

CHAPTER 1



Foundations of Logic, Language, and Mathematics

1. Overview
2. The Language of Logic and Mathematics
3. Sense, Reference, Compositionality, and Hierarchy
4. Frege's Logic
5. Frege's Philosophy of Mathematics
 - 5.1. Critique of Naturalism, Formalism, and Psychologism
 - 5.2. Critique of Kant
 - 5.3. Frege's Definition of Number
 - 5.3.1. Numerical Statements Are about Concepts
 - 5.3.2. But Numbers Are Objects
 - 5.3.3. Objects and Identity
 - 5.3.4. The Number of F's, Zero, Successor, and the Numerals
 - 5.3.5. The Natural Numbers
6. The Logicist Reduction
 - 6.1. The Axioms of Logic and Arithmetic
 - 6.2. Informal Proofs of the Arithmetical Axioms
 - 6.3. Arithmetical Operations
 - 6.4. Further Issues

1. OVERVIEW

The German philosopher-logician Gottlob Frege was born in 1848, graduated with a PhD in Mathematics from the University of Göttingen in 1873, and earned his Habilitation in Mathematics from the University of Jena in 1874, where he taught for 43 years until his retirement in 1917, after which he continued to write on issues in philosophical logic and the philosophy of mathematics until his death in 1925. While he is now recognized as one of the greatest philosophical logicians, philosophers of mathematics, and philosophers of language of all time, his seminal achievements in these areas initially elicited little interest from his contemporaries in mathematics. Though he did attract the attention of, and have an important influence on, four young men—Bertrand Russell,

Edmund Husserl, Rudolf Carnap, and Ludwig Wittgenstein—who were to become giants in twentieth-century philosophy, it took several decades after his death before the true importance of his contributions became widely recognized.

Frege's main goal in philosophy was to ground the certainty and objectivity of mathematics in the fundamental laws of logic, and to distinguish both logic and mathematics from empirical science in general, and from the psychology of human reasoning in particular. His pursuit of this goal can be divided into four interrelated stages. The first was his development of a new system of symbolic logic, vastly extending the power of previous systems, and capable of formalizing the notion of proof in mathematics. This stage culminated in his publication of the *Begriffsschrift* (*Concept Script*) in 1879. The second stage was the articulation of a systematic philosophy of mathematics, emphasizing (i) the objective nature of mathematical truths, (ii) the grounds for certain, a priori knowledge of them, (iii) the definition of number, (iv) a strategy for deriving the axioms of arithmetic from the laws of logic plus analytical definitions of basic arithmetical concepts, and (v) the prospect of extending the strategy to higher mathematics through the definition and analysis of real, and complex, numbers. After the virtual neglect of the *Begriffsschrift* by his contemporaries—due in part to its forbidding technicality and idiosyncratic symbolism—Frege presented the second stage of his project in remarkably accessible, and largely informal, terms in *Die Grundlagen der Arithmetik* (*The Foundations of Arithmetic*), published in 1884. In addition to being among the greatest treatises in the philosophy of mathematics ever written, this work is one of the best examples of the clarity, precision, and illuminating insight to which work in the analytic tradition has come to aspire. The third stage of the project is presented in a series of ground-breaking articles, starting in the early 1890s and continuing at irregular intervals throughout the rest of his life. These articles include, most prominently, “Funktion und Begriff” (“Function and Concept”) in 1891, “Über Begriff und Gegenstand” (“On Concept and Object”) in 1892, “Über Sinn und Bedeutung” (“On Sense and Reference”) in 1892, and “Der Gedanke” (“Thought”) in 1918. In addition to elucidating the fundamental semantic ideas needed to understand and precisely characterize the language of logic and mathematics, this series of articles contains important insights about how to extend those ideas to natural languages like English and German, thereby providing the basis for the systematic study of language, thought, and meaning. The final stage of Frege's grand project is presented in his treatise *Grundgesetze der Arithmetik* (*Basic Laws of Arithmetic*), volumes 1 and 2, published in 1893 and 1903 respectively. In these volumes, Frege meticulously and systematically endeavors to derive arithmetic from logic together with definitions of arithmetical concepts in purely logical terms. Although, as we shall see, his attempt was not entirely successful, the project has proven to be extraordinarily fruitful.

The discussion in this chapter will not strictly follow the chronological development of Frege's thought. Instead, I will begin with his language of logic and mathematics, which provides the starting point for developing his general views of language, meaning, and thought, and the fundamental notions—truth, reference, sense, functions, concepts, and objects—in terms of which they are to be understood. With these in place, I will turn to a discussion of the philosophical ideas about mathematics that drive his reduction of arithmetic to logic, along with a simplified account of the reduction itself. The next chapter will be devoted to critical discussions of Frege's most important views, including the interaction between his philosophy of language and his philosophy of mathematics. In what follows I refer to Frege's works under their English titles—with the exception of the *Begriffsschrift*, the awkwardness of the English translation of which is prohibitive.

2. THE LANGUAGE OF LOGIC AND MATHEMATICS

I begin with the specification of a simple logical language which, though presented in a more convenient symbolism than the one Frege used, is a direct descendant of his. The first step is to specify how the formulas and sentences of the language are constructed from the vocabulary of the language. After that, we will turn to Fregean principles for understanding the language.

THE SYMBOLIC LANGUAGE L_F

Vocabulary

Names of objects: a, b, c, \dots

Function signs: $f(), g(), h(), f'(,), g'(,) h'(, ,), \dots$ These stand for functions from objects to objects. Function signs are sorted into 1-place, 2-place, \dots , and n -place. One-place function signs combine with a single name (or other term) to form a complex term, 2-place function signs combine with a pair of names (or other terms) to form a complex term, and so on. Standardly, the terms follow the function sign, but in the case of some 2-place function signs—like '+' and '×' for addition and multiplication—the function symbol is placed between the terms.

Predicate signs: $() = (), P(), Q(,), R(, ,) \dots$ Predicate signs are sorted into 1-place, 2-place, etc. An n -place predicate sign combines with n terms to form a formula.

Terms

Individual variables (ranging over objects) are terms: $x, y, z, x', y', z', \dots$

Names of objects are terms: a, b, c, \dots

Expressions in which an n -place function sign is combined with n terms are terms: e.g., if a and b are terms, f and h are 1-place function signs, and g is a 2-place function sign, then $\lceil f(a) \rceil$, $\lceil g(a,b) \rceil$, $\lceil h(f(a)) \rceil$, and $\lceil g(a,f(b)) \rceil$ are terms.

Definite descriptions are terms: If Φv is a formula containing the variable v , then $\lceil \text{the } v \Phi v \rceil$ is a term.

Nothing else is a term.

Formulas

An atomic formula is the combination of an n -place predicate sign with n terms. Standardly the terms follow the predicate sign, but in the case of some 2-place predicate signs—like $\lceil () = () \rceil$ for identity—the terms are allowed to flank predicate sign.

Other (non-atomic) formulas

If Φ and Ψ are formulas, so are $\lceil \sim\Phi \rceil$, $\lceil (\Phi \vee \Psi) \rceil$, $\lceil (\Phi \& \Psi) \rceil$, $\lceil (\Phi \rightarrow \Psi) \rceil$ and $\lceil (\Phi \leftrightarrow \Psi) \rceil$. If v is a variable and $\Phi(v)$ is a formula containing an occurrence of v , $\lceil \forall v \Phi(v) \rceil$ and $\lceil \exists v \Phi(v) \rceil$ are also formulas. (Parentheses can be dropped when no ambiguity results.)

$\lceil \sim\Phi \rceil$, which is read or pronounced $\lceil \text{not } \Phi \rceil$, is the negation of Φ ; $\lceil (\Phi \vee \Psi) \rceil$, read or pronounced $\lceil \text{either } \Phi \text{ or } \Psi \rceil$, is the disjunction of Φ and Ψ ; $\lceil (\Phi \& \Psi) \rceil$, read or pronounced $\lceil \Phi \text{ and } \Psi \rceil$, is the conjunction of Φ and Ψ ; $\lceil (\Phi \rightarrow \Psi) \rceil$, read or pronounced $\lceil \text{if } \Phi, \text{ then } \Psi \rceil$, is a conditional the antecedent of which is Φ and the consequent of which is Ψ ; $\lceil (\Phi \leftrightarrow \Psi) \rceil$, read or pronounced $\lceil \Phi \text{ if and only if } \Psi \rceil$, is a biconditional connecting Φ and Ψ ; $\lceil \forall v \Phi(v) \rceil$, read or pronounced $\lceil \text{for all } v \Phi(v) \rceil$, is a universal generalization of $\Phi(v)$; and $\lceil \exists v \Phi(v) \rceil$, which is read or pronounced $\lceil \text{at least one } v \text{ is such that } \Phi(v) \rceil$, is an existential generalization of $\Phi(v)$. $\forall v$ and $\exists v$ are called “quantifiers.”

Sentences

A sentence is a formula that contains no free occurrences of variables. An occurrence of a variable is free iff it is not bound.

An occurrence of a variable in a formula is bound iff it is within the scope of a quantifier, or the definite description operator, using that variable.

The scope of an occurrence of a quantifier $\forall v$ and $\exists v$, or of the definite description operator, *the v* , is the quantifier, or description operator, together with the (smallest complete) formula immediately following it. For example, $\forall x (Fx \rightarrow Gx)$ and $\exists x (Fx \& Hx)$ are each sentences, since both occurrences of ‘ x ’ in the formula attached to the quantifier are within the scope of the quantifier. Note, in these sentences, that (i) Fx does not *immediately* follow the quantifiers because ‘ $($ ’ intervenes, and

(ii) (Fx is not a *complete* formula because it contains ‘(‘ without an accompanying ‘)’). By contrast, $(\forall x Fx \rightarrow Gx)$ and $(\forall x (Fx \& Hx) \rightarrow Gx)$ are not sentences because the occurrence of ‘x’ following ‘G’ is free in each case. The generalization to ‘the x’ is straightforward.

