Notre Dame Journal of Formal Logic
Volume 38, Number 3, Summer 1997

# Grundgesetze der Arithmetik I §§29–32

RICHARD G. HECK, JNR.

**Abstract** Frege's intention in section 31 of *Grundgesetze* is to show that every well-formed expression in his formal system denotes. But it has been obscure why he wants to do this and how he intends to do it. It is argued here that, in large part, Frege's purpose is to show that the smooth breathing, from which names of value-ranges are formed, denotes; that his proof that his other primitive expressions denote is sound and anticipates Tarski's theory of truth; and that the proof that the smooth breathing denotes, while flawed, rests upon an idea now familiar from the completeness proof for first-order logic. The main work of the paper consists in defending a new understanding of the semantics Frege offers for the quantifiers: one which is objectual, but which does not make use of the notion of an assignment to a free variable.

***1 Opening*** In sections 30–31 of the first volume of *Grundgesetze der Arithmetik*, Frege attempts to show that every well-formed expression of his formal language, Begriffsschrift, denotes.[1] Although there has been a fair bit of discussion of these passages, it remains unclear how he intends to do this, why he thinks he needs to do it, and upon what assumptions his argument depends. No real consensus has been reached about the most difficult parts of the argument: and, in so far as there is an agreed view about other parts, it is, I think, mostly mistaken, resting upon a confusion about the nature of the semantic theory Frege offers for Begriffsschrift. Once that has been cleared up, it will be apparent that Frege's argument is not the complete mess it is often thought to be. In fact, we shall see that there are respects in which it anticipates Tarski's theory of truth, and others in which it constitutes an *alternative* to it. With only a modicum of anachronism, we can understand Frege to have formulated an informal theory of truth for (a significant fragment of) Begriffsschrift and informally to have proved its adequacy, in something like Tarski's sense.

Three things about §§30–31 are relatively uncontroversial: first, that Frege is trying to prove that every well-formed expression denotes uniquely; secondly, that the proof proceeds by induction on the complexity of expressions; and, thirdly, that the

*December 6, 1996; revised August 18, 1997*

proof must be fallacious. There are those who would reject the first of these claims, namely, those who hold that Frege could have made no serious use of semantical notions:[2] for if not, he presumably could not have been trying to prove a semantical claim in §§30–31, could not have been using semantical notions in that proof. But I shall not address this sort of question here and shall simply assume that Frege does intend to offer some sort of proof. Once that is accepted, there is no real option but to understand the proof as a proof by induction. Finally, Russell's Paradox would appear to show that not every well-formed expression of Begriffsschrift *can* have been given a unique denotation, whence there must be something wrong with the proof.

The question is how the proof is supposed to go. It is clear that Frege intends, first, to show that the *primitive* expressions of his system denote and then to conclude from this that *every* well-formed expression denotes, by arguing that any expression formed from denoting expressions denotes. Beyond this, there is not much more agreement. But whatever their differences, all interpretations known to me agree about two further things, which are interdependent, as we shall see below: first, that there is *another* induction in the proof, contained in Frege's argument that the smooth breathing—from which names of value-ranges are formed—denotes; and, secondly, that his argument that the (first- and second-order) universal quantifiers denote depends upon his treating them substitutionally, as if they were (infinite) conjunctions of their instances.[3]

So read, Frege's proof fails because second-order quantification in Begriffsschrift is impredicative. One cannot treat '$\forall F.\Phi_x(Fx)$' as the conjunction of its instances, for we can form an instance of '$\forall F.\Phi_x(Fx)$' by instantiating the bound variable '$F$' with *any* complex predicate (subject to the usual restrictions): we could, for example, instantiate '$F\xi$' with '$\forall G \exists R \forall y(Ry\xi \longrightarrow G\xi)$'.[4] The sort of induction this interpretation reads Frege as offering will therefore fail: instantiation does not, in general, lead from a sentence containing, say, $n$ second-order quantifiers to one containing fewer than $n$. Moreover, we now know, not only that the first-order fragment of Frege's system is consistent, but also that various predicative second-order fragments are (see Parsons [21] and Heck [16]). It is thus plausible that the flaw in the proof should relate to the presence of impredicative comprehension in the system. Whether the proof actually works for the first-order fragment will depend upon how, exactly, Frege treats the case of the smooth breathing: similar problems may infect his argument that the smooth breathing denotes, or they may not.

As we shall see below, there are reasons to endorse this sort of interpretation. But, even at first glance, it is apparent that there are serious problems with it. As Frege explains the second-order quantifier, a sentence of the form '$\forall F.M_x(Fx)$' denotes the True if and only if the function denoted by '$M_x(\varphi x)$' has the value True for every argument of the appropriate type ([7], I:24), that is, if every first-level function of one argument falls under '$M_x(\varphi x)$'. On the sort of interpretation under discussion, Frege is either ignoring or contradicting this explanation: it does not follow from the fact that every *instance* of '$M_x(\varphi x)$' is true that $M_x(\varphi x)$ has the value True for every argument. At best, it follows only if every function in the domain has a name. But Frege does not argue for this latter, somewhat implausible claim, and there is no independent reason to think he believed it. Of course, it is possible that Frege simply overlooked something here: but any interpretation of the proof which does not need to make that claim has a definite advantage over the usual ones.

***2 The character of the induction*** At the beginning of §31, Frege writes:

> Let us apply the foregoing in order to show that the proper names, and names of first-level functions, which we can form in these ways out of our simple names introduced up to now, always have a denotation. By what has been said, it suffices to show that our primitive names denote something.

By "the foregoing," Frege means §30. Thus, the induction proceeds by showing that the simple names[5] denote and that any complex name formed from denoting names also denotes (see the first sentence of [7], I:32). The permissible methods of formation are discussed in §30, which has the title "Two ways to form a name" and which provides an account of the formation-rules of the system, that is, of its syntax. But this is not all that is done in §30: for as Frege indicates in the passage quoted, it is because of "what has been said" there that "it suffices to show that our primitive names denote something." Thus, §30 also contains the inductive part of the argument for the conclusion stated at the beginning of §31, which itself contains only the proof of the basis case.

As Frege says, there are two ways to form a name in Begriffsschrift. To form an expression in the first way, one fills one argument-place of a functional expression with an argument of the appropriate type. There are four methods for doing this.

1. One can fill the sole argument-place of a one-place, first-level functional expression with a proper name—for example, fill the argument-place of '$\Phi(\xi)$' with some term '$t$', to get '$\Phi(t)$'.
2. One can fill the sole argument-place of a second-level functional expression with a first-level functional expression of appropriate type—for example, fill the argument-place of the first-order universal quantifier with a one-place, first-level predicate '$\Phi\xi$', to get '$\forall x.\Phi(x)$'.
3. One can fill the sole argument-place of a third-level functional expression with a second-level functional expression of appropriate type—for example, fill the argument-place of the second-order universal quantifier with a one-place, second-level predicate '$M_x(\varphi x)$', to get '$\forall F.M_x(Fx)$'.

In each of these cases, the result is a proper name (which may be a sentence, since sentences, in the language of the Begriffsschrift, are a special sort of name).

4. One can fill *one* of the two argument-places of a two-place, first-level functional expression with a proper name—for example, fill the $\xi$-argument-place of '$\xi = \eta$' with a term '$t$', to get '$t = \eta$'.

In this case, the result is a one-place, first-level functional expression, from which one could go on to form a proper name in accord with method 1 or 2.

The second way is more complicated. To see the need for it, note that one cannot, simply by filling argument-places, form the sentence '$\forall x.x = x$'. The universal quantifier is, as Frege understands it, a one-place, second-level functional expression, so, to form the sentence '$\forall x.x = x$', we must fill the argument-place of the quantifier with the functional expression '$\xi = \xi$': but this expression is not one of the primitive expressions of Frege's system, although the two-place predicate '$\xi = \eta$' is; nor is there any way to form '$\xi = \xi$' by means of the four methods comprising the first way. The only method allowing us to form *anything* from a two-place functional expression is the fourth, in which we fill one of its argument-places with a name, and that will only yield something of the form '$t = \eta$' or '$\xi = t$'. Similarly, it is not clear how we could form the sentence '$\forall x \forall y . x = y$'. To do so, we need to fill the argument-place of the quantifier with the functional expression '$\forall y . \xi = y$'. We could form *this* expression if we were allowed to put a two-place, first-level functional expression in the argument-place of the quantifier, leaving one argument-place unfilled: but, while Frege has provided for something like this in the case of filling one argument-place of a two-place functional expression with a name, he has made no such provision for this case. Clearly, then, some additional method of formation is required.

What Frege allows us to do is to form what he calls a 'composite' predicate (but is more usually called a 'complex' predicate) by removing one or more occurrences of a proper name from another in which it occurs, leaving behind an argument-place (which may have any number of distinct occurrences): "We begin by forming a name in the first way, and then exclude from it, at some or all places, a proper name that is a part of it (or coincides with it entirely) but in such a way that these places remain recognizable as argument-places . . . ." Thus, to use Frege's own example, we can form the sentence '$\Delta = \Delta$' by successively filling the two argument-places of '$\xi = \eta$' with the same term '$\Delta$'—that is, we use method 4 to form '$\Delta = \eta$', and then method 1 to form '$\Delta = \Delta$'. We can then, in accord with the second way, remove both occurrences of the term '$\Delta$', leaving an argument-place in its wake, thus forming the functional expression '$\xi = \xi$', which we can use, in accord with method 2, to form '$\forall x . x = x$'.

Frege is not as careful as he should have been here. He does remark, as we shall see below, that this method of formation applies not only to names formed "in the first way," but more generally: we will need to apply it to names formed by some combination of the first and second ways.[6] But Frege does not mention that we will need to be able to form, not just composite *first-level* predicates in this way—by excluding one or more occurrences of a proper name—but also composite *second-level* predicates—by removing occurrences of a first-level predicate. Thus, to form '$\forall F \forall x (Fx \longrightarrow Fx)$', we need first to form a sentence of the form '$\forall x (Gx \longrightarrow Gx)$', and then remove the two occurrences of the predicate '$G\xi$', thereby forming the second-level predicate '$\forall x (\varphi x \longrightarrow \varphi x)$', with which we can then fill the argument-place of the quantifier, in accord with method 3. These inaccuracies do not, as we shall see, invalidate his proof of the induction step, but the second does lead to one of the two errors in his argument in §31.

It is these "two ways of forming a name" that Frege needs to show preserve referentiality. His proof of this fact is intermingled with his introduction of them. To understand that proof, however, we need to understand what Frege thinks he must show in order to show that an expression denotes. And to understand that, we must turn to §29, in which Frege "answer[s] the question, When does a name denote something?" The conditions laid out fall into two groups: conditions relating to functional expressions and a condition relating to proper names. The latter condition becomes important only later: for now, we shall suppose it understood and consider only the conditions relating to functional expressions. These are all similar in form to that for one-place, first-level functional expressions:

> A name of a first-level function of one argument has a *denotation* . . . if the proper name that results from this function-name, when its argument-places are filled by a proper name, always has a denotation, whenever the name substituted has a denotation. ([7], I:29)

And a functional expression of some other type has a denotation if any proper name formed by filling its argument-places with denoting expressions of appropriate types has a denotation.

There is a way of reading this condition that can make the usual interpretation of Frege's argument seem mandatory. One of the claims for which Frege will be arguing is that the universal quantifier denotes: for it to denote is, by the relevant analogue of the condition quoted, for it to follow from the fact that the functional expression '$\Phi(\xi)$' denotes that '$\forall x.\Phi(x)$' denotes. But the condition quoted above regarding when '$\Phi(\xi)$' denotes appears to be that all of its instances do, an impression reinforced by the argument that the universal quantifier denotes:

> Now, '$\Phi(\xi)$' has a denotation if, for every denoting proper name '$\Delta$', '$\Phi(\Delta)$' denotes something. If this is the case, then this denotation either always is the True (whatever '$\Delta$' denotes), or not always. In the first case '$\forall x.\Phi(x)$' denotes the True; in the second, the False. Thus it follows universally from the fact that the substituted function-name '$\Phi(\xi)$' denotes something, that '$\forall x.\Phi(x)$' denotes something. ([7], I:31)

What Frege *seems* to be arguing here is that it follows from the fact that every *instance* of '$\forall x.\Phi(x)$' denotes that it too denotes. But it is unclear how that could be, if '$\forall x.\Phi(x)$' were not being treated as the (infinite) conjunction of its instances (substitutionally, in effect).

But this interpretation of Frege's argument rests upon a reading of §29 that has its own problems: so read, the condition for a functional expression to denote is inconsistent with one of Frege's most famous views about reference, a doctrine emphasized in the preceding section of *Grundgesetze* ([7], I:28). Infamously, Frege holds that a function-sign denotes only if it has a value for *every* argument: for example, he insists that '$\xi + 1$' has not been given a denotation unless it has been decided what its value is to be for the sun as argument, that is, what the sun plus one is (see, for example, Frege [6], pp. 19–20). That the result of substituting any *name* for '$\xi$' has a denotation shows only that the function-sign has been given a value for every object for which Begriffsschrift has a name—and that is not obviously the same. Hence, either Frege is flatly contradicting himself, or he is assuming that every object has a name in Begriffsschrift,[7] or the usual way of reading these conditions is incorrect.

The first option is not very palatable, and there are problems with the second, too. I have already said that there is no independent reason to think that Frege believed that every object has a name. And the following passage, taken from the introduction to the argument that the smooth breathing denotes, gives us positive reason to think he did not:

> The matter is less simple in this case, for with this we are introducing not merely a new function-name, but simultaneously answering to every name of a first-level function of one argument, a new proper name (value-range name); in fact not just for those [function-names] known already, but for all such that may be introduced in the future. ([7], I:31)

Frege is here explicitly denying that the only value-ranges present in the domain of his theory are those corresponding to function-names present in Begriffsschrift: value-ranges of functions whose names may only "be introduced in the future" are supposed already to be included.[8]

It would thus be nice if there were a way of reading the condition stated in §29 which did not commit Frege to the claim that every object has a name in Begriffsschrift. And so there is: the condition does indeed say that '$\Phi(\xi)$' denotes if and only if every sentence of the form '$\Phi(\Delta)$' denotes, so long as '$\Delta$' does, no matter what it might denote, *but no assumption is made that '$\Delta$' is itself a name formed from the primitive expressions of Begriffsschrift*. To read the condition in this way is to take Frege to be talking about filling argument-places, not with *actual* expressions, but with *auxiliary* expressions. The term '$\Delta$' is not supposed to be a name in Begriffsschrift *at all*: it is a formal device, a *new* name, added to the language, subject only to the condition that it should denote some object in the domain. And if we read Frege's conditions in this way, they say precisely what we would have thought they should say: to say that '$\Phi(\Delta)$' denotes so long as '$\Delta$' denotes, no matter what it denotes, is precisely to say that '$\Phi(\xi)$' has a value for every argument.

