described lie in an extended domain, not in the original one.[6]

## 2 The linguistic data

Our first question is how to make precise the idea that the community uses the extended language precisely as if they are referring to letters. This idea has to be cashed out in the form of a characterization of the linguistic data for which we will later attempt to determine the best semantic interpretation. My general policy will be to allow the data to be characterized in sophisticated and high-level terms—so long as we do not beg any questions about what is the best semantic interpretation of the extended language. For instance, I will regard the syntax of each language as already settled.[7] Now for the details.

The community starts with a base language $\mathcal{L}_0$, which is used to refer to and quantify over various concrete objects, including inscriptions. I assume—simplifying substantially—that $\mathcal{L}_0$ is an ordinary first-order language with identity and with variables $x_i$ for each natural number $i$. I further assume that all parties to the philosophical debate agree on a classical

[2] Contrast (Wright, 1983), (Hale, 1987), and (Hale and Wright, 2001a, ch. 1–2).

[3] In this respect I am closer to (Rumfitt, 2003) than to Dummett, Hale, and Wright.

[4] See for instance (Dummett, 1956) and (Wright, 1983, p. 83), as well as (Dummett, 1991) for criticism. In the sense of (Hodes, 1990), I aim to show that certain abstract singular terms have 'thick' ontological commitment, not just 'thin'.

[5] See for instance (Wright, 1983), (Hale, 1987) (Hale, 1997), and (Rosen, 1993).

[6] However, the last two features of my account resonate with aspects of Dummett's view. See for instance (Dummett, 1981, ch. 14) and (Dummett, 1991).

[7] Contrast Hale and Wright, who attempt to determine the syntactic category of an expression on the basis of various grammatical and inferential considerations; cf. note 2. Here I rely on the fact that my notion of syntactic category is less philosophically laden than theirs, as it is no longer claimed to be prior to any ontological notions.

referential interpretation $I$ of $\mathcal{L}_0$, whose domain, $D_0$, consists of objects that are not currently in dispute. While this is of course a substantive assumption, it is neutral with regard to our principal question of how the extended language is to be interpreted.

It will be convenient to let the extended language, $\mathcal{L}_1$, be a *two-sorted* first-order language.[8] The first of its sorts is exactly like the single sort of $\mathcal{L}_0$: the same variables, individual constants, and predicates. I thus call this *the 0-sort*. The second sort, which I call *the 1-sort*, can be thought of as a sort reserved for talking about letters. This sort has a distinct collection of individual constants and variables $y_i$ for each natural number $i$. There is also an operator § which applies to any 0-term to form a 1-term (where an $i$-term is a term of sort $i$). Each $n$-place predicate of $\mathcal{L}_1$ comes with an index that specifies the sort of term to which each of its $n$ argument places is open. (In practice, this index will often be omitted.) In particular, $\mathcal{L}_1$ has two identity predicates, $=$ and $=_1$, which may be flanked by any two 0-terms and 1-terms respectively; other combinations are deemed not well formed.

Why let $\mathcal{L}_1$ be a two-sorted language? The reason is to postpone the so-called Julius Caesar problem, which concerns mixed identity statements, like 'this inscription = that letter'. Although this problem is no doubt important, it is independent of the main concerns of this article. Let's therefore adopt the strategy of the person who lent his name to the problem: *divide et impera*. This strategy suggests adopting a two-sorted language where the Caesar problem cannot even be formulated, and where our main concerns thus come more sharply into focus. A solution to the Caesar problem will have to await another occasion.[9]

As mentioned, I wish to assume that the community uses the extended language $\mathcal{L}_1$ precisely as if they are referring to and quantifying over letters, obtained by abstraction on inscriptions. In order to spell out what this amounts to, we need to assume that the members of the community have internalized a standard for when two things are to count as inscriptions of the same letter. What can we say about the relation $\sim$ on the domain $D_0$ to which this standard gives rise? If $a$ and $b$ count as inscriptions of the same letter, then obviously so do $b$ and $a$. So the relation has to be symmetric. An analogous argument can be given for its transitivity. If $a$ isn't an inscription at all, however, then *a fortiori* $a$ and $a$ are not inscriptions of the same letter. So we shouldn't demand that the relation be reflexive. I therefore assume that $\sim$ is a symmetric and transitive but not necessarily reflexive relation. (Such relations

[8]See (Enderton, 2001, Section 4.3) for an introduction to many-sorted logic.

[9]See (Hale and Wright, 2001b) for an influential discussion, with which I am broadly sympathetic.

are known as *partial equivalence relations*.)

I said that the relation $\sim$ specifies 'when two things are to count as inscriptions of the same letter'. It is important not to misunderstand this. The idea is not that, for $a \sim b$ to hold, there must *actually be* a letter of which both $a$ and $b$ are inscriptions. This would be far too demanding. The words 'are to count as' are essential: the relation $\sim$ specifies the conditions under which members of the community *speak as if* two objects are inscriptions of the same letter. And *this* relation can clearly be grasped by people without any antecedent understanding of letters; indeed, the relation is, to a high degree of approximation, implemented in existing physical devices such as scanners.[10]

We are now ready to describe the principles that govern the use of $\mathcal{L}_1$. The principles specify the condition under which the speakers regard the assertoric utterance of a formula of $\mathcal{L}_1$ as correct. When this condition is met, we say that the formula is *assertible* in the relevant context.[11] To say that a formula is assertible in a context is thus not to say that it is true, only that the community regards it as correct. I will here settle for a relatively informal characterization of the assertibility conditions; a precise characterization is provided in Appendix A. Since free variables are involved, the assertibility conditions have to be stated relative to a string of objects, which can be thought of as the values of the free variables in question.

Let's begin with the sublanguage $\mathcal{L}_0$, which we have assumed has a semantic interpretation $I$. So the principles governing the use of $\mathcal{L}_1$ had better be compatible with this interpretation. This means that any formula $\phi$ of $\mathcal{L}_0$ must be regarded as assertible of a string of objects from $D_0$ just in case $\phi$ is true of these objects under the interpretation $I$.

The more interesting assertibility conditions are the ones governing formulas of $\mathcal{L}_1$ that are not in the base language $\mathcal{L}_0$. Let's begin with two simple examples. For any objects $a$ and $b$ from $D_0$ we have:

- '$\S x_1 =_1 \S x_2$' is assertible of $a$ and $b$ iff $a \sim b$

  [Roughly: members of the community regard it as correct to utter 'this letter $=_1$ that letter', pointing to $a$ and $b$ respectively, just in case $a \sim b$.]