Frege’s representational view of language provides the general framework for interpreting L_F . On this view, the central semantic feature of language is its use in representing the world. For a sentence S to be meaningful is for S to represent the world as being a certain way—which is to impose conditions the world must satisfy if it is to be the way S represents it to be. Since S is true iff (i.e., if and only if) the world is the way S represents it to be, these are the *truth conditions* of S. To sincerely accept, or assertively utter, S is, very roughly, to believe, or assert, that these conditions are met. Since the truth conditions of a sentence depend on its grammatical structure plus the representational contents of its parts, interpreting a language involves showing how the truth conditions of its sentences are determined by their structure together with the representational contributions of the words and phrases that make them up. There may be more to understanding a language than this—even a simple logical language like L_F constructed for formalizing mathematics and science—but achieving a compositional understanding of truth conditions is surely a central part of what is involved.

With this in mind, we apply Fregean principles to L_F . Names and other singular terms designate objects; sentences are true or false; function signs refer to functions that assign objects to the n-tuples that are their arguments; and predicates designate concepts—which are assignments of truth values to objects (i.e., functions from objects to truth values). A term that consists of an n-place function sign f together with n argument expressions designates the object that the function designated by f assigns as value to the n-tuple of referents of the argument expressions. Similarly, a sentence that consists of an n-place predicate P plus n names is true iff the names designate objects o_1, \dots, o_n and the concept designated by P assigns these n objects (taken together) the value “the True,” or truth. According to Frege, concepts are also designated by truth-functional operators. The negation operator ‘ \sim ’ designates a function from falsity to truth (and from truth to falsity), reflecting the fact that the negation of a sentence is true (false) iff the sentence negated is false (true); the operator ‘&’ for conjunction designates a function that assigns truth to the pair consisting of truth followed by truth (and assigns falsity to every other pair), reflecting the fact that a conjunction is true iff both conjuncts are; the disjunction operator ‘ \vee ’ designates a function that assigns truth to any pair of arguments one of which is truth, reflecting the fact that a disjunction is true iff at least one of its disjuncts is. The operator ‘ \rightarrow ’ used to form what are called *material conditionals* designates the function that assigns falsity to the pair of arguments the first of which is truth and the second

of which is falsity—capturing the fact that a material conditional $\lceil \Phi \rightarrow \Psi \rceil$ is false whenever its antecedent Φ is true and its consequent Ψ is false. The material conditional, employed in the Fregean logical language, is true on every other assignment of truth values to Φ and Ψ . Finally, the biconditional operator ‘ \leftrightarrow ’ designates a function that assigns truth to the pairs of $\langle \text{truth}, \text{truth} \rangle$ and $\langle \text{falsity}, \text{falsity} \rangle$, while assigning falsity to the other two pairs, thereby ensuring that $\lceil \Phi \leftrightarrow \Psi \rceil$ is equivalent to $\lceil (\Phi \rightarrow \Psi) \ \& \ (\Psi \rightarrow \Phi) \rceil$.

Despite Frege’s use of the term ‘concept’—which sounds as if it stands for an idea or other mental construct—concepts, in the sense he uses the term, are no more mental than the people, places, or other objects that are the referents of proper names. Just as different people who use the name ‘Boston’ to refer to the city in Massachusetts may have different images of, or ideas about, it, so the predicate ‘is a city’ may bring different images or ideas to the minds of different people who predicate it of Boston. For Frege, understanding the predicate involves knowing that it designates a concept that assigns truth to an object o iff o is a city, which, in effect, amounts to knowing that to *say of* o “it’s a city” is to say something true just in case o is a city. The truth or falsity of such a statement depends on objective features of o to which the function designated by ‘is a city’ is sensitive. Thus, Frege takes concepts to be genuine constituents of mind-independent reality.

This example brings out a related feature of Frege’s view. Just as predicates and function signs are different kinds of linguistic expressions than names, and other (singular) terms, so, Frege thinks, concepts and other functions are different kinds of things than objects. On the linguistic side, Frege begins with two grammatical categories of what he calls “saturated expressions.” These are sentences and (singular) terms—all of which he, idiosyncratically, calls “Names.”¹ What he calls “Names” are said to refer to objects, including sentences that are said to refer to truth values “the True” and “the False.” In addition, there are different types of “unsaturated expressions”—each of which is thought of as containing one or more gaps, to be filled by expressions of various types in order to produce a “saturated expression”—i.e., a singular term or a sentence. An n -place predicate, for example, is an expression that combines with n terms to form a sentence. For Frege, these include not only those that are called simple “predicate signs” in the specification of the language above, but also compound expressions that result from removing n terms from a sentence, no matter how complex.

¹The singular terms in L_F include ordinary names, definite descriptions $\lceil \text{the } \nu \Phi \nu \rceil$ and expressions formed by combining an n -place function sign with n terms. In the next section I will explain why Frege came to view sentences as designating truth values in much the way that singular terms designate their referents—and so as being a kind of “Name.” For now, however, I mostly ignore this complication.

For example, starting with the sentence ‘Cb & Lbm’, stating that Boston is a city and Boston is in Massachusetts, we may provide analyses that break it into parts in several different, but *equivalent*, ways. It may be analyzed (i) as the conjunction of a sentence formed by combining the one-place predicate ‘C()’—designating the concept that assigns the value the True to an object iff that object is a city—with the name ‘b’, and another sentence formed by combining the two-place predicate L(,)—designating the concept that assigns the True to a pair iff the first is located at the second—with the names ‘b’, and ‘m’; (ii) as a sentence that results from combining the one-place predicate ‘C() & Lbm’—designating the concept that assigns the True to an object iff it is a city and Boston is located in Massachusetts—with the name ‘b’; (iii) as a sentence that results from combining the one-place predicate ‘C() & L(, m)’—designating the function that assigns the True to an object iff it is a city located in Massachusetts—with the name ‘b’; (iv) as a sentence that results from combining the one-place predicate ‘Cb & L(b,)’—designating the concept that assigns the True to an object iff the object is a place, Boston is a city, and Boston is located at that place; (v) as a sentence that results from combining the two-place predicate ‘Cb & L(,)’—designating the concept that assigns a pair of objects the value the True iff Boston is a city and the second object is a place at which the first is located—with the names ‘b’ and ‘m’; and (vi) as a sentence that results from combining the two-place predicate ‘C() & L(,)’—designating a concept that assigns the True to a pair of objects iff the first is a city and the second is a place at which it is located—with the names ‘b’ and ‘m’.²

The fact that concepts are referents of predicates (themselves regarded as “unsaturated” expressions requiring completion), plus the fact that there is no explaining what concepts are except as intermediaries that assign truth values to objects, led Frege to distinguish concepts from objects, taking them to be “incomplete” or “unsaturated” in some manner thought to parallel the way in which predicates are supposed to be incomplete. The same conclusion is drawn for (i) concepts designated by truth-functional operators, the *arguments* (along with the values) of which are truth values, and (ii) functions designated by function signs that combine with (singular) terms to form complex (singular) terms, the *values* of which are ordinary objects (rather than truth values). As we shall see, these distinctions don’t preclude some higher-order functions (and concepts) from taking other functions (or concepts) as arguments. There are such higher-order functions/concepts, which take lower-order functions/concepts as arguments. Corresponding to these function-argument combinations are sentences, or singular terms, formed from expressions each of which designate concepts, or functions, rather than (what Frege calls) “objects.”

² Underlining in clauses (iii) and (vi) indicates that the empty positions are linked, and so to be filled by the same Name.

Examples include sentences containing quantifiers, or the definite description operator. Consider the sentence $[\exists x \Phi x]$, where Φx is a formula (no matter how complex) in which ‘x’ and only ‘x’ occurs free. To say that only ‘x’ occurs free in Φx is to say that Φx counts as a one-place predicate for Frege, since it is an expression which, when combined with a name ‘a’ (in the sense of replacing the free occurrences of ‘x’ with ‘a’), would form a sentence.³ Thus, Φx designates a concept that assigns the True or the False to an object as argument (corresponding to whether the sentence that would result from replacing free occurrences of ‘x’ with a name of the object would be true or false). It is this concept, C_Φ , that combines with the referent of the quantifier to determine a truth value. Given this much, one can easily see what Frege’s treatment of the quantifier ‘ $\exists x$ ’ had to be. On his analysis, it designates the second-order concept, C_\exists , which takes a first-order concept C as argument and assigns it the True iff C assigns truth to some object or objects (at least one). Thus, $[\exists x \Phi x]$ is true iff there is at least one (existing) object o such that *the concept designated by Φx assigns o the value the True iff o “satisfies” the formula Φx iff replacing occurrences of ‘x’ in Φx with a name n for o would result in a true sentence.* By the same token, ‘ $\forall x$ ’ designates the second-order concept, C_\forall , which takes a first-order concept C as argument and assigns it the True iff C assigns the True to every object. Thus, $[\forall x \Phi x]$ is true iff every object o is such that *the concept designated by Φx assigns o the value the True iff o “satisfies” the formula Φx iff replacing occurrences of ‘x’ in Φx with a name n for o would result in a true sentence.* This is Frege’s breakthrough insight—creating the foundation of the new logic of quantification—into how quantificational sentences are to be understood.

His treatment of the definite description operator, ‘the x’, is a variant on this theme. In specifying our symbolic language, we noted that ‘the x’ combines with a formula Φx to form a compound singular term (a Fregean “Name”) $[\text{the } x \Phi x]$ (called “a definite description”). On Frege’s analysis, the definite description operator designates the second-order function (not concept), F_{the} , which takes a first-order concept C as argument and assigns it an object o as value iff C assigns the True to o, and only to o. What should be said in the event that no object o satisfies this condition will be a topic for later. For now, we simply note that the description $[\text{the } x \Phi x]$ is a singular term (Fregean Name) that designates o if *the concept designated by Φx assigns o, and only o, the value the True—i.e., if o, and only o, “satisfies” the formula Φx , which, in turn, will hold if o, and only o, is such that replacing occurrences of ‘x’ in Φx with a name n for o would result in a true sentence.*

Taken together, the above principles constitute the core of a Fregean interpretation of the language L_F . Though conceptually quite simple, the general framework is powerful, flexible, and extendable to languages of much greater complexity, including ordinary spoken languages like

³The variable, in effect, marks the gap—as well as linking related gaps—in the predicate.