Why, if this is the right interpretation of Frege's conditions, does he state them as he does? Actually, Frege's statement of the conditions really is not all that peculiar. If I were going to state such a condition in Tarskian style, it is hard to see that I could do much better than this:

> A (possibly complex) predicate '$\Phi(\xi)$' has an extension if and only if, the open sentence '$\Phi(v)$' ('$v$' new) has a truth-value, whatever might be assigned to '$v$' (its extension being the set of objects whose assignment to '$v$' makes '$\Phi(v)$' true).

Where I speak of assignments to a free variable, Frege speaks of an auxiliary name assumed only to denote some object. It is, therefore, tempting to say that Frege states the condition as he does because he lacks the notion of an assignment—and so to recommend reading such passages charitably, as if Frege's talk of instances were but a poor approximation to Tarski's talk of assignments. For present purposes, this interpretation would probably do, but it is at best uncharitable. Frege's talk of the truth of instances formed using auxiliary names is not an approximation but an *alternative* to Tarski's talk of satisfaction by sequences. Fregean theories of truth are not particularly well known to philosophers, but they have assumed a central place in linguistics, in particular, in theories of anaphora.[9]

My suggestion, then, is that Frege's conditions for functional expressions to denote are exactly what one would expect, but that this has been obscured by a failure to recognize the semantical alternative he is offering us. Evaluating this interpretation of §29 is no easy matter, however, as there are few texts which bear directly upon the issue.[10] In order to provide some additional support for my interpretation then, I am going to discuss Frege's treatment of free variables. If his talk of instances really is an alternative to talk of assignments, we would expect that, in other contexts in which we would now speak of assignments to free variables, Frege should once again speak of instances formed using auxiliary names.

***3 Frege on free variables*** According to Frege, "in the case of a *Roman letter* [free variable] the *scope* shall comprise everything that occurs in the proposition with the exception of the judgment-stroke" ([7], I:17). That is, if a Roman letter occurs in a proposition,[11] say '$\vdash \Phi(x)$', then the scope of the letter comprises the whole of the formula '$-\Phi(x)$'. Roman letters might thus seem to be, for all intents and purposes, tacitly bound by initial (universal) quantifiers; for this reason, it is often said that free variables do not really occur in Frege's system. Of course, as far as the *validity* of a formula is concerned, free variables do indeed act as if they were tacitly bound: an open formula is valid if and only if its universal closure is.

Why, then, does Frege introduce Roman letters into his system at all? He is quite explicit about the reason:

> From the two sentences,
>
> 'All square roots of 1 are fourth roots of 1'
>
> and
>
> 'All fourth roots of 1 are eighth roots of 1'
>
> we can infer
>
> 'All square roots of 1 are eighth roots of 1'
>
> Now if we write the premises in this way:
>
> '$\vdash \forall x(x^2 = 1 \longrightarrow x^4 = 1)$' and '$\vdash \forall x(x^4 = 1 \longrightarrow x^8 = 1)$'
>
> then we cannot apply our methods of inference. We can, however, if we write them thus:
>
> '$\vdash a^2 = 1 \longrightarrow a^4 = 1$' and '$\vdash a^4 = 1 \longrightarrow a^8 = 1$'
>
> Here we have the case of §15 [i.e., transitivity for the conditional]. ([7], I:17)

The use of Roman letters thus brings certain sorts of inferences within the purview of the rules of inference as Frege states them. But note how he speaks here of rewriting the quantified premises, almost as if he regarded the use of Roman letters as a mere notational trick. The common wisdom is that this is all Frege's use of Roman letters amounts to, a convention allowing the omission of initial universal quantifiers.

But Frege sometimes expresses a different view, for example, in the following remarks, taken from his discussion of Peano's formalism. After remarking that he could have used but one style of letter, say, Roman letters, for both free and bound variables, he goes on to say that[12]

> from the point of view of inference, generality which extends over the scope of the entire proposition is of *vitally different significance* from that whose scope constitutes only a part of the sentence. Hence it contributes substantially to perspicuity that the eye discerns these different roles in the different sorts of letters, Roman and German. (Italics mine) (Frege [8], p. 248)

The "vitally significant difference" thus made visible is connected with the role of generality in inference: and it is this that is at issue in the passage from §17 quoted in the last paragraph. But what Frege there goes on to say about the role of Roman letters in inference is perplexing. He writes:

> Our stipulation regarding the *scope* of a *Roman letter* is to set only a lower bound upon the scope, not an upper bound. Thus it remains permissible to extend such a scope over several propositions, and this renders the Roman letters suitable to do duty in inferences, which the Gothic letters, with the strict closure of their scopes, cannot. If we have the premises '$\vdash a^2 = 1 \longrightarrow a^4 = 1$' and '$\vdash a^4 = 1 \longrightarrow a^8 = 1$' and infer the proposition '$\vdash a^2 = 1 \longrightarrow a^8 = 1$', in making the transition we extend the scope of the '$a$' over both of the premises and the conclusion, in order to perform the inference, although each of the propositions still holds good apart from this extention. ([7], I:17)

Frege appears, at first sight, just to be talking nonsense. How can the scope of a variable comprise multiple sentences?[13] Surely his thought is not that something like

$$\forall x \left[ \frac{\begin{array}{c} \vdash x^2 = 1 \longrightarrow x^4 = 1 \\ \vdash x^4 = 1 \longrightarrow x^8 = 1 \end{array}}{\vdash x^2 = 1 \longrightarrow x^8 = 1} \right]$$

is well-formed! But, as we shall see, on closer examination, Frege turns out not only to be talking sense, but to be giving expression to an insight which would not be fully understood for another forty years, until the work of Tarski.

The passage just quoted does not concern the formalization of the system, that is, how its rules are to be stated: it concerns what *justifies* the inference Frege is discussing. And once we ask how he thinks inferences involving Roman letters are to be justified, all begins to fall into place: Frege wants such inferences to be justified by *precisely* what justifies inferences of the same form in which Roman letters do *not* occur. Yet this view, as natural as it might seem, and as attractive as it obviously is, raises certain problems. In the case of the inference under discussion above, for example, transitivity for the conditional, Frege's justification of it reads as follows:

> From the two propositions '$\vdash \Delta \longrightarrow \Gamma$' and '$\vdash \Theta \longrightarrow \Delta$' we may infer the proposition '$\vdash \Theta \longrightarrow \Gamma$'. For $\Theta \longrightarrow \Gamma$ is the False only if $\Theta$ is the True and $\Gamma$ is not the True. But if $\Theta$ is the True, then $\Delta$ too must be the True, for otherwise $\Theta \longrightarrow \Delta$ would be the False. But if $\Delta$ is the True then if $\Gamma$ were not the True then $\Delta \longrightarrow \Gamma$ would be the False. Hence the case in which $\Theta \longrightarrow \Gamma$ is the False cannot arise; and $\Theta \longrightarrow \Gamma$ is the True. ([7], I:15)

This is essentially a justification in terms of truth-tables, and Frege's intention is that it should apply as much in the case in which '$\Gamma$', '$\Delta$', and '$\Theta$' contain Roman letters as it does when they do not (as much when they are "Roman marks" as when they are "proper names"). But now comes the problem. The justification simply does not apply when the propositions in question contain Roman letters, for the simple reason that '$a^2 = 1$' *has no truth-value* (at least when occurring in the proposition '$\vdash a^2 = 1 \longrightarrow a^4 = 1$'): it does not denote a truth-value, but only 'indicates' one.

This problem is, of course, resolved by the standard semantic theories for quantificational languages which Tarski introduced. What we need here is the notion of a *simultaneous assignment* of objects to free variables in different propositions. Thus, for example, the inference from '$\Phi(a) \longrightarrow \Psi(a)$' and '$\Psi(a) \longrightarrow \Pi(a)$' to '$\Phi(a) \longrightarrow \Pi(a)$' is valid, on the Tarskian account, just in case the last sentence is true under a particular assignment whenever the first two are true under the *same* assignment. It is this idea of simultaneity that Frege is trying to express when he says that the scope of a Roman letter is to be extended "over both of the premises and the conclusion": the letter is to indicate the *same* object in each of its occurrences in the three propositions.

That, however, does not speak to the question of how Frege treats the notion of indication itself, that is, how he resolves the problems which led Tarski to employ the notion of an assignment in the first place. But at this point in *Grundgesetze*, Frege has yet to introduce any rules of inference which make essential use of Roman letters. The rule of universal generalization, introduced later in §17, is the first rule which does, and it is in his discussion of what justifies it that Frege confronts the question how inferences involving Roman letters are, in general, to be justified. As we shall see, he uses the same quasi-substitutional language he employs in §29.

In contemporary terminology, Frege's rule of universal generalization allows '$\vdash \Gamma \longrightarrow \forall x.\Phi(x)$' to be inferred from '$\vdash \Gamma \longrightarrow \Phi(x)$', so long as '$x$' is not free in '$\Gamma$' ([7], I:48, rule 5). His argument for the validity of this rule is in three parts. First, '$\vdash \Gamma \longrightarrow \Phi(x)$' is equivalent to '$\vdash \forall x(\Gamma \longrightarrow \Phi(x))$', since a formula containing a Roman letter is equivalent to its universal closure. The second part of the argument is contained in this passage:

> Let us consider the proposition '$\vdash \forall x(\Gamma \longrightarrow \Phi(x))$', in which '$\Gamma$' is a proper name and '$\Phi(\xi)$' is a function-name. $\forall x(\Gamma \longrightarrow \Phi(x))$ is the False if for any argument the function $\Gamma \longrightarrow \Phi(\xi)$ has the False as value. This in turn is the case if $\Gamma$ is the True, and the value of the function $\Phi(\xi)$ is for any argument the False. In all other cases $\forall x(\Gamma \longrightarrow \Phi(x))$ is the True. With this let us compare '$\Gamma \longrightarrow \forall x.\Phi(x)$'. This denotes the False if $\Gamma$ is the True and $\forall x.\Phi(x)$ is the False. But the latter is the case if the value of the function $\Phi(\xi)$ is for any argument the False. In all other cases $\Gamma \longrightarrow \forall x.\Phi(x)$ is the True. The proposition '$\vdash \Gamma \longrightarrow \forall x.\Phi(x)$' thus asserts the same as does '$\vdash \forall x(\Gamma \longrightarrow \Phi(x))$'.

This is a now familiar argument for the equivalence of '$p \longrightarrow \forall x.Fx$' and '$\forall x(p \longrightarrow Fx)$'. But note that it establishes only that '$\vdash \Gamma \longrightarrow \forall x.\Phi(x)$' and '$\vdash \forall x(\Gamma \longrightarrow \Phi(x))$' are equivalent, if '$\Gamma$' and '$\Phi(\xi)$' are *names*—that is, if neither '$\Gamma$' nor '$\Phi(\xi)$' contains a Roman letter—since, otherwise, it would be illegitimate to speak of the truth-value of $\Gamma$. Additional argument is thus required, if the validity of the rule is to be established for the general case, which argument is given in the following passage:

> If for '$\Gamma$' and '$\Phi(\xi)$' combinations of signs are substituted that do not denote an object and function, respectively, but only indicate, because they contain Roman letters, then the foregoing still holds generally if for each Roman letter a name is substituted, whatever this may be. ([7], I:17)

Note how Frege declines to present any new account of the equivalence in this case: the justification given for the simpler case is to apply to this case also. But the question is how it *can* apply to this case: and all Frege says is that, if all Roman letters which occur in the propositions are uniformly replaced with names (of objects or functions, as may be appropriate), the justification will still go through. Consider, for example, the inference from '$\vdash \Psi(a) \longrightarrow \forall x.\Phi(a,x)$' to '$\vdash \forall x[\Psi(a) \longrightarrow \Psi(a,x)]$'. No matter what name we might substitute for '$a$', the justification Frege gave in the simpler case will go through. From this, he concludes that the inference involving Roman letters is also valid.

It may not be immediately clear why Frege draws this conclusion, but an argument can easily be reconstructed. A valid inference is, for Frege, simply one which is truth-preserving. The stipulation mentioned above, about the scopes of Roman letters, does assign truth-values, *simpliciter*, to propositions containing Roman letters: so the rule of universal generalization will be valid just in case the conclusion is true whenever the premise is—that is, as we might put it, if the universal closure of the conclusion is true whenever the closure of the premise is. Now, if the closure of the premise is true, all of its instances are true; hence, by the previous paragraph, every instance of the conclusion will be true; so its universal closure will be true.

One could object, yet again, that this argument hardly shows that the inference is legitimate, since its last step rests upon the assumption that every object in the domain has a name. But it should by now be clear that this objection would miss the point of what Frege is trying to do here. What we would say, in Tarskian terminology, is that the justification given for the case in which no free variables occur can be made to show, not that '$\Psi(a) \longrightarrow \forall x.\Phi(a, x)$' and '$\forall x[\Psi(a) \longrightarrow \Phi(a, x)]$' must have the same truth-value, but that they must have the same truth-value *under any assignment*. Frege does not use (or have) the notion of an assignment; he speaks instead of substituting names (which denote objects) for Roman letters (which merely indicate them). And the *point* of such talk would be obscured if we insisted that the instances of which Frege speaks must be formed using actual expressions of Begriffsschrift: we understand it better if we read him as intending that they should be formed using auxiliary names. The conditions for an expression to denote should be read in the same way: when Frege says that '$\Phi(\xi)$' denotes if and only if '$\Phi(\Delta)$' denotes, for any denoting name '$\Delta$', we should understand '$\Delta$' to be an auxiliary name, added to the language of the theory, which might denote any object in the domain.

***4 The induction step*** Now that we have seen what Frege thinks he needs to do if he is to show a functional expression to denote, we are ready to consider the proof of the induction step given in §30. What Frege needs to show is that the two ways of forming more from less complex names preserve referentiality. It is obvious that the four methods which together comprise the first way do so, and Frege merely remarks that "all names arising in this way succeed in denoting if the primitive simple names do so." For consider method 1, which forms a proper name '$\Phi(t)$' from a functional expression '$\Phi(\xi)$' and a proper name '$t$'. Suppose that both '$\Phi(\xi)$' and '$t$' denote. What it *is* for '$\Phi(\xi)$' to denote is for every expression of the form '$\Phi(u)$' to denote, so long as '$u$' denotes, whatever it may denote. But we have supposed that '$t$' denotes something, so '$\Phi(t)$' certainly must denote. The other three methods can be treated similarly.[14]

It is important to recognize, at this point, that *it is not obvious* that a function-sign formed in the *second* way must denote if the expressions from which it is formed do. Frege therefore needs to argue that it will. He begins with an argument for a syntactic claim, from which he draws a semantical conclusion:[15]

> A proper name can be employed in the present process of formation only by its filling the argument-places of one of the simple or composite names of first-level functions. Composite names of first-level functions arise in the way provided above [that is, in the first way] only from simple names of first-level functions of two arguments, by a proper name's filling the $\xi$- or $\zeta$-argument-places.