[10] Of course, a certain amount of idealization is involved in assuming that people have internalized a partial equivalence relation. As is well known, vagueness poses a threat to the transitivity of the relation.

[11] By characterizing the use of $\mathcal{L}_1$ in terms of assertibility conditions, I avoid having to claim that speakers of $\mathcal{L}_1$ proceed by translating its formulas into $\mathcal{L}_0$, or that these formulas somehow provide 'recarvings' of meanings expressible already in $\mathcal{L}_0$. Compare footnote 5.

- '$\text{Vow}^1(\S x_1)$' is assertible of $a$ iff $a$ is an inscription of a vowel

  [Roughly: members of the community regard it as correct to utter 'this letter is a vowel', pointing to $a$, just in case $a$ is the inscription of a vowel.]

More generally, each $n$-place letter predicate $\mathbf{F}$ is associated with an assertibility condition $\phi_{\mathbf{F}}$ such that:

$$\mathbf{F}(\S x_1, \ldots, \S x_n) \text{ is assertible of } a_1, \ldots, a_n \quad \text{iff} \quad \phi_{\mathbf{F}}(a_1, \ldots, a_n)$$

We also need to assume that $\sim$ is a congruence with respect to $\phi_{\mathbf{F}}$; that is, that $\phi_{\mathbf{F}}$ doesn't distinguish between objects that are equivalent under $\sim$:

(1) $$a_1 \sim b_1 \wedge \ldots \wedge a_n \sim b_n \rightarrow \big(\phi_{\mathbf{F}}(a_1, \ldots, a_n) \leftrightarrow \phi_{\mathbf{F}}(b_1, \ldots, b_n)\big)$$

For instance, if the speakers regard it as correct to assert 'this letter is a vowel', pointing at an inscription $a$, then they must also regard it as correct to assert the sentence when pointing at an equivalent inscription $b$. This ensures that they treat equivalent objects as inscriptions of one and the same letter.

What about formulas with free 1-variables, that is, variables of the sort that function as if they are reserved for letters? It is not an option to quantify over letters when we characterize the assertibility conditions of such formulas. For our goal is to characterize the linguistic data in a way that only mentions concrete objects. Fortunately, the assertibility conditions of formulas with free 1-variables can be characterized relative to a string of inscriptions and other concrete objects. See Appendix A for details.

To complete our informal characterization of the assertibility conditions governing $\mathcal{L}_1$, we must specify how connectives and quantifiers are handled. Here there are no surprises. A negation $\neg\phi$ is assertible relative to a string of objects just in case it is not the case that $\phi$ is assertible relative to this string. A disjunction $\phi \vee \psi$ is assertible relative to the concatenation of two strings just in case $\phi$ is assertible relative to the first string or $\psi$ is assertible relative to the second. Finally, a quantification $\forall v\, \phi$ is assertible relative to a string $s$ just in case $\phi$ is assertible relative to every string $s'$ just like $s$ except additionally making an assignment to $v$. (We assume that the existential quantifier is defined in terms of the universal quantifier in the ordinary way.)[12]

I claim that a community that uses $\mathcal{L}_1$ in accordance with these assertibility conditions speaks precisely as if they are referring to and quantifying over letters. This should already be fairly clear. Additionally, the claim will be made formally precise and given a rigorous defense in the next section, where we give precise content to the phrase 'refer to and quantify over letters'.

## 3 Two competing interpretations

We can now give an improved statement of a question posed in the introduction. Assume that a community expands their language from $\mathcal{L}_0$ to $\mathcal{L}_1$ in the way just described. What, then, is the best semantic interpretation of $\mathcal{L}_1$? On the one hand, the principles governing the use of $\mathcal{L}_1$ make no mention of abstract objects such as letters. So to ascribe reference to such objects to terms of $\mathcal{L}_1$ seems to be to go far beyond what is warranted by the linguistic data. An interpretation that stays closer to the data seems preferable. On the other hand, $\mathcal{L}_1$ is used precisely as if speakers are referring to and quantifying over letters. So it seems appealing to take appearances at face value and interpret the letter terms as actually referring to and ranging over abstract letter types. I will now describe two competing interpretations of $\mathcal{L}_1$, one corresponding to each of the two mentioned considerations.

For reasons that will become apparent, I call the first interpretation *semantic reductionism*. This interpretation seeks to stay as close as possible to the data. In fact, the interpretation is read off directly from the assertibility conditions. Instead of saying that a letter term $t$ is *associated with* an inscription $a$, we say that $t$ *refers to* $a$. And each letter predicate $F$ is interpreted as a predicate of inscriptions, true of a string of inscriptions just in case the community regards the predicate as assertible relative to this string. Finally, connectives and quantifiers are interpreted in the obvious way. As a result, on the reductionist interpretation a formula is true relative to a string of objects just in case the formula is assertible relative to

[12]This assignment of assertibility conditions generalizes readily from the case of letters to other predicative abstractions. It is instructive to compare with an impredicative abstraction, such as Hume's Principle, which says that two Fregean concepts $F$ and $G$ have the same number—in symbols: $\#F = \#G$—just in case the there is a dyadic relation that one-to-one correlates the $F$s and the $G$s—in symbols: $F \approx G$. This is impredicative because the numbers are supposed to belong to the domain on which $F$ and $G$ are defined. But this feature also means that the assertibility condition of $\#F = \#G$, namely $F \approx G$, is not given in unproblematic terms: for this condition's obtaining or not will depend on various numerical questions, perhaps even the one whose assertibility condition we are trying to fix.

that string. In effect, the 'letter talk' of $\mathcal{L}_1$ is interpreted as just a peculiar form of 'inscription talk' where the letter identity predicate $=_1$ is interpreted as $\sim$.

The second interpretation of $\mathcal{L}_1$, which I call *semantic non-reductionism*, takes at face value the fact that the speakers behave precisely as if they are talking about letters. A letter term $t$ associated with an inscription $a$ is interpreted as referring to the abstract letter type of $a$. It is convenient to express this by means of an abstraction operation $a \mapsto \overline{a}$ that sends an inscription to its abstract letter type but is undefined on non-inscriptions.[13] The operation is subject to the following conditional abstraction principle:

$$(\Sigma) \qquad a \sim a \wedge b \sim b \rightarrow \big(\overline{a} = \overline{b} \leftrightarrow a \sim b\big)$$

Since $c \sim c$ just in case $c$ is an inscription, $(\Sigma)$ says that the operation maps two inscriptions $a$ and $b$ to the same letter (and thus, in particular, is defined $a$ and $b$) just in case $a \sim b$. If a letter term $t$ is associated with an inscription $a$, we let its referent be $\overline{a}$, rather than $a$ itself as the reductionist interpretation would have it.