English and German. However, as we shall see in the next section, there are further central elements of Frege's framework that remain to be put on the table.

3. SENSE, REFERENCE, COMPOSITIONALITY, AND HIERARCHY

So far, we have ignored an entire dimension of meaning—namely, what Frege calls “sense.” He introduced his conception of sense in the first few paragraphs of “On Sense and Reference” with an argument involving identity sentences containing names or definite descriptions. Although the argument was a powerful and influential one, its focus on the identity relation—which holds only between an object and itself—introduced unnecessary complications, and spawned confusions, separable from the main point at issue. For that reason, I will present the idea behind Frege's argument in a different way, reserving a critical discussion of his own formulation of the argument until chapter 2. The argument here is based on a famous problem known as “Frege's puzzle,” which involves explaining why substituting one term for another in a sentence sometimes changes meaning, even though the two terms refer to the same thing. The argument takes English sentences (1–3) to be obvious examples of such cases. (The same points could have been made using examples drawn from L_F .)

- 1a. The brightest heavenly body visible in the early evening sky (at certain times and places) is the same size as the brightest heavenly body visible in the morning sky just before dawn (at certain times and places).
- b. The brightest heavenly body visible in the early evening sky (at certain times and places) is the same size as the brightest heavenly body visible in the early evening sky (at certain times and places).
- 2a. Hesperus is the same size as the brightest heavenly body visible in the morning sky just before dawn (at certain times and places).
- b. Hesperus is the same size as Hesperus.
- 3a. Hesperus is the same size as Phosphorus.
- b. Phosphorus is the same size as Phosphorus.

The contention that the (a)/(b) sentences in these examples differ in meaning is supported by three facts. First, one can understand both sentences, and so know what they mean, without taking them to mean the same thing, or even to agree in truth value. For example, understanding the sentences is consistent with taking the (b) sentences to be true and the (a) sentences to be false. Second, one who assertively utters (a) would typically be deemed to say, or convey, something different from, and more informative than, what one would say, or convey, by assertively uttering (b). Third, one would standardly use the (a) and (b) sentences in ascriptions $\lceil A \text{ believes that } S \rceil$, in which (a) and (b) take the place of S , to report what one took to be different beliefs. If these three points are sufficient for

the (a) and (b) sentences to differ in meaning, then principles T1 and T2 cannot be jointly maintained.

- T1. The meaning of a name or a definite description is the object to which it refers.
- T2. The meaning of a sentence S (or other compound expression E) is a function of its grammatical structure plus the meanings of its parts; hence, substituting an expression β for an expression α in S (or E) will result in a new sentence (or compound expression) the meaning of which does not differ from that of S (or E), provided that α and β do not differ in meaning.

Although Frege takes both ordinary names and definite descriptions to be singular terms the referents of which are objects, he rejects T1. For him, the meaning of a name is not its bearer, and the meaning of a definite description is not what it denotes. Instead, meaning—or in his terminology, *sense*—is what determines reference. It is the mode by which the referent of a term is presented to one who understands it. This sense, or *mode of presentation*, is a condition, grasped by one who understands the term, satisfaction of which by an object is necessary and sufficient for that object to be the referent of the term. For example, the sense of the description ‘the oldest living American veteran of World War II’ is a complex condition satisfaction of which requires one both to have been an American soldier in World War II, and to be older than any other such soldier. Although different terms with the same sense must have the same referent, terms designating the same referent may differ in sense, which explains the difference in meaning between the (a) and (b) in sentences (1) and (2). The explanation is expanded to (3) by Frege’s contention that, like definite descriptions, ordinary proper names have senses that determine, but are distinct from, their referents.

The case of proper names is complicated by his admission that it is common for different speakers to use the same name to refer to the same thing, even though they may associate different senses with it. Frege’s examples suggest that he regards the sense of a name n , as used by a speaker s at a time t , to be a reference-determining condition that could, in principle, be expressed by a description.⁴ On this view, n as used by s at t refers to o iff o is the unique object that satisfies the descriptive condition associated with n by s at t . When there is no such object, n is meaningful, but refers to nothing. Although Frege thinks that pains should be taken to avoid such reference failures in a “perfect” language constructed for logic, mathematics, and science, he seems to regard such failures in ordinary speech as tolerable nuisances with limited practical effects. (Comparable points hold for meaningful definite descriptions that fail to designate any object.) In the case of a proper name n , the meaning, for a speaker (at a

⁴ See Frege (1892b), p. 153, including footnote B; also, Frege (1918a), pp. 332–33, both in Frege (1997).

time), of a sentence containing n is the same as that of the corresponding sentence in which the reference-determining description the speaker implicitly associates with n (at the time) is substituted for n . Thus, for Frege, (3a) and (3b) differ in meaning for any speaker who associates the two names with different descriptive modes of presentation.

Although we have followed Frege in using examples involving names and descriptions of ordinary objects—like ‘Hesperus’ and ‘the brightest heavenly body visible in the early evening’—to motivate his distinction between the sense and referent of a term, the distinction is meant to apply to all singular terms. Thus, it should not be surprising that other, quite different, examples—such as those in (4)—could have been used to motivate the distinction.

- 4a. $6^4 > 1295$
- b. $1296 > 1295$

The same can be said about these examples as was said about (1–3)—namely that the contention that they differ in meaning is supported by three facts. First, one can understand them, and so know what they mean, without taking them to mean the same thing, or even agree in truth value. Hence, understanding them is consistent with taking (4b) to be true and (4a) to be false. Second, one who assertively utters (4a) would typically be deemed to say, or convey, something different from, and more informative than, what one would say, or convey, by assertively uttering (4b). Third, one would standardly use (4a) and (4b) in ascriptions ‘A believes that S’, in which they take the place of S, to report what one took to be different beliefs. If these observations justify a distinction between the Fregean senses and referents of the names and definite descriptions occurring in (1–3), then they also justify such a distinction for ‘ 6^4 ’ and ‘1296’.

This is true, *even though (4a), (4b), and the identity ‘ $6^4 > 1295$ ’ are a priori truths of arithmetic that qualify as “analytic” for Frege*. This means that the two expressions ‘ 6^4 ’ and ‘1296’ can have different Fregean senses despite the fact that it is possible for one who understands both to reason a priori from knowledge *that for all x ‘ 6^4 ’ refers to x iff $x = 6^4$ and for all y ‘1296’ refers to y iff $y = 1296$* to the conclusion *that ‘ 6^4 ’ and ‘1296’ refer to the same thing*. This is not a criticism of Frege’s notion of sense, which corresponds quite well in this respect to standard conceptions of linguistic meaning. However, it is also not without consequence for his overall philosophical view, which includes not only his philosophy of language but also his philosophy of logic and mathematics. Since his views about language evolved in the service of his goal of illuminating mathematics and logic, it is an important question, usefully emphasized in Beaney (1996), how well the notion of sense that emerges from his linguistic investigations in “On Sense and Reference” and related essays advances his central project of providing an analysis of number that reduces arithmetic to logic. This is something to keep an eye on as we proceed.

With this in mind, we return to Frege's puzzle, in which he uses the compositionality principle T2 for senses of sentences and other compound expressions, to reject a purely referential conception of meaning. T2 is paralleled by the compositionality of reference principle, T3, for terms, plus Frege's thesis T4 about sentences.

- T3. The referent of a compound term E is a function of its grammatical structure, plus the referents of its parts. Substitution of one coreferential term for another in E (e.g., substitution of '5³' for '125' in 'the successor of 125') results in a new compound term ('the successor of 5³') the referent of which is the same as that of E. Moreover, if one term in E fails to refer, then E does too (e.g., 'the successor of the largest prime').
- T4. The truth or falsity of a sentence is a function of its structure, plus the referents of its parts. Substitution of one coreferential term for another in a sentence S results in a new sentence with the same truth value as S. For example, the sentences in the following pairs are either both true, or both false.

The author of the *Begriffsschrift* was widely acclaimed during his time.
The author of "On Sense and Reference" was widely acclaimed during his time.

The probe penetrated the atmosphere of Hesperus.

The probe penetrated the atmosphere of Phosphorus.

$2^{10} > 6^4$

$1024 > 1296$

As before, I use examples drawn from English. However, perhaps surprisingly, Frege would insist that T3 and T4 hold (along with T2) not just for some terms and sentences of some languages, but for all languages. To be sure, T3 and T4 are already incorporated into our semantic interpretation of L_F . Since the language was constructed by us, we were free to set things up so as to make this so. Following Frege, we defined the truth value of an atomic sentence of L_F consisting of an n-place predicate sign Θ plus n terms to be the value assigned by the concept designated by Θ to the n-tuple of objects that are referents of the terms. Hence if $\lceil \Theta t_1 \dots t_n \rceil$ is true (false), then the result of substituting coreferential terms for any, or all, of $t_1 \dots t_n$ must also be true (false). We also followed Frege in defining the referent of a compound term of L_F consisting of an n-place function sign ϕ plus n terms to be the object assigned by the function designated by ϕ to the n-tuple of objects that are referents of the terms. Hence if o is the referent of $\lceil \phi(t_1 \dots t_n) \rceil$, then o is also the referent of any compound term that results from substituting coreferential terms for any, or all, of $t_1 \dots t_n$. Finally, we defined the operators that combine with sentences Φ and Ψ to form larger sentences as designating functions from truth values to truth values, thereby ensuring that if T4 holds for Φ and Ψ , it will hold for the compound sentences constructed from them using the operators. These points generalize to ensure that T3 and T4 are true of the sentences and terms of L_F .

However, it is one thing to construct a fruitful formal language for logic and mathematics that conforms to these principles, and quite another to show that natural languages like English and German do, or—even more strongly—that all possible languages (perhaps with a certain minimal expressive power) do. It certainly *seems* possible (i) that a language might contain a function sign designating a function that assigns a value to an n -tuple of arguments, one or more of which is the sense, rather than the referent, of the term supplying the argument, or (ii) that the language might contain a two-place predicate Θ occurring in a sentence $\lceil \alpha \Theta \text{'s } \beta \rceil$ designating a concept that maps the referent o of α plus the sense s of β onto the value the True (or the False), depending on the relationship between o and s , or (iii) that the language might have a sentential operator O designating a concept that maps the sense, rather than the truth value, of S onto the truth value of the complex sentence $\lceil O(S) \rceil$. Languages that allow these possibilities will violate T3, T4, or both. Thus, when L is a naturally spoken language—not devised with the purpose of conforming to these and other principles—it would seem to be an empirical question whether the principles are true of L . This question will be examined in chapter 2. For now, we take T3 and T4 for granted.