> Thus the argument-places that remain open in a composite function-name are always also the argument-places of a simple name of a function of two arguments. From this it follows that a proper name that is part of a name formed in [the first] way, wherever it occurs, always stands at an argument-place of one of the simple names of first-level functions. If we now replace this proper name at some or all places by another, then the proper name so arising is likewise formed in the [first way], and thus it also has a denotation, if all the simple names employed also denote something. ([7], I:30)

The syntactic claim is thus this: let '$\Phi(\Gamma)$' be a proper name, which contains the term '$\Gamma$'; let '$\Phi(\Delta)$' be another proper name, in which some (not necessarily all) occurrences of '$\Gamma$' have been replaced by occurrences of a term '$\Delta$'. Then '$\Phi(\Delta)$' can *also* be formed in the first way: since '$\Gamma$' got where it was by filling the argument-places of simple names—since "a proper name that is part of a name formed in [the first] way, wherever it occurs, always stands at an argument-place of one of the simple names of first-level functions"—we may construct '$\Phi(\Delta)$' by mimicking the construction of '$\Phi(\Gamma)$', but filling (some of the) argument-places we had previously filled with '$\Gamma$', with '$\Delta$'.[16] Note that it is implicit in this argument that '$\Delta$' is not *itself* a name formed in the second way: if '$\Delta$' were a name formed in the second way, then '$\Phi(\Delta)$' obviously could not itself be formed in the first way.

It is worth mentioning, before we move on, that Frege's argument for the syntactic claim depends upon another assumption, one that he does make explicit. His discussion of that assumption reveals just how careful he is trying to be here:

> Of course in this we are assuming that the simple names of first-level functions of one argument have only one argument-place, and that the simple names of first-level functions of two arguments have only one $\xi$- and one $\zeta$-argument place. Otherwise it could indeed occur in the case of the replacement just described that related argument-places of simple function-names were filled by different names, and an explanation of the denotation for this case would be lacking. But this can always be avoided: and must be avoided, so as to prevent the occurrence of names which have no denotation. And there would certainly be no point in introducing several $\xi$-argument-places or several $\zeta$-argument-places into simple function-names. ([7], I:30)

Here Frege is considering the possibility that, among the *simple* names in the system, there should have been one of the form '$F\xi\xi$', that is, a primitive symbol for a *one*-place functional expression, which nonetheless had two occurrences of its single argument-place. If there were such a symbol, we could proceed as follows: first, form the sentence '$Ftt$', in accord with method 1, by filling the one argument-place of '$F\xi\xi$' with the term '$t$'; then, replace only the *first* occurrence of '$t$' with one of '$u$', to get '$Fut$'. If, however, '$F\xi\xi$' had been treated *semantically* as a monadic predicate, the stipulation which determined its denotation would not have provided one for such a sentence as '$Fut$' (except in the special case where '$u$' denoted the same object as '$t$'). We would therefore have no guarantee that '$Fut$' *had* a denotation. (Note that it could not have been formed in the first way.) But there is no reason to allow such primitive expressions as '$F\xi\xi$' in the first place, and, as Frege mentions in §31, none of the primitive expressions of Begriffsschrift are of this peculiar sort.

From the syntactic claim, Frege draws the semantical conclusion that '$\Phi(\Delta)$' denotes, if '$\Delta$' denotes and if '$\Phi(\Gamma)$' was formed from primitive expressions which all denote. The reason is that '$\Phi(\Delta)$' can itself be formed from primitive denoting names in the first way, and we already know that any such expression denotes. Frege then introduces the second way of forming a functional expression and notes that any expression so formed must denote:[17]

> We begin by forming a name in the first way, and we then exclude from it at all or some places, a proper name that is a part of it (or coincides with it entirely)—but in such a way that these places remain recognizable as argument-places.… The function-name resulting from this likewise always has a denotation if the simple names from which it is formed denote something; and it may be used further to form denoting names in the first way or the second. ([7], I:30)

For let '$\Phi(\xi)$' be a functional expression formed from denoting expressions in the second way. It denotes if and only if '$\Phi(\Delta)$' denotes, so long as '$\Delta$' denotes. But '$\Phi(\xi)$' was formed by removing occurrences of some expression '$\Gamma$' from '$\Phi(\Gamma)$', where '$\Phi(\Gamma)$' is a name formed from denoting expressions in the first way. But then, by the semantic corollary to the syntactic claim, '$\Phi(\Delta)$' will denote if the primitive names from which it is constructed denote—which they do, since all of the primitive expressions contained in '$\Phi(\xi)$' denote. Hence, '$\Phi(\xi)$' denotes.[18]

If we read Frege's condition for a functional expression to denote in the usual way, to show that '$\Phi(\xi)$' denotes, Frege needs to show that '$\Phi(\Delta)$' denotes, whenever '$\Delta$' is a denoting name *in the language of the Begriffsschrift*. Now, as said earlier, it is implicit in Frege's argument that '$\Phi(\Delta)$' denotes that '$\Delta$' is not itself formed in the second way. But if '$\Delta$' is an arbitrary denoting name in Begriffsschrift, *there is no reason to suppose that it is not formed in the second way*—as, indeed, it would be were it either '$\dot{\epsilon}.\epsilon = \epsilon$' or '$\forall x.x = x$'—whence the syntactic claim simply does not apply, and the argument collapses. And so, on the substitutional reading, Frege's argument would fail for reasons that had nothing to do with the impredicativity of second-order quantification, but rather because of an oversight marring his theory of the syntax (of even the first-order fragment) of the langauge. On the other hand, if the condition for a functional expression to denote is read as I have suggested it should be, there is no problem. To show that '$\Phi(\xi)$' denotes, Frege needs to show that '$\Phi(\Delta)$' denotes, so long as '$\Delta$' denotes, no matter what—and '$\Delta$' here is an auxiliary name, a new *primitive* name, and so is certainly not one formed in the second way. So the argument goes through.

The argument as Frege states it thus does not concern Begriffsschrift proper, but the result of extending that language by adding auxiliary names: in order to conclude that '$\Phi(\xi)$' denotes, we need to know that '$\Phi(\Delta)$' denotes, so long as '$\Delta$' does, no matter what; the name '$\Phi(\Delta)$' is not in the language of the Begriffsschrift, but contains the auxiliary term '$\Delta$'. The point applies as well to the case Frege does not mention, the formation of composite second-level predicates by removing a functional expression from a proper name. An argument parallel to that Frege gives for the case of first-level predicates will show that expressions formed in this way must denote, so long as their primitive parts do,[19] and this argument too will have to make essential use of auxiliary expressions: to conclude that '$\forall x(\varphi x \longrightarrow \varphi x)$' denotes, we need to know that '$\forall x(\Phi x \longrightarrow \Phi x)$' denotes, so long as '$\Phi\xi$' denotes, no matter what, and '$\Phi\xi$' must be taken to be an auxiliary *functional expression*. Frege was not, I think, at all clear about these matters, but the oversight does not affect his semantic theory, as it applies to ordinary sorts of languages. Unfortunately, the language of the Begriffsschrift is no ordinary sort of case, as we shall see.

***5 The basis case: the logical expressions*** Frege explains, at the beginning of §31, that his proof is supposed to show that every well-formed proper name has a denotation; in light of what has been argued in §30, it is enough to show that all the primitive expressions denote. The arguments Frege gives that the primitive expressions denote appeal to stipulations he makes regarding what their denotation is to be, which I shall call his *semantical stipulations* regarding the primitive expressions.

Thus, what the argument in §31 is supposed to show is that the semantical stipulations made about the primitive expressions suffice to assign each of them a denotation.[20] And, given the form that those stipulations typically take, it is not hard to see that they do. But, before beginning the argument, Frege remarks that he will "start from the fact that names of truth-values denote something, namely, either the True or the False" ([7], I:31). This assumption is needed because the conditions stated in §29 are not "definitions of the phrases 'have a denotation' and 'denote something', [since] their application always presupposes that we have already recognized some names as denoting" ([7], I:30). But to what does the assumption amount exactly? I see no option but to suppose that what Frege means is that, if an expression denotes a truth-value, then it denotes *something*. So the assumption amounts to the stipulation that the domain of the theory is not empty, that, in particular, it contains the two truth-values.[21] Frege begins his argument as follows:

> In order now to show that the function-names '$-\xi$' and '$_\top\xi$' denote something, we have only to show that those names succeed in denoting that result from our substituting for '$\xi$' a name of a truth-value (we are not yet recognizing other objects). This follows immediately from our explanations. The names obtained are again names of truth-values.

One might wonder how showing that the result of substituting any *sentence* in the argument-place of the horizontal could show that it has been assigned a denotation. But this Frege very clearly says, and he justifies the restriction by reminding us that "we are not yet recognizing other objects." The restriction does not follow from the remark discussed in the last paragraph, but it is in a similar vein: what Frege is saying is that we are, at this point, to think of the domain not only as containing the True and the False, at least, but as containing *only* these two objects. The initial goal is to show that, if the domain contains only the two truth-values, the stipulations secure a denotation for at least *some* of the primitive names—what I shall call the *logical* expressions of Begriffsschrift. Having shown that, Frege will attempt to extend the result to the complete system.

Given our earlier discussion, we can see that what Frege intends to show here is that the horizontal and negation-sign have a value for every argument, and he intends to show this by showing that, if we substitute a name '$\Delta$' for '$\xi$' in '$-\xi$', the resulting sentence will denote a truth-value, so long as '$\Delta$' denotes a truth-value.[22] But, as Frege says, it clearly follows from the semantical stipulation concerning the horizontal that it does have a value for every argument: he stipulated that '$-\xi$' has the value True for the True as argument; False, for all other arguments. Hence, a sentence of the form '$-\Delta$' will denote, so long as '$\Delta$' denotes, whatever it might denote, since the sentence will denote the True if $\Delta$ is the True; the False, otherwise. Frege disposes of the cases of negation, the conditional, and the identity-sign with no more fuss.

Moreover, we are now in a position to understand Frege's argument concerning the universal quantifier:

> To investigate whether the name '$\forall x.\varphi(x)$' of a second-level function denotes something, we ask whether it follows universally from the fact that the function-name '$\Phi(\xi)$' denotes something that '$\forall x.\Phi(x)$' succeeds in denoting. Now, '$\Phi(\xi)$' has a denotation if, for every denoting proper name '$\Delta$', '$\Phi(\Delta)$' denotes something. If this is the case, then this denotation either always is the True (whatever '$\Delta$' denotes), or not always. In the first case '$\forall x.\Phi(x)$' denotes the True; in the second, the False. Thus it follows universally from the fact that the substituted function-name '$\Phi(\xi)$' denotes something, that '$\forall x.\Phi(x)$' denotes something. Consequently the function-name '$\forall x.\varphi(x)$' is to be admitted into the sphere of denoting names. The same follows similarly for '$\forall F.\varphi_x(Fx)$'. ([7], I:31)

The question is whether '$\forall x.\Phi(x)$' denotes, so long as '$\Phi(\xi)$' denotes, no matter what. Assuming that '$\Phi(\xi)$' denotes amounts to assuming that '$\Phi(\Delta)$' denotes, so long as '$\Delta$' denotes, no matter what, that is, that '$\Phi(\xi)$' has a value for every argument. But if '$\Phi(\xi)$' does have a value for every argument, then that value must either always be the True or not: if so, then by the stipulation Frege has made regarding the denotation of '$\forall x.\varphi(x)$', '$\forall x.\Phi(x)$' denotes the True; if not, the False. Either way, '$\forall x.\Phi(x)$' denotes a truth-value (i.e., is a name of a truth-value) and so denotes something.

Frege does not argue specially for the case of the second-order quantifier. But if this is the correct reading of his argument concerning the first-order quantifier, there is no need for him to do so: the case really is similar. The expression '$\forall F.\varphi_x(Fx)$' denotes if and only if '$\forall F.\Lambda_x(Fx)$' denotes, so long as '$\Lambda_x(\varphi x)$' denotes, no matter what. But if '$\Lambda_x(\varphi x)$' denotes, then '$\Lambda_x(\Phi x)$' denotes, so long as '$\Phi\xi$' denotes, no matter what, which is to say that '$\Lambda_x(\varphi x)$' has a value for every argument. If so, that value must either always be the True, or not: if so, '$\forall F.\Lambda_x(Fx)$' denotes the True; if not, the False.

That completes Frege's demonstration that the logical expressions of Begriffsschrift denote, so long as we assume that the domain contains at least the True and the False—and, officially, so long as we assume that the domain contains only the True and the False. Note two things, however. First, the proofs do not really depend upon this latter assumption: what the part of the proof at which we have so far looked actually shows is that the semantical stipulations assign a denotation to the logical expressions, no matter what objects the domain might contain, so long as it contains the two truth-values—assuming, that is, that enough has indeed been said to determine the truth-values of the atomic sentences formed using names of (or variables whose range includes) these other objects. As we shall see, however, Frege does *not* say enough to determine the truth-values of *all* the atomic sentences which must be considered, if he is to apply the inductive argument given in §30 to Begriffsschrift.

Secondly, read in light of the preceding discussion of Frege's talk of instances, the proofs *really do show* that his semantical stipulations regarding the primitive logical expressions suffice to assign each of them a denotation—if by 'having a denotation' we mean what Frege means (which we do). Given what was argued in §30, Frege has thus shown that the semantical stipulations provide a denotation for any expression formed from the logical expressions of Begriffsschrift, no matter what the domain of the theory might be, so long as it contains the two truth-values. A careful examination of the proof will show that Frege has, in fact, proven more, namely, that a given sentence denotes the True if and only if a related condition obtains, where the sentence stating that condition is a *translation* of the sentence in question. It is for this reason that Frege can say, in §32, that

> not only a denotation, but also a sense, attaches to all names correctly formed from our signs. Every such name of a truth-value *expresses* a sense, a *thought*. Namely, by our stipulations it is determined under what conditions the name denotes the True. The sense of this name—the *thought*—is the thought that these conditions are fulfilled.

What Frege has done is to produce an informal, axiomatic theory of truth for the logical fragment of Begriffsschrift and then to prove, informally, that the theory in question is *adequate*,[23] in roughly Tarski's sense. That is no mean feat.

***6 The basis case: the smooth breathing (I)*** That accomplished, Frege next attempts to show that the smooth breathing denotes. As he remarks, this part of the proof is more complicated:

> For with this we are introducing not merely a new function-name, but simultaneously answering to every name of a first-level function of one argument, a new proper name (value-range name) . . . . ([7], I:31)

At this point we need to expand the domain of the theory: it will not do to take the domain to contain merely the truth-values; it must also contain value-ranges, which are to be the denotations of terms of the form '$\dot{\epsilon}.\Phi(\epsilon)$'.