We define a special letter domain $D_1$ as the range of the abstraction operation, that is, as the set of letters with inscriptions in $D_0$:

$$\text{(B1)} \qquad b \in D_1 \leftrightarrow \exists a(a \in D_0 \wedge b = \overline{a})$$

Just as 0-terms range over and refer to objects in the ordinary domain $D_0$, so the non-reductionist holds that 1-terms range over and refer to objects in $D_1$.

Next we consider the interpretation of predicates. So let $F$ be an $n$-place letter predicate of $\mathcal{L}_1$, which is assertible of $a_1, \ldots, a_n$ just in case $\phi_F(a_1, \ldots, a_n)$. As we have seen, the semantic reductionist appropriates this assertibility condition for her own semantic purposes by laying down that:

$$F \text{ is true of } a_1, \ldots, a_n \text{ on } R \quad \text{iff} \quad \phi_F(a_1, \ldots, a_n)$$

where $R$ is the reductionist interpretation. In keeping with her aims, the reductionist thus

[13] How should we handle cases where the operation is undefined? My preferred option is to adopt a primitive two-place predicate Abst, where Abst$(a, b)$ means that $b$ is the abstract letter type of $a$, and then to regard $\overline{a}$ as a mere shorthand for 'the $b$ such that Abst$(a, b)$'. An alternative is to use a negative free logic.

takes the predicate to be defined only on the concrete objects from $D_0$. The non-reductionist, on the other hand, wishes to interpret the predicate as defined on his new letter domain $D_1$. He achieves this by defining a new condition $\overline{\phi_F}$ on $D_1$ such that:

(B2) $$\forall a_1, \ldots, a_n \in D_0 \; [\overline{\phi_F}(\overline{a_1}, \ldots, \overline{a_n}) \leftrightarrow \phi_F(a_1, \ldots, a_n)]$$

(This is well defined because $\sim$ is a congruence with respect to $\phi_F$.) The non-reductionist then interprets the letter predicate $F$ by means of this new condition $\overline{\phi_F}$:

$$F \text{ is true of } b_1, \ldots, b_n \text{ on } N \quad \text{iff} \quad \overline{\phi_F}(b_1, \ldots, b_n)$$

where $N$ is the non-reductionist interpretation. This treatment of letter predicates achieves what the non-reductionist wants, namely an interpretation that is defined on letters but is compatible with the linguistic data associated with the assertibility conditions.

Finally, the non-reductionist semantics treats connectives and quantifiers in the obvious way. It is worth recalling, however, that the non-reductionist lets the 1-variables range over the special letter domain $D_1$ rather than over the ordinary domain $D_0$, as on the reductionist semantics.

The two interpretations of $\mathcal{L}_1$ that I have just presented are obviously quite different. It is not hard to see that they ascribe different canonical truth-conditions—and therefore different meanings—to many sentences of $\mathcal{L}_1$. Moreover, sentences of $\mathcal{L}_1$ can incur different ontological commitments on the two interpretations. For instance, on the non-reductionist interpretation, the sentence $\exists y_1 \, \textsc{Vow}(y_1)$ is ontologically committed to a letter, while on the reductionist interpretation, it is committed only to an inscription. In another respect, however, the two interpretations are closely related. We already have the resources to pinpoint how. Consider the two 'bridge principles' (B1) and (B2) encountered above, which relate some key ingredients of the two interpretations. It is not hard to verify that, *modulo* these bridge principles, the truth-conditions generated by the two interpretations are equivalent. (The result is given a precise statement as Proposition 2 in Appendix C, where a proof can also be found.)

This technical result is philosophically significant. For one thing, the result enables us to give precise content to, and proof of, my claim that to use $\mathcal{L}_1$ in accordance with the assertibility conditions is to speak precisely as if one is referring to and quantifying over letters. Details are provided in Appendix C. For another, the result sheds light on the nature of the disagreement between the two interpretations. The interpretations are in complete agreement about which sentences are true on which occasions of use. Their disagreement concerns the boundary between semantics and everything else. The reductionist will complain that the non-reductionist semantics involves an unnecessary and problematic detour via abstract semantic values. The bridge principles are then needed to convert the non-reductionist's truth condition to the reductionist's truth-conditions, which are closer to the linguistic data. By contrast, the non-reductionist will complain that the reductionist mistakenly incorporates certain extra-semantic principles into her semantics. Consider for instance the reductionist's claim that the truth of an identity flanked by two letter terms is a matter of the two associated inscriptions' being equivalent. From the non-reductionist's point of view, this claim is a confused amalgamation of the semantic principle that the truth of an identity statement is a matter of the two referents' being identical, and the metaphysical principle that the identity of letters is a matter of the equivalence of the associated inscriptions. I will return to this theme in Section 6.

## 4 Why a non-reductionist interpretation would be preferable

In this section, I assume that both of the interpretations described above are available and discuss whether one of them is better suited as a semantic interpretation of $\mathcal{L}_1$. I formulate some constraints on the choice of an interpretation, and based on these, I defend the non-reductionist interpretation. The assumption that this interpretation is available, despite being ontologically more demanding, will be defended in the next section.

### 4.1 The principle of charity

According to the principle of charity, we should, in the absence of countervailing considerations, interpret a language so as to respect its speakers' dispositions concerning the truth of sentences. The semantic interpretation should be such that sentences which the speakers firmly regard as true are indeed true.

Of course, the presumption in favor of a charitable interpretation can be overridden where there is reason to believe that the speakers are guilty of conceptual or factual errors. I argue in Section 5 that our community of speakers $\mathcal{L}_1$ are not. Regardless, the principle of charity is powerless to differentiate between the reductionist and non-reductionist interpretations of $\mathcal{L}_1$, because (assuming the bridge principles just mentioned) both interpretations satisfy the principle perfectly. On both interpretations, a sentence is true just in case it is assertible according to the community's conventions.

### 4.2 The principle of compositionality

Next, we often require that a semantic interpretation be *compositional*. Let me explain. In semantics and the philosophy of language it is widely assumed that each component of a complex expression makes some definite contribution to the meaning of the complex expression. This contribution is known as the *semantic value* of the component. I will write $[\![\mathbf{E}]\!]$ for the semantic value of an expression **E**. For instance, Frege held that the semantic value of a sentence is its truth-value and that the semantic values of other expressions are their contributions to the truth-values of sentences in which they occur. In particular, the semantic value of a singular term is just its referent. The notion of a semantic value thus generalizes the ordinary notion of reference. Where the notion of reference is intended primarily for singular terms, the notion of semantic value is meant to be applicable to expressions of all grammatical categories.