T5 is a corollary of T4, which is also worth noting.

T5. If one term in a sentence S fails to refer, then S lacks a truth value (is neither true nor false). Examples include:

The present king of France is (isn't) wise.

The largest prime number is (isn't) odd.

Truth value gaps of the sort illustrated here will arise in any language that both allows some singular terms that fail to refer and incorporates the Fregean semantic principles illustrated by L_F . In any such language, the truth value of a sentence consisting of a predicate Θ plus a term α is the truth value assigned to the referent of α by the concept designated by Θ —which is a function from objects to truth values. Since there is no argument for the function to apply to when α fails to refer, the sentence has no truth value. This is significant for Frege's account of negation, since when S lacks a truth value, there is no argument on which the truth function designated by the negation operator can operate—so the negation of S must also be truth valueless. The analysis generalizes to many-place predicates and truth-functional connectives. Reference failure anywhere in a sentence results in its truth valuelessness. Such sentences aren't epistemically neutral. Since the norms governing belief and assertion require truth, asserting or believing something that isn't true is incorrect no matter whether the thing asserted or believed is false or truth valueless. Thus, for Frege, there is something wrong about asserting or believing that either the present king of France is wise or he isn't, or that the largest prime number is odd or it isn't.

All of the sentences we have looked at so far share an important characteristic: in every case, the truth value of the sentence depends on the

referents of its parts. *Noticing this, Frege subsumed T4 and T5 under T3 by holding that sentences refer to truth values—the True and the False—which he took to be objects of a certain kind.* On this picture, the referent (truth value) of a sentence is compositionally determined by the referents of its parts, while its meaning (the thought it expresses) is *composed* of the meanings of its parts. Just as the sentence

5. The author of the *Begriffsschrift* was German.

consists of a subject phrase and a predicate, so (ignoring tense) the sense of the sentence—which Frege calls the *thought* it expresses—consists of the sense of the subject (which determines an object *o* as referent iff *o*, and only *o*, wrote the *Begriffsschrift*), and the sense of the predicate (which determines as referent the function that assigns the True to an individual iff that individual was German, and otherwise assigns the False).⁵ As for the structure of thoughts, he says:

If, then, we look upon thoughts as composed of simple parts, and take these, in turn, to correspond to the simple parts of sentences, we can understand how a few parts of sentences can go to make up a great multitude of sentences, to which, in turn, there correspond a great multitude of Thoughts.⁶

The idea, of course, is that the structure of thoughts mirrors the structure of the sentences that express them. Just as a sentence has a grammatical unity that comes from combining (complete/saturated) nominal expressions that stand for objects with an (incomplete/unsaturated) predicate expression that stands for a concept to form a grammatically unified structure that is more than a mere list, so, Frege thinks, the thought expressed by a sentence has a representational unity that comes from combining complete/saturated senses (which are modes of presentation of objects) with an incomplete/unsaturated sense (which is a mode of presentation of a concept) to form an intentional unity that represents things as being a certain way—and so is capable of being true or false, depending on whether the things in question are, in reality, as the thought represents them to be.⁷

⁵ Frege explicitly extends his sense and reference distinction to predicates in the first page of Frege (1892d), “Comments on Sense and Reference.” The same considerations apply to function signs generally. See Currie (1982), pp. 86–87, for discussion.

⁶ “Compound Thoughts,” published in 1923, translated and reprinted in Geach (1977), pp. 55–77, at page 55.

⁷ The parallel between the structure of sentences and that of the thoughts they express is here illustrated with simple sentences and thoughts. However, it also holds for complex sentences and thoughts of all types—including, for example, sentences in which an incomplete/unsaturated predicate expression (standing for a concept) combines with an incomplete/unsaturated quantifier expression (standing for a higher-level concept) to form a grammatical unity. The thoughts expressed by such sentences are unities in which the incompleteness of the sense that determines the higher-level concept complements the incompleteness of the sense that determines the lower-level concept in just the way needed for

The relationship between the Fregean sense of a predicate and its referent parallels the relationship between the sense of a name or description and its referent. The sense of an expression is always distinct from, and a mode of presentation of, the referent of the expression. This gives us a clue about the identity conditions for concepts, and functions generally. One possible way of conceiving of functions allows them to differ, even if, given the way the world actually is, they assign precisely the same values to precisely the same objects—provided that they do so on the basis of different criteria. On this conception, ϕ might assign the True to an individual in virtue of the individual's being a human being (while assigning the False to everything else in virtue of their not being human beings), and ψ might assign the True to an individual in virtue of the individual's being a rational animal (while assigning the False to everything else in virtue of their not being both rational and an animal). If all and only human beings happen to be rational animals, then, on this conception of functions, ϕ and ψ will assign the same values to the same actual arguments, even though they are different functions—as shown by the fact that they *would have assigned* different values to certain objects if, for example, a nonhuman species of rational animal had evolved. However, Frege didn't think of functions in this way—and, given his insistence that the sense of a predicate, or other functional expression, is always distinct from its referent, he had no need to individuate concepts, or other functions, so finely. Instead, he took concepts in particular, and functions in general, to be identical iff they in fact assign the same values to the same arguments.⁸

Being a Platonic realist about senses, Frege accepted the truism that there is such a thing as *the* meaning of 'is German', and that different speakers who understand the predicate know that it has that meaning. For him, senses, including the thoughts expressed by sentences, are public objects available to different thinkers. There is, for example, one thought—that *the square of the hypotenuse of a right triangle is equal to the sum of the squares of the remaining sides*—that is believed by all who believe the Pythagorean theorem. It is this that is preserved in translation, and this that is believed or asserted by agents who sincerely accept, or assertively utter, a sentence synonymous with the one used to state the theorem. For Frege, thoughts and their constituents are abstract objects, imperceptible to the senses, that are grasped by the intellect. These are the timeless contents in relation to which our use of language is to be understood.

a representational unity to be formed. The rule, for Frege, is that grammatical unities all require at least one incomplete/unsaturated expression, while representational unities all require at least one incomplete/unsaturated sense. Frege's views of this—which are responses to what has come to be known as "the problem of the unity of the proposition"—will be critically discussed in chapter 2. The evolving responses of G. E. Moore and Bertrand Russell to the same problem will be explained and discussed in chapters 3, 7, and 9.

⁸ Frege makes this clear in "Comments on Sense and Reference." See, in particular, Frege (1997), p. 173.

We have seen that the Fregean sense of a singular term is a mode of presentation of its referent, which is an object, while the Fregean sense of a predicate is a mode of presentation of its referent, which is a concept. Something similar can be said about the Fregean sense of a sentence. Since he took the referent of a sentence to be a truth value, he took its sense—the thought it expresses—to be a mode of presentation of the True or the False. Although the analogy is not perfect, the relationship between the sense and referent of a sentence is something like the relationship between the sense and reference of a definite description. Just as the sense of a definite description may be taken to be a condition the unique satisfaction of which by an object is sufficient for that object to be the referent of the term, so a thought expressed by a sentence may be taken to be a condition the satisfaction of which by the world as a whole is sufficient for the sentence to refer to the True. The strain in the analogy comes when one considers what happens when no object uniquely satisfies a description, as opposed to what happens when the world as a whole doesn't "satisfy" the thought expressed by a sentence. In the former case, the description is naturally said to lack a referent, while in the latter case Frege takes the referent of the sentence to be a different object—the False. However, even here the analogy is not entirely off the mark, as is evidenced by his lament that reference failure is a *defect* in natural language to be *remedied* in formal work by *supplying* an arbitrary stipulated referent in cases in which the conventional reference-determining condition fails to be satisfied. Perhaps reference to the False should be understood along similar lines.

This brings us to a more important complication. Frege recognized that, given the compositionality of reference principle T3, he had to qualify his view that sentences refer to truth values. While taking the principle to unproblematically apply to many sentences, he recognized that it doesn't apply to *occurrences of sentences* as content clauses in attitude ascriptions [A asserted/ believed/ . . . that S]. Suppose, for example, that (6a) is true, and so refers to the True.

6a. Jones believes that $2 + 3 = 5$.

Since ' $2 + 3 = 5$ ' is true, substituting another true sentence—'Frege was German'—for it ought, by T3, to give us another true statement, (6b), of what John believes.

6b. Jones believes that Frege was German.

But this is absurd. An agent can believe one truth (or falsehood) without believing every truth (or falsehood). Thus, if the truth values of attitude ascriptions are functions of their grammatical structure, plus the referents of their parts, then the complement clauses of such ascriptions must, if they refer at all, refer to something other than the truth values of the sentences occurring there.

Frege's solution to this problem is illustrated by (7), in which the putative object of belief is indicated by the italicized noun phrase.

7. Jones believes *the thought expressed at the top of page 91*.

Since the phrase is not a sentence, its sense is not a thought. Thus, what is said to be believed—which is itself a thought—must be the referent of the noun phrase that provides the argument of ‘believe’, rather than its sense. This result is generalized in T6.

T6. The thing said to be believed in an attitude ascription [‘A believes E’] (or similar indirect discourse report) is what the occurrence of E in the ascription (or report) refers to.

Possible values of ‘E’ include [‘the thought/proposition/claim that S’], [‘that S’], and S. In these cases what is said to be believed is the thought that S expresses. If T6 is correct, this thought is the *referent* of occurrences of S, [‘that S’], and [‘the thought/proposition/claim that S’] in attitude ascriptions (or other indirect discourse reports). So, in an effort to preserve his basic tenets—that meaning is *always* distinct from reference, and that the referent of a compound is *always* compositionally determined from the referents of its parts—Frege was led to T7.

T7. An occurrence of a sentence S embedded in an attitude ascription (indirect discourse report) refers not to its truth value, but to the thought S expresses when it isn’t embedded. In these cases, an occurrence of S refers to S’s ordinary sense. Unembedded occurrences of S refer to the ordinary referent of S—i.e., its truth value.