Frege does not, however, tell us explicitly what the domain of the theory is to be: he cannot do so, because he is attempting, by means of the smooth breathing and the semantical stipulation he makes concerning it, to *introduce* value-ranges—not just into the system, but, so to speak, entirely.[24] The semantical stipulation governing the smooth breathing is not like the stipulations Frege gives for the other primitives. In the case of the horizontal, for example, he writes:

> I regard it as a function-name, as follows: $-\Delta$ is the True if $\Delta$ is the True; on the other hand it is the False if $\Delta$ is not the True. ([7], I:5)

Frege does, in fact, make a similar stipulation concerning the smooth breathing:[25]

> . . . $\dot{\epsilon}.\Phi(\epsilon)$' denotes the value-range of the function '$\Phi(\xi)$' . . . . ([7], I:9)

However, what he means by a 'value-range' is explained only as follows:

> I use the words 'the function $\Phi(\xi)$ has the same *value-range* as the function $\Psi(\xi)$' generally to denote the same as the words 'the functions $\Phi(\xi)$ and $\Psi(\xi)$ always have the same value for the same argument'. ([7], I:3)

The net effect of all of this is that the only semantical stipulation Frege has actually made is this one:

> '$\dot{\epsilon}.\Phi(\epsilon) = \dot{\epsilon}.\Psi(\epsilon)$' is co-referential with '$\forall x[\Phi(x) = \Psi(x)]$'.

It is this that Frege actually uses in his argument that the smooth breathing denotes[26] (together with an additional stipulation made in §10, which we shall discuss shortly). But it is far from obvious that this stipulation actually *does* assign a denotation to the smooth breathing. This is a large part of the reason Frege needs to prove that every well-formed expression of Begriffsschrift has a denotation: the only part of the proof that is at all difficult is the proof that the smooth breathing denotes. Frege begins that argument as follows:

> To the inquiry whether a value-range name denotes something, we need only subject such value-range names as are formed from denoting names of first-level functions. We shall call these for short *fair* value-range names. ([7], I:31)

By the stipulation made in §29, regarding when a second-level functional expression denotes, Frege must show that every proper name resulting from the substitution of an auxiliary name of a first-level function, for the argument of the smooth breathing, denotes—that is, that '$\dot{\epsilon}.\Phi(\epsilon)$' denotes, so long as '$\Phi(\xi)$' denotes, no matter what. That is why we can restrict attention to 'fair' value-range names: to say that '$\dot{\epsilon}.\varphi(\epsilon)$' denotes *just is* to say that all fair value-range names denote.

It is at this point, then, that we need to consider Frege's condition for a proper name to denote.

> A proper name *denotes* if the proper name that results from its filling the argument-places of a denoting name of a first-level function of one argument always has a denotation, and if the name of a first-level function that results from its filling the $\xi$-argument-places of a denoting name of a first-level function of two arguments always has a denotation, and if the same holds also for the $\zeta$-argument-places. ([7], I:29)

What one would have expected Frege to say would have been something like: a proper name denotes if and only if there is some object it denotes. But the condition does not take anything like this form: given the form of the semantical stipulation governing the smooth breathing, there is no way for Frege to show that any term of the form '$\dot{\epsilon}.\Phi(\epsilon)$' *does* denote an object, for the stipulation does not directly assign denotations to these terms.

Read in the way we read the conditions for functional expressions to denote, the condition for a proper name to denote would be this: a term '$t$' denotes if and only if '$\Psi(t)$' denotes, so long as '$\Psi(\xi)$' denotes, no matter what; in particular, '$\dot{\epsilon}.\Phi(\epsilon)$' will denote just in case '$\Psi(\dot{\epsilon}.\Phi(\epsilon))$' denotes, so long as '$\Psi(\xi)$' denotes. This, however, cannot be what Frege takes the condition to mean, for he does not argue for any such claim in §31. What he argues, at least initially, is that the result of substituting such a term in any of the *primitive, first-level logical expressions* will denote, that is, that they have a value for every value-range as argument.

> We must examine whether a fair value-range name placed in the argument-places of '$-\xi$' and '$\top\xi$' yields a denoting proper name, and further whether, when placed in the $\xi$-argument-places or in the $\zeta$-argument-places of ['$\zeta \longrightarrow \xi$'] and '$\xi = \zeta$', it always forms a denoting name of a first-level function of one argument. ([7], I:31)

We shall look at his argument for an affirmative answer—which I shall call the *central claim*—in the next section.

Our present problem is to understand what the condition for a term to denote is supposed to be, that is, how it is supposed to follow from the central claim that fair value-range names denote. Perhaps surprisingly, though, Frege himself says almost nothing about the matter:

> We have seen that each of our simple names of first-level functions '$-\xi$', '$\top\xi$,' ['$\zeta \longrightarrow \xi$'], and '$\xi = \zeta$', up to now recognized as denoting, produces denoting names upon admission of fair value-range names in the argument-places. Thus the fair value-range names may be admitted into the sphere of denoting names. Thereby, however, the same thing is decided for our function-name '$\dot{\epsilon}.\varphi(\epsilon)$', since it now follows universally from the fact that a name of a first-level function denotes something, that the proper name resulting from its being substituted in '$\dot{\epsilon}.\varphi(\epsilon)$' denotes something. ([7], I:31)

The central claim having been established, the argument is then to be completed thus: it follows that all fair value-range names denote and, from this, that the smooth breathing does. The second inference we have discussed and seen to be unproblematic. What justifies the first?

It is hard to see that there is any alternative to supposing that Frege intends some sort of induction at this point. And the usual, substitutional interpretation of the argument derives a great deal of support from its ability to explain what Frege means to be arguing here. On that reading, the condition for a term '$t$' to denote is that '$\Psi(t)$' should denote, whenever '$\Psi(\xi)$' is a denoting functional expression in Begriffsschrift; what Frege intends to show is that '$\Psi(\dot{\epsilon}.\Phi(\epsilon))$' denotes, if '$\Psi(\xi)$' is a denoting functional expression. The central claim is precisely that this holds if '$\Psi(\xi)$' is a primitive expression, and the proof is to be completed by an application of the argument of §30.[27] On the usual reading, then, two applications are made of that argument, back-to-back: the first, showing that '$\Psi(\dot{\epsilon}.\Phi(\epsilon))$' denotes, whatever denoting functional expression '$\Psi(\xi)$' might be; the second, that all well-formed expressions denote, where the first application of the argument establishes part of the basis of this induction, namely, that the smooth breathing denotes.

We have, of course, already seen that the substitutional interpretation of Frege's argument faces fatal objections. But it will not suffice simply to point to them in response to the interpretation just outlined. In principle, one could concede that Frege's argument cannot, *in general*, be read along substitutional lines, but yet insist that *this part* of the argument must be so read—in particular, that, since the condition for a proper name to denote cannot be read as I have argued the other conditions should be, there is no alternative but to read it substitutionally.

Still, I doubt that this hybrid interpretation can be right. Frege cannot be attempting to show that '$\Psi(\dot{\epsilon}.\Phi(\epsilon))$' denotes, *whatever* functional expression '$\Psi(\xi)$' might be: one such expression is '$\dot{\alpha}(\alpha = \xi)$', and we cannot conclude, as it were, *during* the induction, that '$\dot{\alpha}(\alpha = \dot{\epsilon}.\Phi(\epsilon))$' denotes: we simply have no way to show that yet, since the whole *point* of the argument is to show that fair value-range names denote.[28] Of course, it is part of the condition that '$\Psi(\xi)$' should itself denote, but it is unclear what force this restriction might have: it is not as if '$\dot{\alpha}(\alpha = \xi)$' *doesn't* denote.[29]

The key to understanding Frege's condition for a name to denote is a passage early in §31. He writes:

> We start from the fact that the names of truth-values denote something, namely, either the True or the False. We then gradually widen the sphere of names to be recognized as denoting by showing that those to be adopted, *together with those already adopted*, form denoting names by way of the one's appearing at fitting argument-places of the other. (Italics mine)

Note carefully what Frege says here: that it will suffice to show that a new expression denotes to show that the results of putting it in the argument-places of names "already adopted," that is, *already recognized as denoting*, denote. So, in the context of the argument that the smooth breathing denotes, what needs to be shown is that '$\Psi(\dot{\epsilon}.\Phi(\epsilon))$' denotes, if '$\Psi(\xi)$' is an expression already recognized as denoting—that is, a functional expression constructed *from the logical expressions alone*, since those are the expressions we have, at that point, recognized as denoting. Indeed, since what are "adopted" are *primitive* expressions, the condition might be as weak as this: that the results of putting fair value-range names in the argument-places of the *primitive* logical expressions denote.[30] This reading of Frege's condition fits well with the passage quoted above:

> We have seen that each of our simple names of first-level functions '$-\xi$, '$\top\xi$', ['$\zeta \longrightarrow \xi$'], and '$\xi = \zeta$', *up to now recognized as denoting*, produces denoting names upon admission of fair value-range names in the argument-places. Thus the fair value-range names may be admitted into the sphere of denoting names. (Italics mine)

What Frege *seems* to be saying here is that it follows *immediately* from the central claim that fair value-range names denote. And, on my interpretation, that is exactly right: what it is for a name to denote is that all *atomic* sentences in which it occurs should have a denotation; the central claim all but *is* that fair value-range names denote.

The view I am ascribing to Frege is one to which we know he was attracted at the time he wrote *Die Grundlagen*. There, he explicitly considers the question whether fixing the meanings of all identity-statements in which some term '*t*' occurs will suffice to fix the meaning of that term ([14], 62). The identity-statements in which fair value-range names occur are, of course, among the atomic formulas in which they occur—and, as we shall see below, Frege correctly argues that fixing the truth-values of identity-statements will suffice to fix the denotations of *all* atomic formulas in which fair value-range names occur. In effect, Frege's condition for fair value-range names to denote therefore *is* the condition discussed in *Die Grundlagen*.

In the case of most interest to him, Frege considers the stipulation that a sentence of the form 'the number of Fs is the same as the number of Gs' is to have the same truth-value as 'The Fs are equinumerous with the Gs'. (This is a semantical version of what is now called Hume's Principle.) His objection is not that the procedure is wrong in principle: the objection is the Julius Casear objection, that the stipulation *does not* fix truth-values for such statements as 'Casear is the number of Fs'—not, note, that it *could not* be supplemented so that it would, although Frege did not see how the supplementation might be effected (see [14], 107). The case of the semantical stipulation governing the smooth breathing is, as has often been pointed out, entirely parallel: the two stipulations are of the same form and, as we shall see, a version of the Caesar objection arises in the case of the smooth breathing, too. Less often mentioned is what follows from the parallelism: that, if, in the intervening years, Frege had found an answer to the Caesar objection that satisfied him, he would have been free to claim that the semantical stipulation governing the smooth breathing *does* suffice to assign denotations to all fair value-range terms. As we shall see, Frege's argument for the central claim amounts, in large part, to an argument that he has got such an answer.

One might object to this interpretation that Frege now seems to be offering something like a *contextual definition* of names of value-ranges: yet he explicitly denies doing so and is opposed to such definitions in general. But I am not claiming that Frege intended the semantical stipulation governing the smooth breathing as a *definition* of names of value-ranges[31] but as a means of (partially) fixing their reference. Moreover, Frege himself was aware that the way he introduces value-range names is easily confused with apparently similar procedures to which he objects:

> If there are logical objects at all—and the objects of arithmetic are such objects—then there must also be a means of apprehending, of recognizing, them. This service is performed for us by the fundamental law of logic that permits the transformation of an equality holding generally into an equation . . . . We thus hope to be able to develop the whole wealth of objects and functions treated of in mathematics out of the germ of the eight functions whose names are enumerated in vol. i, §31. Can our procedure be termed construction? Discussion of this question may easily degenerate into a quarrel over words. In any case our construction (if you like to call it that) is not unrestricted and arbitrary; the mode of performing it, *and its legitimacy*, are established once and for all. (Italics mine) ([7], II:147, see also 146.)

Frege does not refer to §31 here just because his eight primitive expressions happen to be listed there: rather, it was there that he intended to establish the legitimacy of the semantical stipulation governing the smooth breathing, by showing that it does suffice to assign it a denotation. What distinguishes his method from those to which he objects is not so much the kind of procedure he uses—some kind of "construction" or "abstraction"—but the fact that he has, first of all, formalized the method and, secondly, established its legitimacy, in that sense.

But one might well wonder whether, if Frege does adopt this condition for a name to denote—that all atomic sentences in which it occurs denote—he can still argue, on that basis, that every well-formed expression denotes. The answer, however, is that he can—or rather that, though he cannot, the reason has nothing to do with the *specific* interpretation I am offering of the condition. The argument of §30 purports to show that, if all the primitive expressions denote, then all expressions correctly formed from them denote: in effect, it purports to show that, if all *atomic* sentences denote, so do all complex sentences. Applied to the case at hand, the argument goes like this. Call an expression which can be constructed in the *first* way, from the logical expressions and the smooth breathing, a *simple expression of rank 0*.[32] The argument that all logical expressions denote, and that expressions formed from denoting expressions in the first way denote, shows that all expressions of rank 0 constructed *without* using the smooth breathing denote. Moreover, an expression like '$\dot{\epsilon}.(-\epsilon)$' is a fair value-range term, since '$-\xi$' denotes: so '$\dot{\epsilon}.(-\epsilon)$' denotes. By the central claim, all such expressions as '$-\dot{\epsilon}.\Phi(\epsilon)$' and '$\dot{\epsilon}.\Psi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon)$' denote, if '$\dot{\epsilon}.\Phi(\xi)$' and '$\dot{\epsilon}.\Psi(\epsilon)$' are fair value-range terms: so expressions like '$-\dot{\epsilon}.(-\epsilon)$' and '$\xi = \dot{\epsilon}.(-\epsilon)$' denote; so expressions like '$\forall x.x = \dot{\epsilon}.(-\epsilon)$' denote. And so on and so forth. So all simple expressions of rank 0 denote.

And now the induction can continue. We can form functional expressions in the *second* way from simple expressions of rank 0. Call these *composite* expressions of rank 0, and call expressions constructed in the first way from the logical expressions, the smooth breathing, and composite expressions of rank 0, *simple expressions of rank 1*.[33] Frege takes himself to have shown, in §30, that if '$F\xi$' is a composite expression of rank 0, it denotes. It follows, first, that all expressions of rank 1 which are constructed without additional uses of the smooth breathing—for example, '$\forall x.Fx$'—denote. Secondly, '$\dot{\epsilon}.F\epsilon$' is a fair value-range term, which therefore denotes; so, by the central claim, '$-\dot{\epsilon}.F\epsilon$' and the like denote. And so on and so forth. So all expressions of rank 1 denote, and off we go.