It is also widely assumed that the semantic value of a complex expression is functionally determined by the semantic values of its components and their syntactic mode of combination.[14] This assumption is known as *compositionality*. For instance, according to Frege the semantic value of a simple sentence such as 'John runs' is determined by the equation:

(2) $$[\![\text{John runs}]\!] = [\![\text{runs}]\!]([\![\text{John}]\!])$$

That is, the semantic value of the sentence 'John runs' is the result of applying the function which is the semantic value of the predicate 'runs' to the argument which is the semantic value of the subject 'John'. More generally, the principle of compositionality can be formulated as follows.

[14]The most important possible exceptions concern idioms and intensional contexts. Since the language with which we are concerned is non-idiomatic and extensional, these possible exceptions need not worry us.

**Compositionality**

For any syntactic operation $C$ that is applicable to syntactic expressions $E_1, \ldots, E_n$, there is a semantic operation $C^*$ corresponding to $C$ such that the semantic value of the result of applying the syntactic operation $C$ to the expressions $E_1, \ldots, E_n$ is identical to the result of applying the semantic operation $C^*$ to the expressions' semantic values:

$$(3) \qquad [\![C(E_1, \ldots, E_n)]\!] = C^*([\![E_1]\!], \ldots, [\![E_n]\!]).$$

How do the our two competing interpretations fare with respect to compositionality? It is easy to verify that, as set out above, both interpretations satisfy the constraint perfectly. However, it turns out that the reductionist interpretation is not at all robust in this regard. Some minor modifications of $\mathcal{L}_1$ cannot be handled compositionally by the reductionist interpretation or anything like it.

The problem arises in connection with generalized quantifiers.[15] The standard semantic treatment of 'most' has '(Most $x : \phi(x)(\psi(x))$' come out true just in case more than half of the objects in the domain that satisfy $\phi(x)$ also satisfy $\psi(x)$. Consider a situation involving five inscriptions, three of the letter 'A' and one of each of 'B' and 'C'. Then the following sentence is false:

(4) Most letters are vowels.

For of the three letters present, only one is a vowel. But by operating with a domain consisting of inscriptions rather than letters, the semantic reductionist cannot easily account for the falsity of (4). For according to the reductionist, the domain consists of five inscriptions, more than half of which *are* vowels. In order to hold on to the ordinary treatment of generalized quantifier such as 'most', we seem to be forced to operate at the level of letters rather than inscriptions.

The reductionist may respond that the contemplated extension of $\mathcal{L}_1$ has *two* versions of the generalized quantifier 'most': one for each of the two sorts. Where the 0-version behaves in the ordinary way, the 1-version behaves unusually in counting only 'up to equivalence'. But

[15] This problem is discussed by (Heck, 2000) in a slightly different context. As Heck observes, similar considerations apply to many other generalized quantifiers as well.

this response is unlikely to succeed. For one thing, it is unpleasantly *ad hoc*. For another, the argument against the reductionist can be recast in the context of a one-sorted language $\mathcal{L}_2$, where the two sorts of $\mathcal{L}_1$ have been merged into one. (See Appendix B for details.) In this one-sorted language there is no room for two different quantifiers 'most', and the reductionist interpretation can be convicted of an outright failure of compositionality.

In order to have an acceptable account of generalized quantifiers, we need the kind of 'coarse graining' of the domain $D_0$ that the non-reductionist obtains from his abstraction mapping $a \mapsto \overline{a}$ that sends an inscription to its abstract letter type. But this still leaves the reductionist some room to maneuver. All that the analysis of generalized quantifiers requires is that the mapping satisfy the abstraction principle ($\Sigma$). Provided this requirement is met, the mapping can do anything whatsoever. In particular, it can send any inscription to some chosen representative from its equivalence class. This raises the possibility that the reductionist can interpret the mapping in precisely this way and thus adapt the non-reductionist interpretation to suit her own purposes.

Again, the non-reductionist has the resources to strike back. Firstly, the choice of representatives would be completely arbitrary.[16] Secondly, as Gideon Rosen (1993) observes in a related context, the trick of choosing a distinct representative from each equivalence class isn't always be available. Assume for instance that the community extends their language, not by speaking as if every inscription determines a letter, but as if any plurality of objects of the 0-sort determine a unique set of the 1-sort. This is a perfectly coherent linguistic practice. The abstraction principle guiding the practice is a *two-sorted* version of Basic Law V, which, unlike its more familiar one-sorted cousin, is known to be consistent. But in this case there are as many equivalence classes as there are pluralities of objects from $D_0$. So by a version of Cantor's theorem, it is impossible to choose a distinct representative from $D_0$ for each equivalence class.

### 4.3 Cognitive constraints on an interpretation

I now turn to a cognitive constraint on the choice of a semantic interpretation. According to this constraint, the truth-condition that the interpretation assigns to a sentence must

[16]And it is not an option for the reductionist to let the mapping send each inscription to its own equivalence class under $\sim$. For this would invoke reference to abstract objects after all.

be one that is adequately grasped by speakers who understand the sentence. For instance, the standard truth-condition for 'Snow is white' satisfies the constraint. For in order to understand this sentence, a speaker must grasp the concepts of snow and whiteness and master the operation of predication; and together, these cognitive skills suffice to grasp the standard truth-condition. But in other cases, the constraint enables us to reject attempted semantic analyses which invoke concepts or principles of which competent speakers do not have an adequate grasp. I will illustrate this point by means of an example.