Here, Frege takes not expressions but their *occurrences* to be semantically fundamental. Unembedded occurrences express “ordinary senses,” which determine “ordinary referents.” Singly embedded occurrences, like those in the complement clauses in (6a) and (6b), express the “indirect senses” of expressions, which are modes of presentation that determine their ordinary senses as “indirect referents.”⁹ The process is repeated in (8).

8. Mary imagines that John believes that *the author of the Begriffsschrift was German*.

The occurrences in (8) of the words in

9. John believes *that the author of the Begriffsschrift was German*

refer to the senses that occurrences of those words carry when (9) is not embedded—i.e., to the ordinary senses of ‘John’ and ‘believes’, plus the indirect senses of the words in the italicized clause. In order to do this, occurrences of ‘John’ and ‘believe’ in (8) must *express* their indirect senses (which are, of course, distinct from the ordinary senses they determine as indirect referents), while occurrences in (8) of the words in the italicized clause must *express* doubly indirect senses, which determine, but

⁹ Because they aren’t embedded, occurrences of italicized words in (7) have their ordinary, not, indirect referents.

are distinct from, the singly indirect senses that are their doubly indirect referents. And so on, *ad infinitum*. Thus, Frege ends up attributing to each meaningful unit in the language an infinite hierarchy of distinct senses and referents.

But if this is so, how is the language learnable? Someone who understands ‘the author was German’ when it occurs in ordinary contexts doesn’t require further instruction when encountering it for the first time in an attitude ascription. How, given the hierarchy, can that be? If s is the ordinary sense of an expression E , there will be infinitely many senses that determine s , and so are potential candidates for being the indirect sense of E . How, short of further instruction, could a language learner figure out which was *the* indirect sense of E ? Different versions of this question have been raised by a number of philosophers from Bertrand Russell to Donald Davidson.¹⁰ These, in turn, have provoked an interesting neo-Fregean answer, to be taken up in chapter 2.

4. FREGE’S LOGIC

The logic invented in the *Begriffsschrift* is the modern *predicate calculus*, which is the result of combining the truth-functional logic of the *propositional calculus*—familiar, in one form or another, from the Stoics onward—with a powerful new account of quantification (‘all’ and ‘some’) supplanting the long-standing, but far more limited, syllogistic logic dating back to Aristotle. The key to Frege’s achievement was his decision to trade in the traditional subject/predicate distinction of syllogistic logic for a clarified and vastly expanded version of the function/argument distinction from mathematics, ingeniously extended to quantification in the manner illustrated by the semantics for L_F in the previous sections.¹¹

A system of logic, in Frege’s modern sense, consists of a formal language of the sort illustrated by L_F , plus a proof procedure, which, in his case, is put in the form of a small set of axioms drawn from the language, plus a small number of rules of inference. A proof in the system is a finite sequence of lines, each of which is an axiom or a formula obtainable from earlier lines by the inference rules. His fundamental idea is that whether or not something counts as a proof in such a system must, in principle, be decidable merely by inspecting the formula on each line, and determining

¹⁰One version of the question is raised by a central argument in Russell’s “On Denoting.” This will be discussed in chapter 8. Another version is what stands behind Davidson’s notion of “semantic innocence” presented in his “On Saying That.” This version of the question will be addressed in chapter 2.

¹¹Informative discussions of the relationship of Frege’s system to syllogistic logic, as well as the logical contributions of Leibniz, Boole, and De Morgan, can be found in Kneale and Kneale (1962) and Beaney (1996). The former also usefully compares Frege’s contribution to that of his contemporary, Charles Sanders Peirce.

(i) whether it is an axiom, and (ii) whether, if it isn't, it bears the required structural relation to earlier lines in order for it to be obtainable from those lines by the rules. For this reason, the axioms themselves must constitute an effectively decidable set—i.e., there must be a purely mechanical procedure capable of deciding, in every case, whether a formula is one of the axioms. Similarly, rules of inference—like *modus ponens*, which allows one to infer B on a line in a proof iff earlier lines include both A and $[A \rightarrow B]$ —must be stated in such a way that the question of whether one formula is obtainable from earlier ones is effectively decidable in the same sense. When these requirements are met, the question of whether or not something counts as a proof in the system can always be uncontroversially resolved—thereby forestalling the need to prove that something is a proof.¹²

Questions can, of course, be raised about the status of the axioms, and about whether the rules of inference successfully transmit that status to formulas provable from them. Formal systems of logic in the modern sense, which employ fundamental notions not available, or at least not explicit, in Frege's time, characterize some sentences (of the languages of those

¹² For Frege, the chief factor motivating these requirements was a perceived practical need in the foundations of mathematics for a logical system powerful enough to eliminate avoidable error and uncertainty by formalizing proof in mathematics. Speaking about this on p. 1 of *The Foundations of Arithmetic*, he says:

After deserting for a time the old Euclidean standards of rigor, mathematics is now returning to them, and even making efforts to go beyond them. In arithmetic . . . it has been the tradition to reason less strictly than in geometry. . . . The discovery of higher analysis only served to confirm this tendency; for considerable, almost insuperable, difficulties stood in the way of any rigorous treatment of these subjects. . . . Later developments, however, have shown more and more clearly that in mathematics a mere moral conviction, supported by a mass of applications, is not good enough. Proof is now demanded of many things that formerly passed as self-evident. Again and again the limits to the validity of a proposition have been in this way established for the first time. The concepts of function, of continuity, of limit and infinity have been shown to stand in need of sharper definition. Negative and irrational numbers, which had long since been admitted into science, have had to submit to a closer scrutiny of their credentials. In all directions these same ideals can be seen at work—rigor of proof, precise delimitation of extent of validity, and as a means to this, sharp definition of concepts.

As Frege points out in section III of the preface of the *Begriffsschrift* (Frege 1997, p. 48), the pursuit of these goals required a precise formal language, or “concept script,” in the service of a formal proof procedure of the sort discussed above.

So that nothing intuitive could intrude here unnoticed, everything had to depend on the chain of reasoning being free of gaps. In striving to fulfill this requirement in the strictest way, I found an obstacle in the inadequacy of the language. . . . Out of this need came the idea of the present “concept script.” It is intended to serve primarily to test in the most reliable way the validity of a chain of inference and reveal every presupposition that tends to slip in unnoticed, so that its origin can be investigated. The expression of anything that is without significance for logical inference has therefore been eschewed.

systems) as *logically true*, and some sentences as being *logical consequences* of others. To understand what this means, one must understand the difference between two classes of symbols. Certain expressions are singled out as “logical vocabulary,” the rest are called “nonlogical.” For example, the logical vocabulary of L_F consists of the quantifiers ‘ \forall ’ and ‘ \exists ’, the definite description operator ‘the’, and the truth functional connectives ‘&’, ‘ \vee ’, ‘ \sim ’, ‘ \rightarrow ’, and ‘ \leftrightarrow ’. The identity predicate ‘=’—which can be treated either as logical or nonlogical, depending on the aims of the system—will here be regarded as a logical symbol. A sentence in the language is said to be *logically true* iff it is true and would remain so (i) no matter what (nonempty) domain of objects its quantifiers were taken to range over, and (ii) no matter how its nonlogical vocabulary was interpreted to apply (or not apply) to those objects (i.e., no matter which objects in the domain its names refer to, no matter which objects its predicates, other than ‘=’, apply to, and no matter which functions from things in the domain to things in the domain its function signs designate).

Although this way of thinking about logical truths wasn’t itself made the subject of precise meta-mathematical investigation until Alfred Tarski formalized the notion of a model, or interpretation, of a formalized language in the 1930s, the foundations of Tarski’s idea were, I think, implicit in Frege.¹³ Using the idea of a model, we say that S is a logical truth iff S is true in all models/interpretations of the language, and that Q is a logical consequence of a sentence (or set of sentences) P iff every model/interpretation that makes P (or the sentences in P) true also makes Q true. Since only the interpretation of the logical vocabulary remains fixed across models, this fits the implicit Fregean ideas (i) that the logical truths, and our knowledge of them, are, in principle, independent of special truths that are unique to any particular domain, and (ii) that when Q is a logical consequence of P , what is required by one who knows P in order to come to know Q on that basis is not further specialized knowledge of the subject matter of P and Q .¹⁴

With this, we return to proof in a logical system. In addition to *consistency*, which requires that one’s proof procedure not enable one to derive

¹³ Tarski’s work will be discussed in volume 2.

¹⁴ According to Frege, the truths of logic (as well as those of arithmetic) are entirely general, and do not depend on any special subject matter. This suggests that any sentence that counts as a logical truth should remain such no matter how its names, function signs, and predicates are interpreted—and, one would think, no matter what domain of objects the quantifiers are chosen to range over (though this may be less than transparent). Since a model is just a formalization of the idea of such an interpretation, it should not be seen as foreign to Frege, even though he lacked any such explicit notion. Of course, this picture also depends on taking definitions, which play the central role in his reduction of arithmetic to logic, as themselves counting as “logical truths.” This is at variance with the Tarskian tradition. Perhaps it would be best to characterize Fregean logical truth as consisting of all sentences that can be turned into more standard logical truths by substitution on the basis of correct definitions. In other words, the logical and the analytic are merged.

contradictions (defined as sentences false in all models) from the axioms of the system, there are three natural demands one might place on a formal proof procedure. The maximal demand is that there should be a *decision procedure* which, given any sentence S of the language of the system, will always decide correctly (in a finite number of steps) whether S is, or is not, a logical truth. The only logical system we will be concerned with that satisfies this demand is the propositional calculus—which consists solely of atomic sentences, plus compound sentences constructed from them using only the truth-functional connectives. A sentence S of this system is a logical truth iff S is a tautology—i.e., a sentence that comes out true no matter what truth values are assigned to the (finitely many) atomic sentences from which it is constructed. That there is a decision procedure for tautology is evident from “the truth-table method”—which consists of writing down every possible assignment of truth values to the atomic sentences from which S is constructed, and then determining the truth value of each compound clause in S , relative to each such assignment—starting with the simplest and working one’s way up to more complex clauses, and finally to S itself—by consulting the function from truth values to truth values designated by the connective used to form each clause. If, at the end of this process, no assignment of truth values to the atomic sentences makes S false, then S is a logical truth; otherwise it isn’t.