There are problems with this argument, but we knew that there would be problems. Consider '$\forall F.F(\dot{\epsilon}.(-\epsilon))$'. To conclude that it denotes, we need to know that the composite second-level functional expression '$\varphi(\dot{\epsilon}.(-\epsilon))$' denotes. This is a composite expression of rank 0, which is constructed from '$-\dot{\epsilon}.(-\epsilon)$' by omitting the first occurrence of the horizontal. Now, '$\varphi(\dot{\epsilon}.(-\epsilon))$' denotes just in case '$\Psi(\dot{\epsilon}.(-\epsilon))$' denotes, so long as '$\Psi(\xi)$' denotes, no matter what. The argument that it does is the argument that expressions formed in the second way denote, that '$\Psi(\dot{\epsilon}.(-\epsilon))$' could have been formed in the first way—that is, that is itself a simple expression of rank 0—and we already know all such expressions to denote. Hence, as was said at the end of §4 above, Frege's argument presupposes that *auxiliary functional expressions are among the primitive expressions of the language* (that is, among those from which simple expressions of rank 0 are constructed). Of course, it is easy enough to include them. But if we do, the argument stalls almost immediately, for we have no way to show that '$\Psi(\dot{\epsilon}.(-\epsilon))$', which is a simple expression of rank 0, denotes, for *any* particular denotation that '$\Psi(\xi)$' might have. The point may also be put in Tarskian terms: what we need to know is that '$F(\dot{\epsilon}.(-\epsilon))$' denotes, no matter what might be assigned to '$F\xi$', and, to determine a truth-value for this formula under such an assignment, we have to know *which* object in the domain '$\dot{\epsilon}.(-\epsilon)$' denotes.[34] But the semantical stipulation governing the smooth breathing does not tell us that—and the condition for a proper name to denote *deliberately* falls short of requiring that it should. That is why I said above that the failure of Frege's argument is independent of my specific interpretation of this condition. It is due, quite simply, to his not reading that condition the way he should—as requiring that '$\Psi(t)$' denotes, whenever '$\Psi(\xi)$' denotes, no matter what.

Note that this part of the argument does, however, work for the *first-order* fragment of the language.[35] The problem with the proof concerns the treatment of formulas which contain auxiliary functional expressions—roughly, free second-order variables—and there is no need to make use of such expressions in giving a semantic theory for the first-order fragment. And so, if Frege's argument that the smooth breathing denotes also worked—that is, if his argument for the central claim worked—he would have given a *correct* proof of the soundness of the first-order fragment of the Begriffsschrift. Unfortunately, as we shall see, Frege's proof of the central claim is irremediably flawed.

***7 The basis case: the smooth breathing (II)*** As said earlier, Frege's argument that the smooth breathing denotes is an argument for what I have called the *central claim*, that "each of our simple names of first-level functions '$-\xi$, '$_\top\xi$', ['$\zeta \longrightarrow \xi$'], and '$\xi = \zeta$', up to now recognized as denoting, produces denoting names upon admission of fair value-range names in the argument-places." He argues that it suffices to show that the result of substituting a fair value-range name into '$\xi = \zeta$' denotes. For, if so, then

> it is also known that we always obtain a denoting name from the function-name '$\xi = (\xi = \xi)$', if we put in the argument-places a fair value-range name. Since now, according to our stipulations, the function $-\xi$ always has the same value for the same argument as the function $\xi = (\xi = \xi)$, it is also known of the function-name '$-\xi$' that a proper name of a truth-value always results from it by substitution of a fair value-range name. By our stipulations the names '$_\top\Delta$' and ['$\Delta \longrightarrow \Gamma$'] always have denotations if the names '$-\Delta$' and '$-\Gamma$' denote something. Since this is now the case if '$\Gamma$' and '$\Delta$' are fair value-range names, we always obtain denoting proper names from the function names '$_\top\xi$' and ['$\zeta \longrightarrow \xi$'] by placing fair value-range names or names of truth-values in the argument-places.

Since '$-\Delta$' has the same denotation as '$\Delta = (\Delta = \Delta)$' (see [7], I:10), and since '$\dot{\epsilon}.\Phi(\epsilon) = (\dot{\epsilon}.\Phi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon))$' denotes, '$-\dot{\epsilon}.\Phi(\epsilon)$' denotes. The other operators then take care of themselves, since they embed a horizontal.

It will thus suffice to show that the result of substituting a fair value-range name for one argument of the identity-function denotes. Frege begins that argument as follows:

> The question is whether '$\xi = \dot{\epsilon}.\Phi(\epsilon)$' is a denoting name of a first-level function of one argument,[36] and to that end it is to be asked in turn whether all proper names, that result from our substituting in the argument-place either a name of a truth-value or a fair value-range name, denote. By our stipulations, that '$\dot{\epsilon}.\Phi(\epsilon) = \dot{\epsilon}.\Psi(\epsilon)$' is always to have the same denotation as '$\forall x[\Phi(x) = \Psi(x)]$', that [the True is identical with its own unit class], and that [the False is identical with its own unit class], a denotation is thus secured in every case for a proper name of the form '$\Gamma = \Delta$' . . . .

The functional expression '$\xi = \epsilon.\Phi(\epsilon)$' will denote if and only if '$\Delta = \dot{\epsilon}.\Phi(\epsilon)$' denotes, so long as '$\Delta$' denotes, no matter what. Frege supposes that it suffices to consider only instances of two sorts: those of the form '$\dot{\epsilon}.\Psi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon)$', where '$\dot{\epsilon}.\Psi(\epsilon)$' is itself a fair value-range name, and those of the form '$\Gamma = \dot{\epsilon}.\Phi(\epsilon)$', where '$\Gamma$' denotes a truth-value. The denotations of instances of the first sort are supposed to be determined by the semantical stipulation governing the smooth breathing. In order to provide a denotation for instances of the second sort, though, Frege will need to specify whether the truth-values are value-ranges and, if so, which they are—whence the second case should reduce to the first. It is in [7], I:10 that Frege discusses this question. He argues, first, that the semantical stipulations made to that point do not decide it and, moreover, that he is free to stipulate that the True and the False should be any distinct value-ranges at all. Eventually, he stipulates that the True and the False are to be identified with their own unit classes. Frege claims that this stipulation, together with that governing the smooth breathing, then decides the truth-values of all instances of the second sort.[37]

It is tempting to think that it is in restricting attention to instances of these two sorts that Frege makes his second mistake. For suppose that there are objects other than truth-values and value-ranges in the domain of the theory—as many would think there must be, since it is widely held that, for Frege, the quantifiers always range over all the objects there are. If so, then showing that instances of these two sorts denote does not show that '$\Delta = \dot{\epsilon}.\Phi(\epsilon)$' denotes so long as '$\Delta$' denotes, *no matter what*—in particular, it does not show that this sentence denotes if '$\Delta$' denotes something other than a truth-value or value-range, say, Julius Casear. One might have thought otherwise: one might suggest, as Moore and Rein do[38] that

> Frege was concerned only with questions which could be stated within his formalism. Since his formalism contains proper names only for value-ranges and the two truth-values, questions involving other objects . . . cannot be formulated within the system. ([19], note 9)

But this simply isn't true. We can formulate the question whether every object is a value-range: is it the case that $\forall x \exists F.x = \dot{\epsilon}.F(\epsilon)$? Now, Frege's goal is to show that every well-formed expression denotes: and to show that '$\forall x \exists F.x = \dot{\epsilon}.F(\epsilon)$' denotes, he must show that the functional expression '$\exists F.\xi = \dot{\epsilon}.F(\epsilon)$' denotes. To show that, Frege needs to show that '$\exists F.\Delta = \dot{\epsilon}.F(\epsilon)$' denotes, so long as '$\Delta$' does, and so also to show that '$\Delta = \dot{\epsilon}.\Phi(\epsilon)$' denotes, so long as '$\Delta$' and '$\Phi(\xi)$' do. That is: by Frege's own lights, '$\forall x \exists F.x = \dot{\epsilon}.F(\epsilon)$' will not denote unless '$\Delta = \dot{\epsilon}.\Phi(\epsilon)$' does, so long as '$\Delta$' does, *no matter what*. If Caesar is a member of the domain, the case in which '$\Delta$' denotes him must be considered; it is quite irrelevant whether there is a name in Begriffsschrift that denotes him. It follows that, if we accept that, for Frege, the domain of quantification is always unrestricted,[39] we must convict him of here having made his second mistake.

There is an alternative: we can take Frege tacitly to be restricting the domain of the theory to truth-values and value-ranges. This might seem unmotivated. But it was in order to handle cases in which '$\Delta$' denotes a truth-value that Frege needed to make a special stipulation in §10 concerning the truth-values. Had there been objects other than truth-values and value-ranges in the domain, similar stipulations would have had to be made regarding whether they were value-ranges, and if so which ones, *for exactly the same reason*. One might suggest that Frege thought that he needed to make a stipulation concerning the truth-values only because there are actual terms in Begriffsschrift which denote them—and so convict him of having made the same mistake made by Moore and Rein. But it is utterly implausible that Frege should have made this mistake. I did not choose the sentence '$\forall x \exists F.x = \dot{\epsilon}.F(\epsilon)$' arbitrarily. Although it is hard to know whether Frege thought much about it, he undoubtedly thought *a great deal* about the predicate '$\exists F.\xi = \grave{\epsilon}.F(\epsilon)$', for the question whether the semantical stipulation governing the smooth breathing has provided this predicate with a denotation *just is the Julius Caesar problem*.[40]

Frege's discussion of that problem, at [14], 66–67, concerns sentences of the form '$t = Nx : Fx$'. ('$Nx : \varphi x$' is here an expression of the same type as '$\grave{\epsilon}.\varphi(\epsilon)$', completions of which are intended to denote cardinal numbers.) At this point in *Die Grundlagen*, Frege is considering whether the stipulation that '$Nx : Fx = Nx : Gx$' is to have the same truth-value as '$F\xi$ is equinumerous with $G\xi$' suffices to explain, not only names of individual numbers, but *the concept of number*: the title of that chapter, "To obtain the concept of number, we must fix the sense of a numerical identity," indicates that the plan is to fix the truth-values of numerical identities by means of this 'contexual definition' and then to use that to define the concept of number.

The problem Frege raises, the "third doubt which may make us suspicious of our proposed definition," is that the relevant stipulation does not fix the content of a sentence of the form '$t = Nx : Fx$', unless '$t$' is itself of the form '$Nx : Gx$':[41]

> Naturally, no one is going to confuse Julius Caesar with the number belonging to the concept F; but that is no thanks to our definition of number. That says nothing as to whether the proposition
>
> > 'the number belonging to the concept $F$ is identical with $q$'
>
> should be affirmed or denied, except for the one case where $q$ is given in the form 'the number belonging to the concept $G$'. What we lack is the concept of number; for if we had that, then we could lay it down that, if $q$ is not a number, our proposition is to be denied, while if it is a number, our original definition will decide whether it is to be denied or affirmed. So the temptation is to give as our definition:
>
> > $q$ is a number, if there is a concept $F$ whose number is $q$.
>
> But then we have obviously come round in a circle. For in order to make use of this definition, we should have to know already in every case whether the proposition
>
> > 'the number belonging to the concept $F$ is identical with $q$'
>
> was to be affirmed or denied. ([14], 66)

Frege is here using the Caesar objection to argue that the stipulation governing '$Nx : \varphi x$' fails to fix the concept of number. If it had fixed it, the content of the predicate '$\xi = Nx : Fx$' would itself have been fixed, that is, the truth-value of every sentence of the form '$q = Nx : Fx$' would have been fixed: but, as the case of Caesar shows, there are sentences of that form whose truth-value has not been fixed. The most obvious way to try to fix the concept of number—to define it as '$\exists F.\xi = Nx : Fx$'—thus fails, precisely because the stipulation has not fixed the content of '$\xi = Nx : Fx$'.

The same problem arises in *Grundgesetze*, only this time the functional expression is not '$Nx : \varphi x$', but '$\grave{\epsilon}.\varphi(\epsilon)$'; the stipulation is not that introduced in [14], 63, but that governing the smooth breathing; and the problematic predicate is not '$\exists F.\xi = Nx : Fx$', but '$\exists F.\xi = \grave{\epsilon}.F(\epsilon)$'. But the two situations are entirely parallel. Frege's solution in *Die Grundlagen*—to define numbers explicitly in terms of extensions of concepts—is obviously not available here, since extensions of concepts are among the things he is attempting to introduce. But nine years later, Frege still has no general solution to offer, so he attempts to finesse the issue by making a specific stipulation about the truth-values: if that is to work, the domain has to be limited to truth-values and value-ranges, and Frege must admit that he has not fixed the concept of a value-range completely, but has only "determined the value-ranges so far as is here possible" ([7], I:10). Since this stipulation is made precisely in order to resolve a special case of the Caesar problem, by providing for the case in which '$\Delta$' denotes a truth-value, it is extremely unlikely that Frege could have overlooked the need to make similar stipulations about any objects other than truth-values and value-ranges which might have been in the domain.

I thus see no option but to suppose that Frege intended to limit the domain of the theory to truth-values and value-ranges.[42] If so, his considering only the two sorts of instances he does is not a flaw in his argument. Every object in the domain is either a truth-value or the denotation of '$\dot{\epsilon}.\Psi(\epsilon)$', where '$\Psi(\xi)$' denotes a function, since every value-range is the value-range of some function. Hence, to show that '$\Delta = \dot{\epsilon}.\Phi(\epsilon)$' denotes, whatever '$\Delta$' denotes, it is enough to show that '$\Gamma = \dot{\epsilon}.\Phi(\epsilon)$' denotes, so long as '$\Gamma$' denotes a truth-value, and that '$\dot{\epsilon}.\Psi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon)$' denotes, so long as '$\Psi(\xi)$' denotes, no matter what. (That is why we can restrict attention to fair value-range names on the *left*-hand side of the identity-statement.) And now it might seem like Frege is home: the stipulations made in §10 reduce the first case to the second; and the semantical stipulation governing the smooth breathing tells us that '$\dot{\epsilon}.\Psi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon)$' is true just in case '$\forall x(\Psi(x) = \Phi(x))$' is true.

Unfortunately, matters are not so simple. Let us proceed slowly. Frege's intention is to determine whether '$\Delta = \dot{\epsilon}.\Phi(\epsilon)$' is true, where '$\Delta$' denotes a particular value-range—call it $\tau$—by asking whether '$\dot{\epsilon}.\Psi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon)$' is true, when '$\Psi(\xi)$' denotes a function whose value-range is $\tau$. There is *some* suitable function for '$\Psi(\xi)$' to denote, $f(\xi)$, say, since every value-range is the value-range of some function. The stipulation governing the smooth breathing then tells us that '$\dot{\epsilon}.\Psi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon)$' is true if and only if '$\forall x[\Psi(x) = \Phi(x)]$' is true, where '$\Psi(\xi)$' denotes $f(\xi)$. Suppose, however, that $\tau$ is the value-range of two *different* functions. Then '$\Delta = \dot{\epsilon}.\Phi(\epsilon)$' is also true just in case '$\dot{\epsilon}.\Psi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon)$' is true, where '$\Psi(\xi)$' denotes some other function, $g(\xi)$, say. But then the semantical stipulation governing the smooth breathing makes '$\Delta = \dot{\epsilon}.\Phi(\epsilon)$' both true and false when '$\Phi(\xi)$' denotes $f(\xi)$ and '$\Delta$' denotes $\tau$.[43] Frege is thus tacitly supposing that no object in the domain is the value of '$\dot{\epsilon}.\Phi(\epsilon)$' for more than one assignment to '$\Phi(\xi)$'. And that is impossible, since it requires that the objects in the domain be in one-one correspondence with the concepts true or false of them, that is, with the power set of the domain, *contra* Cantor's Theorem.