Imagine another community, this time one whose base language is so impoverished that it features no genuine reference to properly individuated objects. All that can be expressed are simple 'feature-placing' claims such as 'hot here' and 'wet there'. But later, perhaps over many generations, the community extends their language to one that functions precisely as if it is referring to and quantifying over physical bodies, such as sticks and stones. The extended language contains demonstratives such as 'this body' and 'that body'. On each occasion of use, a demonstrative of this sort is associated with a piece of matter. For instance, when a speaker utters 'this body', attending to a strick protruding from the ground, then the demonstrative is associated with the piece of matter with which the speaker perceptually interacts. One of the assertibility conditions that governs the extended language says that an identity of the form 'this body = that body' is assertible, relative to two pieces of matter, just in case the pieces of matter are related through a continuous stretch of solid stuff, all of which belongs to the same unit of independent motion (roughly in the sense that, if you wiggle one part, the other part follows along). Call this relation between pieces of matter $C$.[17]

What is the appropriate semantic interpretation of the extended language? Again, there are two competing alternatives. For present purposes, it suffices to consider their respective analyses of a simple identity of the form 'this body = that body', where the demonstratives are associated with two pieces of matter. According to the reductionist interpretation, the sentence should be interpreted as referring to the two pieces of matter and asserting that they stand in relation $C$. According to the competing non-reductionist interpretation, the sentence should be interpreted as referring to the two physical bodies picked out by the pieces of matter and asserting that these bodies are identical. The cognitive constraint may help us decide between these two interpretations. Assume that the speakers have no explicit understanding

[17]Readers who are dissatisfied with my characterization of $C$ are of course free to substitute their own.

of the relation $C$. Although they are generally able to tell whether two pieces of matter are $C$-related, this ability is based on various subpersonal mechanisms and low-level practical skills, not on a grasp of any concept of $C$-relatedness. If so, then the cognitive constraint gives us reason to favor the non-reductionist interpretation.

Let's return to our main example of the apparent 'letter-talk' of $\mathcal{L}_1$. Can the cognitive constraint be as useful here as it was in the example just considered? The question is whether the speakers have an adequate grasp of the relation $\sim$ and are guided by this grasp, or merely draw on a practical ability to tell whether two objects stand in the relation $\sim$. Since our example is an imagined one, we are free to stipulate the latter, in which case the cognitive constraint will favor the non-reductionist interpretation. (In fact, I believe this stipulation is a decent approximation of our ordinary practice of 'letter-talk'. But a defense of this belief would take us too far afield.)

## 5 Why a non-reductionist interpretation is available

In the previous section, I argued that, *if available*, a non-reductionist interpretation of $\mathcal{L}_1$ would be superior to its reductionist rival in two respects, having to do with compositionality and a cognitive constraint on how to draw the boundary between semantic principles and non-semantic ones. I will now argue that a non-reductionist interpretation is indeed available, despite its commitment to abstract letters. The argument will turn on a sense in which the use of a language is prior to semantic theorizing about it.

Consider again our community that wishes to extend their language from $\mathcal{L}_0$ to $\mathcal{L}_1$, used as described above. When looking at some inscriptions, it would be convenient to be able to ask each other, for instance, whether most of the letters are vowels, rather than having to resort to a more longwinded formulation involving counting only up to the equivalence $\sim$. So the community goes ahead and adopts the extended language, governed by the assertibility conditions. Is their new linguistic practice in good order? It is hard to see why not. The community isn't very sophisticated and take no interest in theoretical semantics. So they have no view on how the extended language is to be interpreted. What matters to them is simply to be able to use the extended language in accordance with the assertibility conditions. Given the unspecific nature of their demand and the fact that each sentence has an assertibility condition expressed in unproblematic terms that make no mention of directions or other abstract objects, it is hard to see how the linguistic practice could be flawed.[18]

Of course, things would be very different if the assertibility conditions of the extended language were subject to non-trivial presuppositions. Consider, for instance, a community of seventeenth-century chemists who wish to extend their language so as to talk about phlogiston. For instance, they wish to regard it as assertible of any process of combustion that phlogiston has been released. However, the community intends this assertion to be more than just a fancy way of saying that combustion is taking place. They intend the term 'phlogiston' to pick out a substance that plays a number of roles that are characterizable already in their base language. For instance, the substance is to play a crucial role in the explanation of combustion and to be something that can, at least in principle, be isolated and observed. The success of their new linguistic practice is thus premised on the existence of a substance that plays at least most of these roles. By contrast, the speakers of $\mathcal{L}_1$ intend nothing more than to use their new language in accordance with the mentioned assertibility conditions, which are stated in unproblematic terms and subject to no additional presuppositions.

What happens next? As the centuries go by, the speakers hold on to their expressive resources and add new ones in an analogous way. Gradually their sophistication increases, and they begin to take an interest in theoretical semantics. One day they ask how best to interpret their own earlier language $\mathcal{L}_1$—or, if you prefer, a qualitatively identical language spoken by another community that has also undergone the sort of development described in Section 2. In particular, our speakers wonder what objects, if any, the singular terms of $\mathcal{L}_1$ refer to. When discussing these questions, they use as their metalanguage their own language at the time in question. This will be some extension of $\mathcal{L}_1$, which we may call $\mathcal{L}_1^M$.[19] Crucially, this metalanguage will be one that functions precisely as if it is capable of referring to and quantifying over letters (as well as other kinds of objects). This means that both of the competing interpretations of $\mathcal{L}_1$ that we have discussed can be formulated and are available to the community. For the reasons set out in the previous section, the community finds the non-reductionist interpretation preferable. So they decide to adopt a non-reductionist interpretation of $\mathcal{L}_1$ that ascribes to its 1-terms genuine reference to abstract letters.

[18] This cannot be said about cases of impredicative abstraction; cf. footnote 12.

[19] Alternatively, they may use an extension of the one-sorted language $\mathcal{L}_2$, described in Appendix B, that results from merging the two sorts of $\mathcal{L}_1$. This possibility doesn't materially affect what follows.

A skeptic may challenge the community's right to use a metalanguage which, like $\mathcal{L}_1$, functions as if it is capable of referring to letters. Would it not be better to use a more modest metalanguage which, like $\mathcal{L}_0$, refrains from using potentially problematic linguistic resources such as terms that function as if they are referring to abstract objects? If such a modest metalanguage was used, however, the community would be unable to formulate the non-reductionist interpretation and hence would regard it as unavailable.

The skeptic's challenge can be answered. If it was permissible, centuries ago, for the community to adopt $\mathcal{L}_1$ as their main language, and if later extensions were permissible in an analogous way, why should it not also be permissible for the community to rely on the resulting language when doing science, in particular, when doing semantics? Still, the challenge points to something interesting. In order for an interpretation of some object language to be available in a metalanguage, the metalanguage needs to have sufficient expressive resources. For instance, in order to interpret an object language as referring to ordinary physical bodies, we need to be able to engage in precisely this sort of reference in the metalanguage. And the point generalizes. If a metalanguage cannot refer to $F$s, it cannot be used to interpret an object language as referring to $F$s. It follows that, if upheld, the skeptic's challenge could be used to undermine *any* ascription of reference. We should therefore reject it.

Let's take stock. Something remarkable has happened. A simple community has in an acceptable way extended their language so as to speak precisely as if there are abstract letter types. Later, while still relying on a language with this feature, they look back at what they have accomplished and reach the conclusion that they have succeeded in referring to abstract objects. The language that has been adopted is thus *stable under semantic reflection* in the sense that, once adopted, its speakers can look back and make good semantic sense of what they have done. Finally, because reference to $F$s cannot be ascribed to an object language without relying on such reference in the metalanguage, this form of stability under semantic reflection is the best we can hope for.