Since the method always terminates, an answer is always reached. Hence it is a decision procedure. However, employing it, when the number of atomic sentences in S gets large, can be quite laborious. Hence authors of logic texts sometimes formulate a small number of axioms drawn from the language of the system and one or two rules of inference, from which all the tautologies can be derived. Frege’s axiomatic proof procedure for the propositional calculus in the *Begriffsschrift* is particularly simple and elegant—more so, he argued, than those of some of his illustrious predecessors, like George Boole.¹⁵ However, apart from such modest improvements, this well-trodden ground is not where he made his revolutionary advance.

Once we move beyond the propositional calculus, the maximal demand—for a decision procedure—is too strong for most interesting systems of logic, though the minimal demand—that every sentence provable from the axioms be true in all models, and hence be a genuine logical truth—is within the reach of any system of logic worthy of the name. Systems meeting this requirement (including Frege’s version of the predicate calculus) are called *sound*. The most interesting, and modestly ambitious,

¹⁵ Frege’s system of the propositional calculus takes just two connectives (‘ \sim ’ and ‘ \rightarrow ’) as primitive, defining the others from these. There are two rules of inference, *modus ponens* and substitution, plus six axioms. He argues—in Frege (1880–81) and (1882)—that his system is superior to, and more explanatory than, Boole’s formalization, due to its fewer axioms and primitives—a point discussed in Beaney (1996).

demand to be placed on such systems goes beyond this in requiring the specified proof procedure to provide a formal proof (from the axioms) of every logical truth. A logical system that satisfies this demand is called *complete*.

Here the facts are more complex, and require us to distinguish *the first-order predicate calculus* from *the second-order predicate calculus*. The language L_F that I have used to illustrate Frege's ideas is a first-order language—which means that the only quantifiers it employs are those that combine with *individual variables* (a type of singular term), and range over objects. However, this is just a fragment of Frege's total system, which qualifies as a version of the second-order calculus by virtue of allowing quantifiers to combine with predicate and function variables that range over concepts and functions. Still, the first-order fragment is a significant part of his system. Although the metatheorem that there are complete proof procedures for the first-order predicate calculus wasn't established until Kurt Gödel did so in his doctoral dissertation in 1929 (which is repeated in more succinct form in Gödel 1930), Frege's proof procedure for the first-order fragment of his logical system is, in fact, complete. What he didn't, and couldn't, know is that no higher-order version of the predicate calculus of the sort he offered can be both sound and complete—which is a corollary of Gödel (1931).¹⁶

To convert the first-order language L_F into second-order language L_{F+} we add new n -place predicate variables— $\mathcal{P}, \mathcal{Q}, \mathcal{R}$, etc.—for arbitrary n , and/or new n -place function variables— f, g, h , etc. (Frege has both.) Syntactically the predicate variables combine with (singular) terms to form atomic formulas, and the function variables combine with terms or other function symbols to form (singular) terms. These new variables require the addition of new rules for forming quantified formulas that parallel those for the first-order quantifiers. When $\Phi(\mathcal{P})$ is a formula containing the predicate variable \mathcal{P} , $\lceil \forall \mathcal{P} \Phi(\mathcal{P}) \rceil$ is a formula that is a universal generalization of $\Phi(\mathcal{P})$, and $\lceil \exists \mathcal{P} \Phi(\mathcal{P}) \rceil$ is a formula that is an existential generalization of $\Phi(\mathcal{P})$. Similarly, when $\Phi(f)$ is a formula containing the function variable f , $\lceil \forall f \Phi(f) \rceil$ is a formula that is a universal generalization

¹⁶ In addition to being complete, first-order logic is also *compact*—in the sense that every inconsistent set of first-order sentences (i.e., every set the members of which are not jointly true in any model) has a finite subset that is inconsistent. By contrast, second-order logic is neither complete nor compact. The reason that logical truth is not decidable in complete systems of first-order logic is that although each such truth can be proven in finitely many steps, there is no upper bound on the number of steps required in searching for a proof of an arbitrary sentence. We know that if S is a logical truth, we will ultimately find a proof, and if S isn't a logical truth, we will never find a proof that it is. What we don't, and can't, know in the general case is whether, after determining that there is no proof with n , or fewer, steps, there is proof with more than n steps waiting to be found. Hence, although the proof procedure is an effective positive test for logical truth, no end point can, in general, be established for failure, which means that we have no effective negative test for logical truth, and hence no decision procedure.

of $\Phi(\lambda)$, while $\lceil \exists \lambda \Phi(\lambda) \rceil$ is a formula that is an existential generalization of $\Phi(\lambda)$. (Free and bound occurrences of variables are determined exactly as before.)

In Frege's system, predicate variables are, of course, used to quantify over concepts, while function variables are used to quantify over (other) functions. Thus, the semantics for the higher-order quantifiers parallels the semantics of the first-order quantifiers.

Existential Quantification

1st-Order: '∃x' designates the 2nd-level concept C_2 , which takes a 1st-level concept C as argument and assigns it the value the True iff C assigns this value to at least one *object*. So $\lceil \exists x \Phi x \rceil$ is true iff there is at least one *object* o such that the concept designated by Φx assigns o the value the True iff o "satisfies" the formula Φx iff replacing occurrences of 'x' in Φx with a *name* n for o would result in a true sentence. Example: '∃x (Number x & Even x & Prime x)' is true iff at least one object falls under the concept *being a number that is both even and prime* iff there is something of which it is true that it is an even prime number.

2nd-Order: '∃ \mathcal{P} ' designates the 3rd-level concept C_3 , which takes a 2nd-level concept C_2 as argument and assigns it the value the True iff C_2 assigns this value to at least one 1st-level *concept*. So $\lceil \exists \mathcal{P} \Phi(\mathcal{P}) \rceil$ is true iff there is at least one 1st-level concept C such that the 2nd-level concept designated by $\Phi(\mathcal{P})$ assigns C the value the True iff C "satisfies" the formula $\Phi(\mathcal{P})$ iff replacing occurrences of ' \mathcal{P} ' in $\Phi(\mathcal{P})$ with a *predicate* designating C would result in a true sentence. Example: '∃ \mathcal{P} (\mathcal{P} Aristotle & \mathcal{P} Plato & $\sim \mathcal{P}$ Pericles)' is true iff at least one 1st-level concept falls under the 2nd-level concept *being something under which Plato and Aristotle but not Pericles fall* iff there is something that is true of both Plato and Aristotle but not Pericles.

2nd-Order: '∃ λ ' designates the higher-level concept C_3 , which takes a concept C_2 as argument and assigns it the value the True iff C_2 assigns this value to at least one 1st-level *function*. So $\lceil \exists \lambda \Phi(\lambda) \rceil$ is true iff there is at least one 1st-level function F such that the 2nd-level concept designated by $\Phi(\lambda)$ assigns F the value the True iff F "satisfies" the formula $\Phi(\lambda)$ iff replacing occurrences of ' λ ' in $\Phi(\lambda)$ with an expression designating F would result in a true sentence. Example: '∃ λ [$\forall x \forall y ((Ax \ \& \ Ay \ \& \ x \neq y) \rightarrow (B/\lambda(x) \ \& \ B/\lambda(y))$)]

$\&/(x) \neq/(y))]$] is true iff at least one function from objects to objects falls under the concept *being something that maps every A onto a distinct B* iff it is possible to put all of the A's into 1-1 correspondence with some of the Bs.

The rules for the corresponding types of universal quantification are exactly analogous.

However, Frege's system doesn't stop here. For all finite n , his system allows n^{th} -order quantification over concepts/functions of level $n - 1$. For example, call any first-level concepts A and B *equal* iff the objects falling under each can be put in one-to-one correspondence with the objects falling under the other—in the sense that all the A's can be put into one-to-one correspondence with some of the B's (as spelled out in the final example of second-order quantification above), and all the B's can be put in one-to-one correspondence with some of the A's (as spelled out by interchanging 'A' and 'B' in the example). In Frege's system, this two-place predicate—'() is equal to ()'—will syntactically combine with a pair of *first-order predicates* to form an atomic formula, e.g., 'A is equal to B'. The new predicate will designate a *second-level concept* that maps a pair of first-level concepts onto the True iff the objects falling under each can be exhaustively paired off 1 to 1 with the objects falling under the other. Next we ascend the quantificational hierarchy by replacing the equality predicate with a two-place, *second-level predicate variable* \mathcal{R}^2 that ranges over *second-level concepts*. This gives us the formula 'A \mathcal{R}^2 B' that designates the third-level concept C_3 that assigns the value the True to any two-place, second-level concept C_2 that assigns the True to the pair of first-level concepts A and B. The existential quantifier ' $\exists \mathcal{R}^2$ ' will then designate a fourth-level concept C_4 that takes a third-level concept as argument and assigns it the value the True iff that concept assigns this value to *at least one two-place, second-level concept*. This means that the quantified sentence ' $\exists \mathcal{R}^2 [A \mathcal{R}^2 B]$ ' that results from attaching the quantifier to the formula 'A \mathcal{R}^2 B' is true iff the fourth-level concept designated by the quantifier assigns the True to the third-level concept C_3 , designated by the formula (which assigns the value the True to any two-place, second-level concept C_2 , that assigns the True to the pair of first-level concepts A and B). Thus, the quantified sentence will be true iff *there is at least one second-level concept that is true of the first-level concepts A and B*. This is an example of *third-order quantification, which is quantification over second-level concepts*. (The "order level of the quantification" comes from the level of the concept that is the argument of the higher-level concept designated by the quantifier.)

Although it is worth knowing how this higher-order system works for each n , the system itself is complicated. Fortunately, our purposes don't require close examination of the higher reaches of the system. However, the differences between first- and second-order quantification can't be ignored. As mentioned above, while it is an important Gödelian metatheorem

that the first-order predicate calculus has sound and complete formalizations, it is an equally important Gödelian metatheorem that no sound formalization of the second-order calculus can be complete. If a given proof procedure for the second-order calculus is sound, then (i) although every sentence provable from its axioms will be a logical truth, there will be many logical truths (sentences true in all models) that are not provable from the axioms, and (ii) although every sentence Q derivable in the system from a set of sentences P will be a logical consequence of P (in the model-theoretic sense), there will be many logical consequences of P that are not derivable in the system.