If we leave matters there, however, it looks as if the problem with the proof lies in the nature of the semantical stipulation Frege actually makes.[44] But suppose we say, not that '$\dot{\epsilon}.\Psi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon)$' is to have the same truth-value as '$\forall x[\Psi(x) = \Phi(x)]$', but that '$Nx : \Psi x = Nx : \Phi x$' is to have the same truth-value as '$\Psi(\xi)$ is equinumerous with $\Phi(\xi)$'. Suppose, further, that we restrict the domain to truth-values and *numbers*, and offer a solution to the Caesar problem for that case parallel to the one Frege offered in [7], I:10 (say, stipulate that the True is $Nx : x = x$; the False, $Nx : x \neq x$). Would that make any difference? The answer is "No," for we can just repeat the discussion in the previous paragraph, *mutatis mutandis*. The problem, this time, will be that the argument tacitly supposes that no number is the number of non-equinumerous functions. Not that this is impossible: as is by now well known, if the domain is Dedekind infinite, it *is* possible. The difficulty is that the argument *assumes* precisely what it is supposed to *prove*. For the tacit assumption is that '$Nx : \varphi x$' has a denotation consistent with the semantical stipulation in question, when what the argument is supposed to show is precisely that the stipulation suffices to assign it one. The proof is, therefore, viciously circular.

To summarize: the argument that the smooth breathing denotes amounts to an argument that the semantical stipulation governing the smooth breathing suffices to fix the denotations of sentences of the form '$\Delta = \dot{\epsilon}.\Phi(\epsilon)$'; the fundamental idea behind the proof is that, since the domain contains nothing but value-ranges, the denotation of such a sentence can be identified with that of a corresponding sentence of the form '$\dot{\epsilon}.\Psi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon)$'. But this move depends upon a tacit supposition that every object is the value-range of but one function. The moral of the story is thus this: despite all Frege's effort and ingenuity, he still has not resolved the Caesar Problem. This is not because there is a lingering problem about identity-statements involving objects *other* than value-ranges, but because his stipulations do not suffice to fix the truth-values of sentences such as '$\Delta = \dot{\epsilon}.\Phi(\epsilon)$', even if we assume that every object *is* a value-range. The situation could hardly be more ironic: it was, to answer one of Dummett's questions, the Caesar Problem that was the serpent in Eden; yet it was it that led Frege to introduce value-ranges in the first place.

***8 Closing*** I began by raising a series of questions: how Frege intended to show that every well-formed expression in Begriffsschrift denotes, why he thought he needed to show this, and upon what assumptions his argument depended. We have discussed the nature of Frege's argument in detail and have seen that it depends, at crucial points, upon assumptions he had no right to make. On the other hand, however, we have seen that the argument fails at very *particular* points and that large parts of it are salvageable: in particular, once the nature of Frege's talk of instances has been understood, we can see that he has given an informal theory of truth for the logical fragment of Begriffsschrift and informally proved its adequacy. The problems with the proof arise in connection with the smooth breathing and are due to peculiarities of the semantical stipulation governing it.

As for why Frege offered such a proof at all, we have uncovered a couple of reasons. First, the proof is to show that the system satisfies the 'fundamental principle' that well-formed names "must always denote something" ([7], I:28), most importantly, that the smooth breathing denotes, which is far from obvious, given the form of the semantical stipulation governing it. More interestingly, the argument constitutes a partial resolution of the Caesar problem: it purports to show that the predicate '$\exists F.\xi = \dot{\epsilon}.F(\epsilon)$' has been given a reference, at least when the domain of the theory is restricted to truth-values and value-ranges. Answering this objection allows Frege to do what he could not do in *Die Grundlagen*, namely, to claim that the semantical stipulation governing the smooth breathing *of itself* assigns denotations to the value-range terms and simultaneously determines the domain of the theory. It would have been an extraordinary coup, had it worked.

Indeed, the coup would have been one of a now familiar sort. Frege's plan was, in effect, to fix the denotations of all atomic formulas—identity-statements containing value-range terms being the central case—and then to let that generate a domain for the theory. Far from being silly, this is the same idea that lies behind Henkin's proof of the completeness of the first-order functional calculus. In that case, we take a consistent theory $T$ and expand it to a theory $T'$ that is maximal consistent and 'has witnesses'.[45] We then consider the following equivalence relation between terms:

$$t \sim u \quad \text{iff} \quad `t = u\text{'} \in T';$$

take the domain to consist of equivalence classes under this relation; let a term denote its equivalence class; and so forth. The problems with Frege's argument, seen from this perspective, are two: such a procedure is not available in the case of second-order theories, and Frege does not, in fact, succeed in assigning denotations to all the atomic sentences, in the first place. Indeed, since this sort of construction works only if $T$ is consistent, Frege's attempt to use such a construction to prove the soundness of this theory is, once again, viciously circular.

There is another sort of purpose which §31 is often said to have had: namely, to demonstrate the *consistency* of the Begriffsschrift. According to the interpretation I have offered, this was not among Frege's primary motivations. On the other hand, however, I do think that Frege was aware that it followed from what he had argued that the Begriffsschrift was consistent. The consistency of the system does not, of course, follow from the fact that every well-formed expression has a reference: one could simply stipulate that every name is to denote the True; that every first-level functional expression (including the logical expressions) is to denote the function whose value, for any argument, is the True; and so forth. That would assign every expression a denotation, but it is quite irrelevant to the question of consistency.[46]

But Frege's conclusion was not that all well-formed expressions have been given a denotation, but that *the semantical stipulations made regarding the primitive expressions* provide them with denotations. And Frege argues, elsewhere in Part I of *Grundgesetze*, that those stipulations assign the value True to all axioms of the theory and validate its rules of inference (see [15], §2). Given that there is at least one sentence which is assigned the value False by the stipulations, the consistency of the system follows (or would follow). Indeed, Frege himself was fond of pointing out something similar, that the best (if not the only) way to show that a formal theory is consistent is to show that its axioms are true and that its rules are truth-preserving (Frege [10], pp. 277–78, op. 324 and [11], p. 325, op. 394; see also Dummett [2]).

Moreover, there is solid textual evidence that Frege knew that the argument of §31 would have implied the consistency of the Begriffsschrift. In his first letter to Russell, Frege writes:[47]

> Your discovery of the contradiction has surprised me beyond words . . . . It seems accordingly that the transformation of the generality of an identity into an identity of value-ranges . . . is not always permissible, that my law V . . . is false, and that my explanations in sect. 31 do not suffice to secure a reference for my signs in all cases.

This is a list of three things which Frege thinks follow from Russell's discovery of the inconsistency of Axiom V: among these are that Axiom V is false and that the argument given in [7], I:30–31 does not work. Note how this latter fact is treated as being as obvious a consequence of Russell's discovery as the former, which really is obvious. The reason is simple: it follows from the argument given in §31 that the Begriffsschrift is consistent, and yet "Herr Russell hat einen Widerspruch aufgefunden . . . ."

***9 Appendix: Outline of a Fregean theory of truth*** We begin with an object-language $\mathcal{L}$, say, the language of first-order arithmetic.[48] We shall define truth for an infinite collection of extensions of $\mathcal{L}$. We assume that we have a fixed, countably infinte list of new, auxiliary constants, $a_1$, $a_2$, and so forth. The relevant extensions of $\mathcal{L}$ are those containing finitely many of these auxiliary constants, with respect to which the languages are *interpreted*. The set of all such extensions we denote $\mathcal{L}^*$, and we write '$a_i \in L$' to mean that $a_i$ is a term in the language $L$. We write '$L <_i L'$' to mean that $L'$ extends $L$ precisely by containing the auxiliary constant $a_i$ not contained in $L$; '$\text{den}(t, L)$', to mean the denotation of $t$ in the language $L$; '$\text{true}(S, L)$', to mean that $S$ is true in $L$. And we assume that '$L$' and similar variables range only over languages in $\mathcal{L}^*$.

Much of the truth-theory is familiar. For the primitive, nonlogical vocabulary of $\mathcal{L}$, we adopt a series of axioms stating their semantic values, not just in $\mathcal{L}$, but in all languages $L \in \mathcal{L}^*$. Thus

$$\begin{aligned}
\text{den}(\text{`0'}, L) &= 0 \\
\text{den}(\ulcorner St \urcorner, L) &= S(\text{den}(t, L)) \\
\text{den}(\ulcorner t + u \urcorner, L) &= \text{den}(t, L) + \text{den}(u, L) \\
\text{den}(\ulcorner t \times u \urcorner, L) &= \text{den}(t, L) \times (\text{den}(u, L)
\end{aligned}$$

For the logical vocabulary, other than the quantifiers, we have similar axioms:

$$\begin{aligned}
\text{true}(\ulcorner t = u \urcorner, L) &\quad \text{iff} \quad \text{den}(t, L) = \text{den}(u, L) \\
\text{true}(\ulcorner A \& B \urcorner, L) &\quad \text{iff} \quad \text{true}(A, L) \ \& \ \text{true}(B, L) \\
\text{true}(\ulcorner \neg A \urcorner, L) &\quad \text{iff} \quad \neg\text{true}(A, L)
\end{aligned}$$

Obviously, parallel axioms can be written down for the other connectives.

It is only when we get to the quantifiers that things get interesting. The idea is to say that a sentence $\ulcorner \forall v.A(v) \urcorner$ is true if and only if $\ulcorner A(t) \urcorner$ is true, whatever $t$ might denote. Here, $t$ is to be an auxiliary term, and 'whatever it might denote' is to be cashed out in terms of the extensions of $L$, in which $t$ denotes the various objects in the domain. The clause we need is thus something like

$$\text{true}(\ulcorner \forall v.A(v) \urcorner, L) \quad \text{iff} \quad \forall L' >_i L.\text{true}(\ulcorner A(a_i) \urcorner, L').$$

Although this will do, it is best not to formulate the condition in this form, as it does not extend smoothly to non-first-order quantifiers, such as 'Most'. A more useful variant is

$$\begin{aligned}
&\text{true}(\ulcorner \forall v.A(v) \urcorner, L) \quad \text{iff} \\
&\qquad \forall x \in \text{Dom}(\mathcal{L}) \exists L' >_i L[\text{den}(a_i, L') = x \ \& \ \text{true}(\ulcorner A(a_i) \urcorner, L')].
\end{aligned}$$

And similarly,

$$\text{true}(\ulcorner \exists v.A(v) \urcorner, L) \quad \text{iff}$$

$$\exists x \in \text{Dom}(\mathcal{L})\exists L' >_i L[\text{den}(a_i, L') = x \;\&\; \text{true}(\ulcorner A(a_i)\urcorner, L')]$$

Note how the quantifier '$\forall x$' has simply been replaced by '$\exists x$': a parallel clause is available for other quantifiers.[49]

In order for these clauses to work, we need to make certain assumptions about the languages in $\mathcal{L}^*$. First of all, for any language $L$, for any auxiliary constant $a_i$ not contained in it, and for any object $x$ in the domain, there needs to be a language $L'$ such that $L' >_i L$ and $\text{den}(a_i, L') = x$. Secondly, we need to assume that every language $L$ assigns a unique denotation to each of the auxilary constants $a_i$ contained in it. And finally, we assume that each such language contains only finitely many of the auxiliary constants.[50] Given these assumptions, it is easy to see that the theory of truth sketched above is, as Tarski would say, formally and materially adequate. Indeed, the easiest way to see this is just to note that nothing *in* the theory of truth tells us anything at all about what the 'languages' in question are, nor what the 'auxiliary constants' are supposed to be, nor what it means to say that one of them is 'in' a language. So far as the formal theory is concerned, languages might as well be sequences (better yet, functions from arbitrary finite sets of natural numbers to objects in the domain); auxiliary constants could be free variables; and a constant's being 'in' a language could just be its having been assigned a value by the sequence. The Fregean theory is, therefore, mathematically equivalent to the Tarskian one, although the ideas behind them are quite different (and importantly so in certain contexts).[51]

There are two crucial results that can be proven at this point, both of them analogues of results that Tarski proves. The first, in Tarski, is that, if $S$ is a formula and $\sigma$ and $\tau$ are sequences differing, if at all, only in what they assign to variables not free in $S$, then $S$ is satisfied by $\sigma$ if and only if it is satisfied by $\tau$. In the context of the Fregean theory, the result is that, if $S$ is a sentence and $L$ and $L'$ are languages which agree on the denotations of all auxiliary constants appearing in $S$, then $S$ is true in $L$ if and only if it is true in $L'$. The second is a simple corollary of the first: in Tarski, it is that, if $S$ is a closed sentence, then it is satisfied by all sequences or by none. Here, the result is that, if $S$ does not contain any auxiliary constants, it is true in all languages or in none.

That completes the sketch of one sort of Fregean truth-theory. Although it does make use of Frege's idea of an auxiliary constant, however, it is not entirely true to Frege's thought: it does not, in particular, reflect his doctrine that expressions *of all types* have denotations, and it is for this reason that the notion of an auxiliary constant appears in connection with quantification, rather than with the notion of denotation (as it does in [7], I:29). It is therefore worth sketching the details of an alternative theory, one which, in effect, formalizes the semantic theory Frege presents in Part I of *Grundgesetze*.

We will here make use of a series of denotation-predicates. Thus, we shall have a predicate '$\text{den}(t, x, L)$', to be read '$t$ denotes $x$ in $L$'; '$\text{den}_x(f, \varphi x, L)$', to be read '$f$ denotes the function $\varphi\xi$ in $L$'; '$\text{den}_F(M, \Phi_x Fx, L)$', to be read '$M$ denotes the (second-level) concept $\Phi_x \varphi x$ in $L$'; and so forth.[52]

For the nonlogical vocabulary, we have the following axioms.[53]

$\text{den}(\text{‘}0\text{’}, 0, L)$
$\text{den}_x(\text{‘}S\xi\text{’}, Sx, L)$
$\text{den}_{x,y}(\text{‘}\xi + \eta\text{’}, x + y, L)$
$\text{den}_{x,y}(\text{‘}\xi \times \eta\text{’}, x \times y, L)$

For the logical constants:[54]

$\text{den}_{x,y}(\text{‘}\xi = \eta\text{’}, x = y, L)$
$\text{den}_{p,q}(\text{‘}\xi \mathbin{\&} \eta\text{’}, p \mathbin{\&} q, L)$
$\text{den}_p(\text{‘}\neg\xi\text{’}, \neg p, L)$
$\text{den}_F(\text{‘}\forall v.\varphi v\text{’}, \forall x.Fx, L)$

And we need a series of compositional axioms.

| | | |
|---|---|---|
| $\text{den}(\text{‘}\varphi t\text{’}, x, L)$ | iff | $\exists f \exists y[\text{den}_z(\text{‘}\varphi\xi\text{’}, fz, L) \mathbin{\&} \text{den}(\text{‘}t\text{’}, y, L) \mathbin{\&} x = fy]$ |
| $\text{den}(\text{‘}\Phi t\text{’}, p, L)$ | iff | $\exists F \exists y[\text{den}_z(\text{‘}\Phi\xi\text{’}, Fz, L) \mathbin{\&} \text{den}(\text{‘}t\text{’}, y, L) \mathbin{\&} p \equiv Fy]$ |
| $\text{den}(\text{‘}\varphi tu\text{’}, x, L)$ | iff | $\exists g \exists v \exists w[\text{den}_{z,y}(\text{‘}\Phi\xi\eta\text{’}, gzy, L) \mathbin{\&} \text{den}(\text{‘}t\text{’}, v, L) \mathbin{\&} \text{den}(\text{‘}u\text{’}, w, L) \mathbin{\&} x = gvw]$ |
| $\text{den}(\text{‘}M_y\Phi y\text{’}, p, L)$ | iff | $\exists \Psi \exists F[\text{den}_G(\text{‘}M_y\varphi y\text{’}, \Psi_z Gz, L) \mathbin{\&} \text{den}_z(\text{‘}\Phi\xi\text{’}, Fz, L) \mathbin{\&} p \equiv \Psi_z Fz]$ |

And so forth.