## 6 Metaphysically lightweight objects

The mentioned linguistic development has consequences for ontology as well. Having extended their language, the community regards it as correct to assert a variety of sentences that carry ontological commitment to abstract objects—or at least do so on a semantic analysis that takes the apparent structure of the language at face value. For instance, they assert '$\exists y_1 \, \textsc{Vow}(y_1)$'. Moreover, once the community engages in the mentioned semantic reflection, they ascribe truth, not just correctness, to the sentences in question. They also defend the mentioned semantic analysis and thus agree that the sentences in question are ontologically committed to abstract objects. Since all of this takes place in an acceptable manner, the community ends up correctly concluding that there are abstract objects.

This way of introducing objects into one's rational discourse will no doubt strike some philosophers as too easy. Surely, such philosophers will think, the view that there are abstract objects is a substantive thesis whose truth requires the cooperation of reality and not just the adoption of some language. I disagree. I believe that letters (and other abstract objects introduced in an analogous way) are *metaphysically lightweight*—in a sense that I will now attempt to spell out.[20]

Loosely characterized, the idea is that lightweight objects do not require much of reality for their existence. Their existence requires only the obtaining of some condition which does not mention the objects in question and which is thus comparatively unproblematic. For instance, the existence of a letter requires nothing more than that there be two equivalent inscriptions. The idea of metaphysically lightweight objects has long been attractive to philosophers of mathematics.[21] It is not hard to see why. If the existence of mathematical objects does not require very much, it is unsurprising that there should be so many of them. For the less that is required for the existence of object of some kind, the more such objects there will be. The idea of lightweight objects promises to help with the epistemology of mathematics as well. If it suffices for the existence of some mathematical objects that some condition is met, and if this condition is unproblematic enough to be within our ken, then this may explain how we obtain epistemological 'access' to the objects in question.

Of course, the loose characterization just outlined needs to be spelled out in appropriate detail. The most sustained attempt to do so is by the neo-Fregeans, who build on Frege's idea that matching instances of the two sides of an abstraction principle are just different

[20] For further references and discussion of the contents of the next three paragraphs, see (Linnebo, ).

[21] The neo-Fregeans provide one example. Another comes from a version of structuralism which holds that the existence of a mathematical structure requires nothing more than the coherence of the theory that categorically describes the structure. See for instance (Shapiro, 1997).

ways of 'carving up' one and the same content. Consider for instance the statements that the directions of the lines $l_1$ and $l_2$ are identical and that $l_1$ and $l_2$ are parallel. These statements are said to share a common content which they 'carve up' in different ways. If correct, this will provide a clear sense in which the existence of the directions requires nothing more than the parallelism of the lines.

Not surprisingly, the problem is to articulate the relevant notion of content. The contents in question must be coarse-grained enough to ensure that matching instances of the two sides of an abstraction principle have the same content. In particular, it must be possible for two sentences to have the same content although they refer to different objects. But simultaneously, the contents must be fine-grained enough to transmit epistemological status. For we want knowledge of (or justified belief in, or apriority of) an instance of the right-hand side of an abstraction principle to ensure knowledge of (or justified belief in, or apriority of) the corresponding instance of the left-hand side. It is unclear whether there is any notion of content that is capable of satisfying both demands.

This article makes available a different way of spelling out the idea of lightweight objects. As in the neo-Fregeans' case, the challenge is to explain the relation between matching instances of the two sides of an abstraction principle. But unlike the neo-Fregeans, I have no need for the idea that an instance of one side and a matching instance of the other share a common content. Indeed, on the semantic interpretation that I favor, such sentences have different truth-conditions and therefore different meanings.[22] On my view, the relation between matching instances of the two sides of a predicative abstraction principle has to do with assertibility conditions, not with any kind of shared semantic content. More precisely, it is permissible to extend one's language so as to include the instances of the left-hand side and such that each such instance has as its assertibility condition the corresponding instance of the right-hand side.

My view clearly has a reductionist aspect. Each sentence of the extended language $\mathcal{L}_1$ has an assertibility condition that makes no mention of letters or any other abstract objects. But unlike existing forms of reductionism, mine is not located at either the level of syntax or of semantics. Compare for instance the reductionist semantics discussed in Section 3, on which each sentence of $\mathcal{L}_1$ has a truth-condition that makes no mention of letters or other

[22]The only form of content that such sentences share is a truth-value. But truth-values are of course much too coarse-grained to transmit epistemological status.

abstract objects. On this view, it can be perfectly true to speak as if there are letters. But the apparent reference to letters is deceptive. All reference is really to inscriptions and other concrete objects. By contrast, since my reductionism is tied to the assertibility condition rather than to semantic truth-conditions, I can take the apparent reference to letters at face value. On their proper semantic interpretation, the sentences of $\mathcal{L}_1$ genuinely refer to letters.

The study of the facts in virtue of which linguistic expressions come to possess their semantic properties is often called *metasemantics*. Consider for instance my claim that the correct semantic interpretation of $\mathcal{L}_1$ is the non-reductionist one. There is an associated metasemantic question. What makes this interpretation the correct one? My answer is: the fact that the use of $\mathcal{L}_1$ is guided by the assertibility conditions. By locating the reductionism at the level of the assertibility conditions, my account may thus be called *metasemantic reductionism*. On this view, $\mathcal{L}_1$ genuinely refers to abstract letters. But what is responsible for this semantic truth are some simpler truths about the use of $\mathcal{L}_1$ which make no mention of letters or other abstract objects.

# Appendices

## A Assertibility conditions for $\mathcal{L}_1$

Recall that $\mathcal{L}_0$ has an interpretation $I$ that we assume as given. We work in a metalanguage $\mathcal{L}^M$, which for convenience we assume to be one-sorted. However, it would be straightforward to adapt what follows to the case where $\mathcal{L}^M$ has a separate sort for letters.

*Predicates.* For every $n$-place predicate $F$ there is a formula $\phi_F$ in $\mathcal{L}^M$ with $n$ free variables such that, for any $a_1, \ldots, a_n$ from $D_0$, $F$ is assertible of $a_1, \ldots, a_n$ iff $\phi_F(a_1, \ldots, a_n)$. We make the following further assumptions as well.