This failure of *logical truth* and *logical consequence* to be fully formalizable would have surprised Frege, and—I believe—deeply troubled him. A central goal of his construction of a fully explicit and precise “concept script,” accompanied by a clearly understood semantics, plus a rigorous proof procedure, was to formalize proof in mathematics in a way that would allow one to eliminate all appeal to fallible intuition, uncheckable insight, or unacknowledged presupposition in demonstrating the results of mathematical discovery. His system of the first-order predicate calculus can be said to have achieved this goal for a large and important range of cases. However, we now know that the goal is not fully realizable in the sense of providing a system capable of proving all second-order logical truths.

The significance of this result for his goal of reducing arithmetic, and with it much higher mathematics, to logic is not entirely straightforward. As we will see, the axioms of arithmetic to be reduced are, with one exception, capable of being stated by first-order sentences, all of which can be derived by a version of Frege’s system of logic plus definitions. The one exception is the principle of mathematical induction, which, informally put, states that if (i) the number zero has a property, and (ii) whenever a natural number has a property, the successor of that number also has the property, then (iii) every natural number has the property. There are two well-known ways of formalizing this principle. The first formulation—used in first-order theories of arithmetic—is an *axiom schema*, adoption of which counts as adopting each of its infinitely many instances as an arithmetical axiom.

The First-Order Axiom Schema of Mathematical Induction

$$[F(0) \ \& \ \forall x (NNx \rightarrow (Fx \rightarrow \exists y (Sxy \ \& \ Fy)))] \\ \rightarrow \forall x (NNx \rightarrow Fx)$$

Here ‘ NN ’, ‘ S ’, and ‘ 0 ’ are, respectively, the primitive arithmetical predicate true of all natural numbers, the primitive arithmetical predicate true of all pairs the second of which is the successor of the first, and the arithmetical name of the number zero. Instances of the schema are

obtained by replacing ‘F’ with any first-level formula of the language of arithmetic in which ‘x’ (and only ‘x’) occurs free, and replacing ‘F(0)’ with the sentence of the language that results from replacing all free occurrences of ‘x’ in the formula that replaces ‘F’ with occurrences of ‘0’. Informally, each instance of this axiom says *that if zero “is so-and-so,” and if whenever a natural number “is so-and-so” its successor is too, then every natural number “is so-and-so”*—where the range of the “so-and-so’s” is the class of Fregean concepts designated by formulas of the language of arithmetic with one free variable.

The second formulation of the principle of mathematical induction—used in second-order theories of arithmetic—is a universally quantified second-order sentence of the language of second-order arithmetic.

The Second-Order Axiom of Mathematical Induction

$$\forall \mathcal{P} [[\mathcal{P}(0) \ \& \ \forall x (\text{NN}x \rightarrow (\mathcal{P}x \rightarrow \exists y (\text{S}xy \ \& \ \mathcal{P}y)))] \rightarrow \forall x (\text{NN}x \rightarrow \mathcal{P}x)]$$

Here, ‘ \mathcal{P} ’ is a predicate variable ranging over all first-level Fregean concepts—i.e., all functions assigning truth values to individuals in the domain of the first-order quantifiers. The axiom tells us that any such concept true of zero, and true of the successor of a natural number whenever it is true of the natural number itself, is true of all natural numbers.

The difference between these two is that the second-order sentence is stronger than the collection of the infinitely many instances of the first-order axiom schema. The reason for this is that there are more Fregean concepts assigning truth values to natural numbers than there are such concepts *that are designated by the formulas of first-order arithmetical theories*. The basis for this claim will be explained in chapter 7, which will include a discussion of Bertrand Russell’s reaction to Georg Cantor’s proof that *the power set* of any set s (which is the set of all subsets of s) is larger than s —in the sense that although the members of s can always be exhaustively paired off 1 to 1 with a proper subset of the members of its power set, the members of its power set cannot be so paired off with the members of s . For now, it is enough to note (i) that the set of formulas (in which only ‘x’ occurs free) of the language of first-order arithmetic can be exhaustively paired off with the set of natural numbers—which means that the set of Fregean concepts designated by those formulas is the same size as the set of natural numbers; and (ii) that since there is an exact correspondence

between an arbitrary set s and the Fregean concept that assigns the True to all and only members of s , the set of Fregean concepts that assign truth values to the natural numbers is the same size as the power set of the set of natural numbers. It follows from (i) and (ii) plus Cantor's result that the set of Fregean concepts that assign truth values to the natural numbers is larger than the set of such concepts that are designated by the formulas of first-order arithmetical theories. Hence the second-order axiom of mathematical induction is stronger (more inclusive) than the set of instances of the first-order axiom schema of mathematical induction.

The effect of this on formal theories of arithmetic is striking. Let T be a theory consisting of a formal language L_T plus a decidable set of axioms drawn from L_T . Let the theorems of T be all and only the logical consequences of the axioms of T . Call the theory L_T -complete iff for every sentence S of L_T either S or its negation is a theorem of T . (Note, this is a different notion of completeness than that which applies to systems of logic.) Since we are strongly inclined, at least pre-theoretically, to think that every sentence S in the language L_A of arithmetic is either true or false, and also that the negation of S is true iff S is false, we are strongly inclined to think that for every sentence S in L_A , either S or the negation of S is true. This means that in order for a theory to capture all arithmetical truths it must be L_A -complete. However, it is a fact, deriving ultimately from Gödel (1931), that *every consistent axiomatizable theory of first-order arithmetic (employing the axiom schema of mathematical induction) is incomplete in this sense*. Thus, no axiomatizable first-order theory is capable of capturing all and only the arithmetical truths. Interestingly, this result does not carry over to second-order arithmetic. So, if we substitute the second-order axiom of mathematical induction for the first-order schema, we get a theory that is L_A -complete. The cost, of course, is that although (i) this second-order theory can in fact be *derived* from (a modest reconstruction of) Frege's logic plus his definition of arithmetical concepts in logical terms (using the proof procedure provided by the system), and (ii) every arithmetical truth is in fact a *logical consequence* of the arithmetical theory derived from the logic, (iii) not every arithmetical truth can be *derived* from (a suitably modified version of) Frege's logical axioms by the proof procedure, because second-order logical consequence is not fully formalizable.

These striking metalogical results—proved years after Frege's death—are now settled facts. What is not conclusively settled is their philosophical significance—including their significance for his philosophical project of grounding the certainty, objectivity, and a priori knowability of mathematics in the fundamental laws of logic. In order to approach this issue, we must delve further into his philosophical views of logic and mathematics, and how they influenced his attempted reduction of the latter to the former.

Although there is no denying that Frege was one of the chief architects of our contemporary, essentially mathematical, understanding of symbolic logic, he was also an ambitious philosophical epistemologist whose views

about logic were strongly tied to its role in justifying knowledge.¹⁷ As is well known, he was fond of characterizing the goal of logic as the discovery of the laws of truth, in something like the way in which the aim of physics is the discovery of the laws of heat, or light.¹⁸ But what are these laws? Not, as he correctly and repeatedly insisted, the psychological laws by which we think and reason. Rather, he contended, they are laws by which reasoning is justified.¹⁹ A good statement of this view is found in a draft, written sometime between 1879 and 1891, of a logic text that never appeared.

Logic is concerned only with those grounds of judgment which are truths. To make a judgment because we are cognizant of other truths as providing a justification for it is known as inferring. There are laws governing this kind of justification, and to set up these laws of valid inference is the goal of logic.²⁰

Although the first statement in this passage is a bit misleading, the general import of Frege's position is clear. A fundamental goal of logic—perhaps its fundamental goal—is to lay down laws of truth preservation. We now know that truth preservation and the preservation of justification do not always coincide. However, they often do, which seems to be what Frege had in mind. To be sure, the logical laws of truth preservation will also relate false premises to conclusions derived from them, but since no genuine knowledge is thereby achieved, that is not where the interest and value of logic is to be found.²¹ Rather, it is its role in extending and justifying our knowledge that provides logic with its *raison d'être*.

This is the perspective from which Frege's program of reducing arithmetic and (much of) the rest of mathematics must be viewed. Commenting in 1896 on the ambitious program initiated in the *Begriffsschrift*, he says:

I became aware of the need for a *Begriffsschrift* when I was looking for the fundamental principles or axioms upon which the whole of mathematics rests. Only after this question is answered can it be hoped to trace successfully the springs of knowledge upon which this science thrives.²²

The thought here is that finding the principles on which mathematics rests—by constructing logically valid proofs, in the manner of the *Begriffsschrift*, of the propositions of mathematics from those principles—will show how to arrive at justified knowledge of mathematics from antecedent

¹⁷ Frege's deep concern with the epistemological role of logic is rightly emphasized by Gregory Currie. See, in particular, chapters 1 and 4 of Currie (1982).

¹⁸ See, for example, Frege (1918a), p. 325, and Frege (1897), p. 128.

¹⁹ Frege (1918a), p. 326 of Frege (1997).

²⁰ Frege (1879–91), p. 3.

²¹ Frege's recognition that in *formulating* the laws of logic connecting premises and conclusions we have no interest in the truth of the premises is made explicit in Frege (1906), p. 175, where he says, "The task of logic is to set up laws according to which a judgment is justified by others, irrespective of whether they themselves are true." However, the value of this task is to be found in its employment on known truths to extend our knowledge.