These axioms, however, do not suffice: although they assign denotations to all of the *primitive* expressions, and to *some* complex expressions compounded from them, they do not suffice to assign denotations to *all* expressions. Indeed, as a little experimentation will show, they assign denotations to all and only expressions constructed in Frege's ‘first way’. We therefore require an axiom stating what the denotation of a *complex predicate* (or functional expression) is to be, that is, what the denotations of expressions constructed in Frege's ‘second way’ are. And it is at precisely this point, interestingly enough, that the auxiliary constants become important: the idea is that a complex predicate ‘$\Phi\xi$’ will be true of an object $x$ just in case $\ulcorner \Phi a_i \urcorner$ is true, when ‘$a_i$’ is taken to denote $x$. The required axiom is this:

$$\text{den}_x(\text{‘}\Phi\xi\text{’}, Fx, L) \quad \text{iff} \quad \forall x \in Dom(\mathcal{L}) \exists L' >_i L[\text{den}(\text{‘}a_i\text{’}, x, L') \mathbin{\&} \text{den}(\ulcorner \Phi a_i \urcorner, \top, L') \equiv Fx]$$

Or more concisely

$$\text{den}_x(\text{‘}\Phi\xi\text{’}, \exists L' >_i L[\text{den}(\text{‘}a_i\text{’}, x, L') \mathbin{\&} \text{den}(\ulcorner \Phi a_i \urcorner, \top, L')], L)$$

This axiom is closely related to Frege's condition for a functional expression to denote: for there will be a function denoted by ‘$\Phi\xi$’ if and only if $\ulcorner \Phi a_i \urcorner$ denotes, so long as ‘$a_i$’ denotes, no matter what; moreover, the function denoted will be that whose value, for argument $x$, is the value of $\ulcorner \Phi a_i \urcorner$, when ‘$a_i$’ denotes $x$.

It is easy enough to see that the theory just presented is again formally and materially adequate. Obviously, there are many possible variations on this theme: some of the variations may be of more interest than this one. My present purpose has simply been to show that truth-theories can be given which take Frege's semantic doctrines seriously. How *best* to formulate such a theory, for various purposes, is a question I shall not take up.

**Acknowledgments** Thanks to George Boolos, Tyler Burge, Michael Kremer, Charles Parsons, Ian Proops, Michael Resnik, Jason Stanley, and Jamie Tappenden, for discussion and reactions to earlier drafts, and for most welcome encouragement. The paper also benefitted from the comments of the referees. Special thanks are due to George, for encouraging me to write this paper in the first place, expanding on a more limited discussion of §31 in early drafts of "Frege and Semantics." Most of this material was presented in my seminar on *Grundgesetze*, held in the Spring of 1997 at Harvard University. Thanks to all who participated for their comments. An abridged version was read to the colloquium "From Frege to Wittgenstein," held at Boston University in April 1997.

## NOTES

1. Frege [7], part 1, with which I shall be concerned, was translated by Montgomery Furth and published as [13]. I shall, without comment, make some changes to Furth's translation, and to others, mostly to unify terminology (but sometimes to make the translation more readable). References will be in the text, marked '[7]', followed by a Roman numeral for the volume number and a section number.

   Frege does not clearly distinguish between his formal *language* and the formal *theory* he states in that language. I shall not, therefore, make much of this distinction, but it is worth respecting it, at least verbally. I shall therefore use the term 'the Begriffsschrift' to refer to the theory; 'Begriffsschrift', without the article, to refer to the language.

2. I have in mind such authors as Dreben, Goldfarb, Ricketts, and Weiner. See, for example, Ricketts [23], and Weiner [25]. For critical discussion, see my [15], and Stanley [24].

3. For examples of this sort of interpretation, see Resnik [22], Martin [17], and Dummett [4], pp. 215–22. For an early, and influential, discussion, see Parsons [20], pp. 159–60.

4. Frege uses lower case Greek letters, such as '$\xi$' and '$\zeta$', as placeholders, in Quine's sense: their purpose is to indicate the positions of argument-places within predicates. Frege uses '$\varphi$' similarly, but to indicate second-order argument-places. I shall follow him in this.

5. Note that Frege here uses the word "name," as elsewhere in [7], to mean simply "well-formed expression": a name could be either a proper name—which includes the sentences—or a 'function-name', a one- or more-place functional expression, of any logical type.

6. For example, to form: $\forall y(\forall x.x = x \longrightarrow Fy)$.

7. Note that the issue concerns not just whether all *functions* have names, but whether all *objects* have names. To the best of my knowledge, no one has ever claimed that Frege held this view.

8. Of course, Frege cannot be talking here about the defined symbols to be introduced later in Part I: names of these functions are already present in the language. Compare the passage from [7], I:10 quoted in note 30.

9. See the appendix for an outline of such a theory. Such a treatment of quantification is given in Mates [18], the first edition of which was published in 1965. The connection to Frege's ideas was, to the best of my knowledge, first made in Dummett [3], pp. 15–19. The potential interest of such theories for empirical semantics was first noted in Evans [5] (see especially pp. 83–87).

10. Frege does sometimes speak in a way such as that I shall be employing on his behalf. In concluding his argument that the smooth breathing denotes, he speaks of its following "universally from the fact that a name of a first-level function denotes something that the proper name resulting from its being substituted in '$\dot{\epsilon}.\varphi(\epsilon)$' denotes something": and he might just as well have said here that '$\dot{\epsilon}.\Phi(\epsilon)$' denotes, so long as '$\Phi(\xi)$' denotes, no matter what. Other such passages will be quoted below. But this is hardly conclusive.

11. A 'proposition', as Frege uses the term, is "the presentation of a judgment by use of the sign '$\vdash$' " ([7], I: 5).

12. I have slightly altered the translation.

13. I think we can safely discount the possibility that Frege is anticipating the notion of dynamic binding.

14. Here and below, we are effectively assuming that to say that a proper name denotes something just is to say that there is some object in the domain which it denotes. As we shall see below, the condition Frege specifies for when a proper name denotes is not this one, but another, much more complicated one. See note 21 and the text to which it is attached.

15. Furth makes some alterations to the text, on the basis of his understanding of what Frege is trying to say. But it is Furth's reading, and not the text, which is incorrect.

16. Consider the expression '$Ft \longrightarrow \forall x.Rxt$', which is constructed as follows: first, we form '$Ft$', by filling the argument-place of '$F\xi$' with '$t$'; '$R\xi t$', by filling the second argument-place of '$R\xi\eta$' with '$t$'; '$\forall x.Rxt$', by filling the argument-place of the universal quantifier with '$R\xi t$'; and our target, by filling the $\xi$-argument-place of '$\xi \longrightarrow \eta$' with '$Ft$', the $\zeta$-argument-place, with '$\forall x.Rxt$'. Suppose we now replace both occurrences of '$t$' with ones of '$u$': then the resulting sentence, '$Fu \longrightarrow \forall x.Rxu$', can be formed in a similar manner: we need only fill the argument-place of '$F\xi$', and the $\eta$-argument-place of '$R\xi\eta$', at the appropriate points in the construction, with '$u$' instead of '$t$'.

17. Note that Frege does indicate here that the second way can also be applied to expressions themselves formed (at least in part) in the second way. He does not argue that expressions so formed must denote, but the argument for that claim is a simple induction precisely parallel to that given for the basic case.

18. Resnik objects to this argument that the functional expression " . . . '$(x)(x = \xi)$' must be formed using the 'second way', that is, by forming the name '$(x)(x = A)$', where $A$ is an object name, and then dropping the occurrence of $A$. But there is no analogous method for obtaining the reference of the name ['$(x)(x = \xi)$']. We cannot start with the *object* $(x)(x = A)$ and then 'knock out' the object $A$ in analogy to Frege's second way of forming names" [22]. But the denotation of '$(x)(x = \xi)$' is determined by the

denotations of instances formed using the auxiliary names, not by "knocking out" objects from truth-values. Its value, for argument $\Gamma$, is to be the denotation of the expression '$(x)(x = \Gamma)$', where '$\Gamma$' denotes $\Gamma$. To put the point in Tarskian style, '$(x)(x = \xi)$' is true of $y$ if and only if '$(x)(x = y)$' is true when '$y$' is assigned $y$.

19. Let '$M_x(\varphi x)$' be a (second-level) functional expression formed from denoting expressions in the second way. It was formed by deleting occurrences of some (first-level) functional expression '$\Phi\xi$' from '$M_x(\Phi x)$', itself formed in the first way. '$M_x(\Phi x)$' certainly denotes: moreover, '$M_x(\Psi x)$' also denotes, if '$\Psi\xi$' denotes, since '$M_x(\Psi x)$' could have been formed in the first way, and any expression so formed from denoting expressions denotes. Hence, '$M_x(\varphi x)$' denotes.

20. If one would like more evidence for this claim, note Frege's remark in [7], I:32, that "by our stipulations it is determined under what conditions [a name of a truth-value] denotes the True." Note, too, the discussion in [7], I:30, of the difficulties raised by primitive expressions with multiple $\xi$-argument-places: the worry is that "an explanation of the denotation for this case would be lacking," that is, that no stipulation would have been made about what the denotation of an expression of a particular form is to be.

21. It is for this reason that one can, for much of the time, ignore the actual condition Frege specifies regarding when a proper name denotes. One might have thought that, instead, Frege was speaking of a syntactic category of expressions, that he is assuming that sentences always denote. But this cannot be right. Some of his arguments accord with such a reading. But his discussion of the universal quantifier does not. Anything of the form '$\forall x.\Phi(x)$' is certainly a sentence: so, if what Frege were assuming was that all sentences denote, he could simply have noted that fact and been done with it. But he does not proceed in that way.

22. The question whether there are any actual, primitive terms in *Begriffsschrift itself* which denote the True and the False is beside the point, for the argument involves the consideration of expansions of the language of the theory.

23. The notion of adequacy here is that appropriate to a theory of truth given for a previously *uninterpreted* language, a theory of truth which itself *specifies* the interpretation the language is to have. In such a case, we do not seek a theory which generates T-sentences whose right-hand side is a translation of the sentence mentioned on the left-hand side: there is no such sentence, prior to the formulation of the semantic theory. What we seek is just a theory which generates T-sentences, whatever these may be, for all sentences of the language—one which, that is, suffices to assign truth-conditions to all sentences of the language. This is what Frege has shown his semantic theory does.

24. Frege does make one explicit statement about what objects the domain of his theory is to contain, namely, the truth-values: but he thinks he has an argument that so much as engaging in the practice of judgment commits us to the existence of *those* objects. See here Frege [9], p. 163, op. 34: "Every assertoric sentence concerned with what its words refer to is therefore to be regarded as a proper name, and its reference, if it has one, is either the True or the False. These two objects are recognized, if only implicitly, by everybody who judges something to be true—and so even by a sceptic." Part of the reason he hopes not to have to make any further such statements is that he has no similar argument that thought, or reasoning, does commit us to the existence of value-ranges: at best, certain prevalent forms of mathematical reasoning are committed to their existence. (Frege is fond of making this point: see, for example, [7], II:147.)

25. The full context is: “If I say generally that ‘$\dot{\epsilon}.\Phi(\epsilon)$’ denotes the value-range of the function $\Phi(\xi)$ then this requires a supplementation like that in §8 above, in our explanation of ‘$\forall x.\Phi(x)$’; that is, the question is which function in each case is to be regarded as the *corresponding* function $\Phi(\xi)$.” One might have thought that the conditional character of the sentence suggested that Frege did not intend to endorse this stipulation. But that is wrong: all Frege is saying is that the stipulation requires supplementation. This passage is exactly parallel to one in §8, where he specifies the interpretation of the universal quantifier: “If we now set up the definition as follows: ‘$\forall x.\Phi(x)$’ is to denote the True if for every argument the value of the function $\Phi(\xi)$ is the True, and otherwise is to denote the False, a supplemenation is required . . . ” ([7], I:8). Thus, Frege really is endorsing this stipulation.

26. I do not mean to say that Frege does not take us to have *any* understanding of what value-ranges are other than what is provided by the explanation in §3. My point here is that the *argument* depends only upon that explanation. One might put this point by saying that Frege's remarks about how we are to *think* about value-ranges, on the analogy with extensions of concepts, really are ‘elucidatory’.

27. See, e.g., [22], p. 187. I take it that this is how Dummett reads the proof, too.

28. Moreover, if we have to show that ‘$\dot{\alpha}.(\alpha = \dot{\epsilon}.\Phi(\epsilon))$’ denotes to conclude that ‘$\dot{\epsilon}.\Phi(\epsilon)$’ does, we should also have to show that ‘$\dot{\eta}.[\eta = \dot{\alpha}.(\alpha = \dot{\epsilon}.\Phi(\epsilon))]$’ denotes to show that ‘$\dot{\alpha}.(\alpha = \dot{\epsilon}.\Phi(\epsilon))$’ does, and the induction would not be well founded.

29. Indeed, one might well worry that the sort of argument Frege would have to give for the claim that ‘$\Psi(\dot{\epsilon}.\Phi(\epsilon))$’ denotes, whenever ‘$\Psi(\xi)$’ is a denoting expression of Begriffsschrift, would *already* establish that every well-formed expression denotes, so that the second induction would be unnecessary. For let A be some sentence (say) and consider ‘$A = \dot{\epsilon}.\Phi(\epsilon)$’. Clearly, this cannot denote unless $A$ does, and ‘$A = \xi$’ is a perfectly good functional expression.

30. What if the langauge is expanded? “As soon as there is a further question of introducing a function that is not completely reducible to functions known already, we can stipulate what value it is to have for value-ranges as arguments; and this can then be regarded as much as a further determination of the value-ranges as of that function” ([7], I:10).