- If $F$ is in $\mathcal{L}_0$, then $\phi_F$ is simply the satisfaction condition that $I$ assigns to $F$.
- If the $i$'th argument place of $F$ is of type 1, then for any $a_1, \ldots, a_n$ and $b_i$ from $D_0$ we have:
$$a_i \sim b_i \rightarrow \big(\phi_F(a_1, \ldots, a_n) \leftrightarrow \phi_F(a_1, \ldots, a_n)[b_i/a_i]\big)$$
where $\psi[e'/e]$ is the formula that results from $\psi$ by a uniform substitution of $e'$ for $e$.
- If $F$ is $=_1$, then $\phi_F$ is $\sim$.

*Singular terms.* So far, we have talked about a formula $\phi$ being assertible of a finite sequence of objects. As usual, it will be convenient to state some of these claims in terms of the notion a variable assignment. So let $\sigma$ be an assignment of an element of $D_0$ to each variable of $\mathcal{L}_1$. Relative to $\sigma$, every singular constant $t$ of $\mathcal{L}_1$ is associated with an element $a$ of $D_0$ subject to the following constraints.

- If $t$ is a constant of sort 0, then $a$ is the referent of $t$ under $I$.
- If $t$ is a variable, then $a$ is what $\sigma$ assigns to $t$.
- If $t$ is of the form $\S s$ for a singular term $s$, and if $s$ is associated with $a$ relative to $\sigma$, then $t$ too is associated with $a$ relative to $\sigma$.

*Formulas.* The assertibility conditions for formulas of $\mathcal{L}_1$ are now as follows.

- Let $\phi$ be an atomic formula $F(t_1, \ldots, t_n)$, where each singular term $t_i$ is associated with an element $a_i$ of $D_0$ relative to $\sigma$. Then $\phi$ is assertible relative to $\sigma$ iff $\phi_F(a_1, \ldots, a_n)$.
- The clauses for truth-functional connectives are the obvious ones.
- Let $\phi$ be of the form $\forall v\, \psi$, where $v$ is a variable of either sort. Then $\phi$ is assertible relative to $\sigma$ iff for every variable assignment $\tau$ that differs from $\sigma$ at most in its assignment to $v$, $\psi$ is assertible relative to $\tau$.

## B Assertibility conditions for $\mathcal{L}_2$

The formal language $\mathcal{L}_2$ is obtained from $\mathcal{L}_1$ by deleting all type indices and ignoring all restrictions on well-formedness having to do with such indices.

When stating the assertibility conditions governing $\mathcal{L}_2$, we will make use of a domain $D_l$ obtained from $D_0$ by attaching to each element of $D_0$ a type index as a label. Formally, we let $D_l = D_0 \times \{\text{`0'}\} \cup S \times \{\text{`1'}\}$, where $S$ is the subset of $D_0$ consisting of inscriptions. Intuitively, the idea is that the elements of $D_l$ serve as *presentations* of objects. A presentation of the form $\langle a, \text{`0'}\rangle$ is an *ordinary presentation*, which we may think of as simply presenting the object $a$. A presentation of the form $\langle a, \text{`1'}\rangle$ is a *letter presentation*, which we may think of as presenting the abstract letter type of $a$.

*Predicates.* For every $n$-place predicate $F$ there is a formula $\phi_F$ in $\mathcal{L}^M$ with $n$ free variables such that, for any $c_1, \ldots, c_n$ from $D_l$, $F$ is assertible of $c_1, \ldots, c_n$ iff $\phi_F(c_1, \ldots, c_n)$. We make the following further assumptions as well.

- If $F$ is in $\mathcal{L}_0$ and $\psi_F$ is the satisfaction condition that $I$ assigns to $F$, then for any $\langle a_i, l_i\rangle$ from $D_l$ we have:

$$\phi_F(\langle a_1, l_1\rangle, \ldots, \langle a_n, l_n\rangle) \longleftrightarrow l_1 = \ldots = l_n = 0 \wedge \psi_F(a_1, \ldots, a_n)$$

[Intuitively, any predicate of $\mathcal{L}_0$ is regarded as assertible only of objects from $D_0$.]

- For any elements $a_1, \ldots, a_n$ and $b_i$ of $D_0$ and any labels $l_0, \ldots, l_n \in \{`0\text{'}, `1\text{'}\}$, we have:

$$l_i = `1\text{'} \wedge a_i \sim b_i \rightarrow \big(\phi_F(\ldots) \leftrightarrow \phi_F(\ldots)[\langle b_i, `1\text{'}\rangle \langle a_i, `1\text{'}\rangle]\big)$$

  where '$\ldots$' abbreviates the string '$\langle a_1, l_1\rangle, \ldots, \langle a_n, l_n\rangle$'.

  [Intuitively, if $F$ is regarded as assertible of a string of presentations, the $i$'th of which is a letter presentation, then $F$ is also regarded as assertible of any other string of presentations that presents the same objects, that is, where the first coordinate of any letter presentation may be replaced by any $\sim$-equivalent inscription.]

- Let $\phi_=$ be the condition associated with the identity predicate $=$. Then:

$$\phi_=(\langle a_1, l_1\rangle, \langle a_2, l_2\rangle) \leftrightarrow \big((l_1 = l_2 = `0\text{'} \wedge a_1 = a_2) \vee (l_1 = l_2 = `1\text{'} \wedge a_1 \sim a_2)\big)$$

  [Intuitively, two presentations are regarded as presenting identical objects just in case both are ordinary presentations with identical first coordinates, or both are letter presentations with $\sim$-equivalent first coordinates.[23]]

*Singular terms.* Let $\sigma$ be an assignment of elements of $D_l$ to each variable of $\mathcal{L}_2$. Relative to $\sigma$, every singular constant $t$ of $\mathcal{L}_2$ is associated with an element $c$ of $D_l$ subject to the following constraints.

- If $t$ is a constant in $\mathcal{L}_0$, then $c = \langle a, `0\text{'}\rangle$ where $a$ is the referent of $t$ under $I$.
- If $t$ is a variable, then $c$ is what $\sigma$ assigns to $t$.
- If $t$ is of the form $\S s$ for a singular term $s$, and if $s$ is associated with $\langle a, l\rangle$ relative to $\sigma$, then $t$ is associated with $\langle a, `1\text{'}\rangle$ relative to $\sigma$.[24]

*Formulas.* The assertibility conditions for formulas of $\mathcal{L}_2$ are now as follows.

- Let $\phi$ be an atomic formula $F(t_1, \ldots, t_n)$, where each singular term $t_i$ is associated with an element $c_i$ of $D_l$ relative to $\sigma$. Then $\phi$ is assertible relative to $\sigma$ iff $\phi_F(c_1, \ldots, c_n)$.
- The clauses for truth-functional connectives are the obvious ones.
- Let $\phi$ be of the form $\forall v\, \psi$, where $v$ is a variable. Then $\phi$ is assertible relative to $\sigma$ iff for every variable assignment $\tau$ that differs from $\sigma$ at most in its assignment to $v$, $\psi$ is assertible relative to $\tau$.

[23] This way of handling 'mixed' identities—and thus of resolving the Caesar problem—strikes me as one of the more plausible. However, what follows could easily be adapted to any other formally acceptable way of resolving the Caesar problem.

[24] What matters is of course the case where $l = `0\text{'}$. Allowing the operation to be defined where $l = `1\text{'}$ is merely a matter of convenience, corresponding to letting the letter of any letter be that very letter.

## C Comparing the two interpretations of $\mathcal{L}_1$

Two semantic interpretations of $\mathcal{L}_1$ are explained in the main text: a reductionist interpretation $R$, which is modeled on the assertibility conditions for $\mathcal{L}_1$; and a non-reductionist interpretation $N$, which interprets letter terms as actually referring to and ranging over letters, obtained by abstraction on inscriptions under $\sim$. When talking about a variable assignment $\sigma$ in connection with $N$, we always leave it tacitly understood that $\sigma$ assigns elements of $D_i$ to $i$-variables for $i = 0, 1$, and *mutatis mutandis* for $R$. We write $R \models_\sigma \phi$ for the claim that $\phi$ is satisfied by the variable assignment $\sigma$ on $R$, and likewise for $N$.

**Proposition 1** Let $\phi$ be any formula of $\mathcal{L}_1$ and $\sigma$ any assignment of elements of $D_0$ to its variables. Then:

$$\phi \text{ is assertible relative to } \sigma \quad \text{iff} \quad R \models_\sigma \phi.$$

*Proof.* An obvious induction on syntactic complexity. $\dashv$

Recall the two 'bridge principles' which relate the domains and interpretations of predicates used by the two interpretations.

$$\text{(B1)} \qquad b \in D_1 \leftrightarrow \exists a(a \in D_0 \wedge b = \overline{a})$$

$$\text{(B2)} \qquad \forall a_1, \ldots, a_n \in D_0 \; [\overline{\phi_F}(\overline{a_1}, \ldots, \overline{a_n}) \leftrightarrow \phi_F(a_1, \ldots, a_n)]$$

**Proposition 2** Let $\sigma$ be an assignment of elements of $D_0$ to all the variables of $\mathcal{L}_1$. Let $\tau$ be assignment of elements of $D_0$ to the 0-variables and of elements of $D_1$ to the 1-variables. Assume that for each $i$, we have $\sigma(x_i) = \tau(x_i)$ and $\sigma(y_i) = \overline{\tau(y_i)}$. Then (B1) and (B2) entail that, for any formula $\phi$ of $\mathcal{L}_1$, we have:

$$R \models_\sigma \phi \quad \text{iff} \quad N \models_\tau \phi$$

*Proof.* We begin with the base-case where $\phi$ is atomic. Then (B2) and the definitions of $R$ and $N$ ensure that the equivalence holds for $\phi$. When the main logical operator in $\phi$ is a connective, the induction step is trivial. The only non-trivial induction step is for the universal quantifier. (Recall that '$\exists$' is regarded as an abbreviation of '$\neg\forall\neg$'.) So consider the case where $\phi$ is $\forall v\, \psi$. We prove only the left-to-right direction of the equivalence; the other direction is analogous. So assume that $R \models_\sigma \phi$. We want to show that $N \models_\tau \phi$. By the definition of $N$, it suffices to show that $N \models_{\tau'} \psi$ for any variable assignment $\tau'$ that differs from $\tau$ at most in its assignment to $v$. Assume $v$ is a 1-variable; the case of 0-variables is analogous but easier. Let $\tau'$ be like $\tau$ except that it assigns $b$ to $v$. By (B1) there is an $a \in D_0$ such that $b = \overline{a}$. Let $\sigma'$ be a variables assignment which is like $\sigma$ except that it assigns $a$ to $v$. Then our assumption and the definition of $R$ ensure that $R \models_{\sigma'} \psi$, whence the induction hypothesis yields $N \models_{\tau'} \psi$, as desired. $\dashv$

The following corollary is now immediate.

**Corollary 1** Assume that $\sigma$ and $\tau$ are variable assignments as in Proposition 2. Then (B1) and (B2) entail that, for any formula $\phi$ of $\mathcal{L}_1$, we have:

$$\phi \text{ is assertible relative to } \sigma \quad \text{iff} \quad N \models_\tau \phi$$

This corollary makes good on the claim that the assertibility conditions are such that the language is used precisely as if it was talking about abstract letters.

## References


Dummett, M. (1956). Nominalism. *Philosophical Review*, LXV:491–505. repr. in (Dummett, 1978).

Dummett, M. (1978). *Truth and Other Enigmas*. Harvard University Press, Cambridge, MA.

Dummett, M. (1981). *Frege: Philosophy of Language*. Harvard University Press, Cambridge, MA, second edition.

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

Enderton, H. B. (2001). *A Mathematical Introduction to Logic*. Academic Press, San Diego, second edition.

Frege, G. (1953). *Foundations of Arithmetic*. Blackwell, Oxford. Transl. by J.L. Austin.

Hale, B. (1987). *Abstract Objects*. Blackwell, Oxford.

Hale, B. (1997). *Grundlagen* §64. *Proceedings of the Aristotelian Society*, 97(3):243–61. Reprinted with a postscript in (Hale and Wright, 2001a).

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

Hale, B. and Wright, C. (2001b). To Bury Caesar .... In (Hale and Wright, 2001a).

Heck, R. (2000). Syntactic reductionism. *Philosophia Mathematica*, 8(2):124–149.

Hodes, H. (1990). Ontological commitment thick and thin. In Boolos, G., editor, *Meaning and Method: Essays in Honor of Hilary Putnam*, pages 235–260. Cambridge University Press, Cambridge.

Linnebo, Ø. Metaontological minimalism. Forthcoming in *Philosophy Compass*.

Rosen, G. (1993). The refutation of nominalism (?). *Philosophical Topics*, 21(2):149–186.

Rumfitt, I. (2003). Singular terms and arithmetical logicism. *Philosophical Books*, 44(3):193–219.

Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford University Press, Oxford.

Wright, C. (1983). *Frege's Conception of Numbers as Objects*. Aberdeen University Press, Aberdeen.