31. It is worth remembering at points like this one that Frege almost always uses the word ‘definition’ in such contexts in the sense of a *formal* definition, within a formal language.

32. My discussion here owes some of its spirit to [21].

33. Thus, ‘$\xi = \xi$’ is a composite expression of rank 0, formed in the second way from ‘$-\epsilon = -\epsilon$’, itself a simple expression of rank 0; so ‘$\forall x(x = x)$’ and ‘$\dot{\epsilon}.\epsilon = \epsilon = \forall x(x = x)$’ are simple expressions of rank 1.

34. In fact, matters are worse. If we do not know what the domain is, we do not know what assignments can be made to ‘$F\xi$’ in the first place. Frege hoped that fixing the truth-values of identity-statements containing fair value-range terms would fix their denotations and *thereby* fix a domain—that comprising the denotations of fair value-range terms. Unfortunately, the fair value-range terms in question have to contain auxiliary functional expressions: that is, the domain has to contain the denotation of ‘$\dot{\epsilon}.\Phi(\epsilon)$’, for any assignment to ‘$\Phi(\xi)$’. But, once again, we don't know what assignments we can make to ‘$\Phi(\xi)$’ unless we already know what the domain is.

35. As it stands, the proof does not work for the *predicative* fragment either. However, the methods used in my [16] could be used to extend such a proof to cover the predicative fragment.

36. Obviously, the symmetry of identity will take care of expressions of the form '$\dot{\epsilon}.\Phi(\epsilon) = \xi$'.

37. For, if the True is $\dot{\epsilon}.(\epsilon = \forall x(x = x))$, then an identity of the form '$\top = \dot{\epsilon}.\Phi(\epsilon)$' will have the same denotation as '$\dot{\epsilon}.(\epsilon = \forall x(x = x)) = \dot{\epsilon}.\Phi(\epsilon)$'.

    It is unclear whether these stipulations do together suffice to determine the truth-values of *all* identity-statements. Let $JC$ be a value-range which is its own unit class, and of which is not already known whether it is the True. Then $JC$ is the True if, and only if:

$$(\dot{\epsilon}.\epsilon = JC) = (\dot{\epsilon}. - \epsilon)$$

    The semantical stipulation governing the smooth breathing tells us that this sentence is true if, and only if

$$\forall x(x = JC \equiv -x)$$

    And this, in turn, will be true if and only if $JC$ is the True. So if we do not know whether $JC$ is the True, we cannot proceed any further. More generally, the sort of stipulation we are considering will certainly tell us that the True is not the same as any object which is *not* a unit class. But it will not allow us to distinguish it from any other object which *is* identical with its own unit class, unless we could already distinguish it from that object.

38. Moore and Rein [19] are actually talking about aspects of Frege's proof in [7], I:10, but the point I am about to make applies equally to both cases. (I should add, in fairness, that Moore and Rein are hardly the only commentators to have made this sort of remark.)

39. There is a way of attempting to reconcile this view with what is about to be argued. If objects such as Caesar were identified with their unit classes, the domain could contain them and *still* contain only truth-values and value-ranges. But, first of all, although Frege considers such an identification, he does not endorse it: he argues that no general principle has been coherently formulated. The problem is to say what 'objects such as Caesar' are; one can hardly say that objects which are not value-ranges are their own unit classes, since they then would be value-ranges! The best one can do is say that such objects as are not given as value-ranges are to be identified with their unit-classes, but "it is intolerable to allow this to hold only for such objects as are not given us as value-ranges; the way in which an object is given must not be regarded as an immutable property of it, since the same object can be given in different ways" ([7], I:10; see also [14], 67).

    Moreover, consider the following passage from Frege [12], letter XV/8 (xxxvi/8, in the German edition), at p. 142:

    > You [Russell] ask how it can be known that something is a value-range. This is indeed a difficult point. Now, all objects of arithmetic are introduced as value-ranges. Whenever a new object to be considered is not introduced as a value-range, we must at once answer the question whether it is a value-range, and the answer is probably always no, since it would have been introduced as a value-range if it was one.

    It would follow from this remark that the objects of geometry are probably *not* value-ranges—if, as seems plausible, Frege would not have introduced them as such, since they are not logical objects, but objects known by intuition.

40. It is also worth mentioning that open formulas of the form '$x = \dot{\epsilon}.F(\epsilon)$' play an important role in Frege's theory, most notably, perhaps, in the most fundamental definition Frege makes, that of the application-operator, which is the analogue, for value-ranges, of the notion of membership, as applied to extensions of concepts:

$$x \frown u =_{\mathrm{df}} \backslash \dot{\alpha}.\exists F(u = \dot{\epsilon}.F(\epsilon)\ \&\ \alpha = Fx)$$

Note the occurrence of '$u = \dot{\epsilon}.F(\epsilon)$' here. The point of the definition is revealed by Frege's theorem 1, which is a version of naïve abstraction:

$$x \frown \dot{\epsilon}.F(\epsilon) = Fx$$

Now, as Russell notes, we cannot infer '$u = v$' from '$\forall x(x \frown u = x \frown v)$', but only if $u$ and $v$ are value-ranges themselves. And similarly, we do not, in general have '$u = \dot{\epsilon}.\epsilon \frown u$', but, again, only if $u$ is a value-range. It is this that leads him to ask how it can be known if something is a value-range (see [12], letter XV/6 (xxxvi/6), p. 139), to which question Frege responds in the passage quoted in note 39. Frege himself remarks, as well, that if $u$ is not value-range, then $a \frown u$ is the empty class ([7], I:34). So it seems clear that he was forced, for technical reasons, to think about the predicate mentioned in the text, or at least about the predicate '$\xi = \dot{\epsilon}.F(\epsilon)$'.

41. It is by now uncontroversial that Frege's discussion here, although framed in terms of a similar principle governing names of directions, is to be applied *mutatis mutandis* to the stipulation governing names of numbers which he introduces in [14], 63. I have therefore silently transposed Frege's discussion, omitting all the square brackets, as their inclusion would make the passage almost unreadable.

42. Note that, if the *only* things in the domain are value-ranges, only value-ranges can be members of value-ranges. This might suggest that the domain is to contain only what we might call 'pure' value-ranges, analogously with the pure sets of set-theory. On the other hand, if Caesar is his unit class, then he *is* a value-range whose only member is a value-range, so he is still not excluded.

What then *does* Frege intend to do about Caesar? Frege might have hoped he could be left to take care of himself. Whether he is in or out will depend upon whether he is a value-range, but it may not matter which: if he's out, we don't have to worry about him; if he's in, the semantical stipulation govening the smooth breathing will assign denotations to all identity statements concerning him. There are problems with this line of thought: for example, if Caesar is identical with his own unit class, then the question whether he is one of the truth-values, and if so which one he is, is not decided by any of the stipulations made in *Grundgesetze*. (See note 40.) But perhaps that worry could be addressed.

43. For then '$t = \dot{\epsilon}.\Phi(\epsilon)$' will be true if and only if '$\dot{\epsilon}.\Psi(\epsilon) = \dot{\epsilon}.\Phi(\epsilon)$' is true, when '$\Psi(\xi)$' denotes $f(\xi)$ and also when '$\Psi(\xi)$' denotes $g(\xi)$—that is, if and only if '$\forall x(\Psi(x) = \Phi(x))$' is true when '$\Psi(\xi)$' denotes $f(\xi)$, but also when '$\Psi(\xi)$' denotes $g(\xi)$. But, by hypothesis, '$\forall x(\Psi(x) = \Phi(x))$' *is* true when '$\Psi(\xi)$' denotes $f(\xi)$, but false when '$\Psi(\xi)$' denotes $g(\xi)$.

44. And it would remain unclear whether a similar objection would apply to a version of Frege's argument given for the *first-order* fragment of the theory. Moreover, it would remain unclear whether Frege's argument could be salvaged if we allowed nonstandard models of second-order logic. We ought not to rest with an understanding of why there is no *model* of Frege's theory: for that does not, on its own, show why the theory is

*syntactically* inconsistent. Compare Boolos [1]. What Boolos ought to have said is that Cantor's Theorem is *provable* in second-order logic and that Axiom V is inconsistent with it.

45. That is, which is such that, for every existential formula '$\exists x.A(x)$', there is a term $t$ such that the theory proves: $(\exists x)A(x) \longrightarrow A(t)$.

46. The importance of this point was made clear to me by Resnik. See [22], pp. 189–91.

47. Frege [12], p. 132, letter 36/2. Note how Frege here distinguishes Russell's having shown that Axiom V is false from his having shown that the semantical stipulation governing the smooth breathing is illegitimate.

48. The extension to second-order languages poses no difficulty and is left to the reader. No use will be made of the fact that this language contains a name for every object in the domain.

49. Thus,

$$\begin{aligned}&\text{true}(\ulcorner(\text{Most } v)(A(v); B(v))\urcorner, L) \quad \text{iff}\\ &\quad (\text{Most } x \in \text{Dom}(\mathcal{L}))\{\exists L' >_i L[\text{den}(a_i, L') = x \;\&\; \text{true}(\ulcorner A(a_i)\urcorner, L')];\\ &\quad \exists L' >_i L[\text{den}(a_i, L') = x \;\&\; \text{true}(\ulcorner B(a_i)\urcorner, L')]\}\end{aligned}$$

50. This assumption can be traded in for complications elsewhere in the theory.

51. Note that the theory does not tell us what the denotations of expressions containing auxiliary constants are in languages to which those constants do not belong. It does not matter what these are taken to be.

52. We assume—that is, adopt axioms to the effect—that no primitive expression denotes more than one entity.

53. An alternative theory would make use of auxiliary constants in the axioms for functional expressions. Thus, compressing a bit, the axiom for successor would read

$$\text{den}(\ulcorner Sa_i\urcorner, S(\iota x.\text{den}(a_i, x, L)), L);$$

that for identity

$$\text{den}(\ulcorner a_i = a_j\urcorner, \iota x.\text{den}(a_i, x, L) = \iota x.\text{den}(a_j, x, L), L)$$

The denotations of the functional expressions themselves would then be given only by the axiom to be discussed below in connection with complex predicates. This would be closer to Frege's own method in *Grundgesetze*, as he uses auxiliary constants in the presentations of the semantical stipulations themselves.

54. '$p$' and '$q$' are zero-place second-order variables.

## REFERENCES

[1] Boolos, G., "Whence the contradiction?," *Proceedings of the Aristotelian Society*, sup. vol. 67 (1993), pp. 213–33. Zbl 0961.03529

[2] Dummett, M., "Frege on the consistency of mathematical theories," pp. 1–16 in *Frege and Other Philosophers*, Clarendon Press, Oxford, 1991. MR 58:21379

[3] Dummett, M., *Frege: Philosophy of Language*, Harvard University Press, Cambridge, 1973.

[4] Dummett, M., *Frege: Philosophy of Mathematics*, Harvard University Press, Cambridge, 1991.

[5] Evans, G., "Pronouns, quantifiers, and relative clauses (I)," pp. 76–153 in *Collected Papers*, Clarendon Press, Oxford, 1985.

[6] Frege, G., "Function and concept," pp. 137–56 in *Collected Papers*, translated by P. Geach, edited by B. McGuinness, Basil Blackwell, Oxford, 1984.

[7] Frege, G., *Grundgesetze der Arithmetik*, Georg Olms, Hildesheim, 1962. MR 35:2715

[8] Frege, G., "On Mr. Peano's conceptual notation and my own," pp. 234–48 in *Collected Papers*, translated by V. H. Dudman, edited by B. McGuinness, Basil Blackwell, Oxford, 1984.

[9] Frege, G., "On sense and reference," pp.157–77 in *Collected Papers*, translated by M. Black, edited by B. McGuinness, Basil Blackwell, Oxford, 1984.

[10] Frege, G., "On the foundations of geometry: first series," pp. 273–84 in *Collected Papers*, edited by B. McGuinness, Basil Blackwell, Oxford,1984.

[11] Frege, G., "On the foundations of geometry: second series," pp. 293–40 in *Collected Papers*, edited by B. McGuinness, Basil Blackwell, Oxford, 1984.

[12] Frege, G., *Philosophical and Mathematical Correspondence*, edited by G. Gabriel et al., translated by H. Kaal, University of Chicago Press, Chicago, 1980. Zbl 0502.03002 MR 81d:03001

[13] Frege, G., *The Basic Laws of Arithmetic: Exposition of the System*, translated by M. Furth, University of California Press, Berkeley, 1964. Zbl 0155.33601 MR 31:1173

[14] Frege, G., *The Foundations of Arithmetic*, translated by J. L. Austin, Northwestern University Press, Evanston, 1980. Zbl 0037.00602 MR 50:4227

[15] Heck, Jnr., R., "Frege and semantics," in *The Cambridge Companion to Frege*, edited by T. Ricketts, Cambridge University Press, Cambridge, forthcoming.

[16] Heck, Jnr., R., "The consistency of predicative fragments of Frege's *Grundgesetze der Arithmetik*," *History and Philosophy of Logic*, vol. 17 (1996), pp. 209–20. Zbl 0876.03032 MR 98i:03079

[17] Martin, E., "Referentiality in Frege's *Grundgesetze*," *History and Philosophy of Logic*, vol. 3 (1982), pp. 151–64. MR 84g:03005

[18] Mates, B., *Elementary Logic*, 2d edition, Oxford University Press, Oxford, 1972. Zbl 0146.24601 MR 48:45

[19] Moore, A., and A. Rein, "*Grundgesetze* Section 10," pp. 375–84 in *Frege Synthesized* edited by L. Haaparanta and J. Hintikka, Dordrecht Reidel, Boston, 1986.

[20] Parsons, C., “Frege’s theory of number,” pp. 150–75 in *Mathematics in Philosophy*, Cornell University Press, Ithaca, 1983. Zbl 0900.03011 MR 1 376 396

[21] Parsons, T., “On the consistency of the first-order portion of Frege’s system,” *Notre Dame Journal of Formal Logic*, vol. 28 (1987), pp. 161–68. Zbl 0637.03005 MR 88h:03002

[22] Resnik, M., “Frege’s proof of referentiality,” pp. 177–95 in *Frege Synthesized*, edited by L. Haaparanta and J. Hintikka, Dordrecht Reidel, Boston, 1986.

[23] Ricketts, T., “Objectivity and objecthood: Frege’s metaphysics of judgement,” pp. 65–95 in *Frege Synthesized*, edited by L. Haaparanta and J. Hintikka, Dordrecht Reidel, Boston, 1986.

[24] Stanley, J., “Truth and meta-theory in Frege,” *Pacific Philosophical Quarterly*, vol. 77 (1996), pp. 45–70.

[25] Weiner, J., *Frege in Perspective*, Cornell University Press, Ithaca, 1990. Zbl 0937.03500 MR 92m:03009

*Department of Philosophy*
*Harvard University*
*Emerson Hall*
*Cambridge, MA 02138*
*email: heck@fas.harvard.edu*