²² Translated in Frege (1969), p. 1.

knowledge of the underlying principles. Of course, the strength of the justification thereby transferred will depend on the strength of the principles from which the mathematical truths are derived. Where might the strongest justification be found? About this, Frege is unequivocal. The strongest justification is provided by deriving a proposition from the fundamental laws of logic. Thus, he says in the *Begriffsschrift*,

The firmest proof is obviously the purely logical, which, prescind- ing from the particularity of things, is based solely on *the laws on which all knowledge rests*. Accordingly, we divide all truths that require justification into two kinds: those whose proof can be given purely logically and those whose proof must be grounded on empirical facts.²³

These fundamental laws “on which all knowledge rests” are not just any statements that turn out to be true in all models; nor are they just any rules for “inferring” one statement from others that turn out to be truth preserving no matter what the model. Rather, they are the foundational laws of a logical system—its axioms and rules of inference—from which other, non-obvious logical truths can be formally derived, and other, derived or secondary, rules of inference can be constructed. How are these foundational laws and principles known to be true (or truth preserving)? Frege briefly addresses this question in the introduction to volume 1 of *The Basic Laws of Arithmetic*.

Now the question of why and with what right we acknowledge a logical law to be true, logic can only answer by reducing it to another logical law. Where that is not possible, logic can give no answer. Leaving aside logic, we may say: we are forced to make judgments by our nature and external circumstances; and if we make judgments, we cannot reject this law—of identity, for example; we must recognize it if we are not to throw our thought into confusion and in the end renounce judgment altogether. I do not wish to either dispute or endorse this view and only remark that what we have here is not a logical implication. What is given is not a ground [reason] for [something’s] being true, but of our holding [it] as true.²⁴

The picture of justification suggested here is foundational. Some logical principles are justified by deriving them from other, more fundamental ones. The process of justification ends with the most basic logical laws, which are self-evidently true, and knowable without any further justifying reason. In addition to being self-evident, Frege takes these fundamental laws to be the most pragmatically significant general truths underlying all of our reasoning. It is because he understands them to have this status, while also taking arithmetic to be derivable from them, that he claims, in *The Foundations of Arithmetic*, that the same is true of “the fundamental propositions of the science of number,” holding that “we have only to

²³ Sec. 3 of the Preface, p. 48, my emphasis.

²⁴ Preface, sec. 17, p. 204.

deny any one of them and complete confusion ensues. *Even to think at all seems no longer possible.*²⁵

This is the epistemological bedrock on which stands Frege's project of establishing the a priori certainty of mathematics by reducing it to logic. In the end, he simply presupposes that logic itself is both certain and knowable a priori. Having no doubt that this is so, he recognizes that there are limits to how far we can go in explaining why it is. Regarding justification, he thinks that all there is to say about the most fundamental laws of logic is that they are self-evident; hence they neither need, nor are susceptible to, justification by anything more certain than they are. For those who find this position less than fully satisfying, I recommend holding off judgment until we have had a chance, in volume 2, to see how hard it is to advance beyond it—by examining the difficulties encountered by Rudolf Carnap's view that the truths of logic, and all other a priori truths, can be known to be true simply by understanding them, because their truth is guaranteed by their meaning alone.

Whether or not Frege's epistemological position is ultimately correct, some suitably qualified version of it may have some merit—especially if one emphasizes the generality and indispensability of at least some logical laws, gestured at in his remarks in *The Basic Laws of Arithmetic*.²⁶ It is a characteristic of logical laws that their domain is universal, and so not dependent on any special subject matter. To the extent that they are indeed, indispensable, they apply to, and are needed in, all domains of thought. Consequently, to reduce a mathematical theory to such laws is to ground it in principles needed for reasoning in every domain, and hence to render it immune from special skeptical doubts arising from any specific domain. Although a showing of the indispensability of certain logical "laws" for the thought of beings like us would not guarantee their correctness—or even add to our justification for taking them to be such—it might at least render such laws, plus that which can be formally derived from them, resistant to the kind of *reasoned refutation* that itself must presuppose the very logical principles it seeks to undermine. In this way, a showing of indispensability might provide as secure a bulwark against the sincere arguments of actual skeptics about logic and mathematics as one might reasonably hope for.

Nevertheless, it is important not to let one's epistemological goals become too expansive. The incompleteness of first-order theories of arithmetic, plus the incompleteness (in a related sense) of systems of second-order logic—and hence the unformalizability of second-order

²⁵ *The Foundations of Arithmetic* (Frege 1950), sec. 14, p. 21.

²⁶ The modern proliferation of alternatives to classical logic makes it unreasonable to suppose that all of its principles are indispensable to our thought. Nor would it be easy to precisely identify some restricted core that was indispensable in the required sense. However, it is also not obvious (to me) that there is nothing to this line of thought, or that no significant sort of justification could be made to emerge from it.

logical consequence—means that, if the language of arithmetic is indeed bivalent, then any hope of providing justifying knowledge of all arithmetical truths by deriving them from the fundamental laws of logic must be given up in favor of something weaker. How much of the heavily epistemological motivation of Frege's grand project can be salvaged is, at this point, an open question. One factor to bear in mind as we proceed to his proposed reduction of arithmetic to logic is the extremely high epistemic bar that Frege sets for the self-evidence and indispensability of any logical axiom needed for the reduction. This, as we will see, is an important source of potential difficulties.

Finally, epistemology aside, there are other perspectives from which one can view Frege's attempted reduction of arithmetic, and (much) of higher mathematics, from logic. For example, one may think one understands what makes propositions about ordinary middle-sized objects, like houses, true because one knows what houses are, and what it is for them to have one or another property, while at the same time being puzzled about what makes arithmetical, or other mathematical, propositions true, because one has no idea what numbers—e.g., 7 or 0—are, and what it is for them to have properties. It is not inconceivable that a reduction of higher mathematics to arithmetic, and arithmetic to logic, might effectively dispel this sort of metaphysical puzzlement, even if not all of Frege's epistemological goals for the reduction can be fulfilled.

5. FREGE'S PHILOSOPHY OF MATHEMATICS

5.1. Critique of Naturalism, Formalism, and Psychologism

Though the significance of Frege's philosophy of mathematics extends beyond the epistemological, both his positive views and his criticisms of those of others are, as Gregory Currie contends, driven in large part by his epistemology.²⁷ For Frege, our knowledge of mathematics is certain—as well as not being dependent on, or refutable by, experience in the way in which empirical propositions always are. For this reason, the laws of arithmetic cannot be high-level empirical generalizations supported by induction from past experience. The target of repeated attacks in Frege's 1884 *The Foundations of Arithmetic*, the view that they are such generalizations was prominently defended by John Stuart Mill in *A System of Logic*. The flavor of Mill's views about arithmetic is illustrated by the following passage from that work.

What renders arithmetic the type of deductive science is the fortunate applicability to it of a law so comprehensive as "The sums of equals are equals":

²⁷ Currie (1982).

or (to express the same principle in less familiar but more characteristic language), “Whatever is made up of parts, is made up of the parts of those parts.” This truth, *obvious to the senses in all cases which can be fairly referred to their decision*, and so general as to be co-extensive with nature itself, being true of all sorts of phenomena (for all admit to being numbered), must be considered an *inductive truth*, or law of nature, of the highest order. And every arithmetical operation is an application of this law, or of other laws deduced from it. *This is our warrant for all calculations. We believe that five and two are equal to seven, on the evidence of this inductive law, combined with the definitions of those numbers.*²⁸

The most fundamental of Frege’s many criticisms of this view is that it mistakes the *application* of arithmetic to experience for *inductive dependence* of arithmetic on experience. Consider, for example, the claim that a collection of seven things can (always) be conceptually divided into a collection of five and a collection of two (which we may take to be an obvious corollary of the arithmetical claim that $7 = 5 + 2$). Properly understood, this is true, and knowable a priori. However, to understand it in this way—which Frege insists is how we really do understand it—one must not confuse it, as Mill seemingly does, with the claim that whenever one can discern seven parts that exhaustively make up some physical whole, it is always possible to physically break up that whole, without loss, into something exhaustively made up of five of those parts, plus something else exhaustively made up of two of them. Frege criticizes this view as suffering from the problems (i–iii):²⁹

- (i) The things to which arithmetical claims apply aren’t limited to physical things; events, ideas, concepts, thoughts, lines, points, and sets can all be numbered, even if they don’t exist in space or time, and/or there is no physical operation of breaking them up.
- (ii) Even when physical objects are involved, the truth of an arithmetical claim doesn’t depend on assumptions about what we can perceive or imagine, or the parts we can *discern* in them.
- (iii) If *o* is a physical object that consists of seven non-overlapping parts, whether or not it is possible to physically break up *o*, without loss, into one thing exhaustively made up of five of the parts and another non-overlapping thing made up of the other two, is an empirical claim which we cannot know with the certainty that attaches to an arithmetical claim, and which may turn out to be false for non-arithmetical reasons.

For Frege, these problems point in the same direction. Rather than being something on which the truth of the purely arithmetical proposition that

²⁸ Mill (1843), book III, chapter 24, sec. 5, p. 401, my emphasis.

²⁹ These criticisms of Mill, along with others, can be found in sections 7–10 of *The Foundations of Arithmetic*, Frege (1950).

$7 = 5 + 2$ epistemically depends, the truth of the empirical proposition *that the seven coins on the table are made up of five coins belonging to Mary and two belonging to John* epistemically depends, in part, on the truth of the arithmetical proposition. Indeed, the content of the empirical claim is the conjunction of the empirical propositions *that the number of coins on the table that are Mary's = 5* and *that the number of coins on the table that are John's = 2* with the a priori truth *that $7 = 5 + 2$* .

This point is buttressed by an interesting argument about induction thought of not as a psychological process by which general beliefs arise from particular beliefs—but rather as a theory of when, and to what degree, experience provides evidence for an empirical claim by making the truth of the latter more probable than it otherwise would have been. Frege says:

The procedure of induction, we may surmise, can itself be justified only by means of general propositions of arithmetic—unless we understand by induction a mere process of habituation, in which case it has of course absolutely no power whatever of leading to the discovery of truth. The procedure of the sciences, with its objective standards, will at times find a high probability established by a single confirmatory instance, while at others it will dismiss a thousand as almost worthless; whereas our habits are determined by the number and strength of the impressions we receive and by subjective circumstances, which have no sort of right at all to influence our judgment. *Induction [then, properly understood,] must base itself on the theory of probability, since it can never render a proposition more than probable. But how probability theory could possibly be developed without presupposing arithmetical laws is beyond comprehension.*³⁰

The point, I take it, is that revising our assessment of the probable truth of a putatively empirical proposition in light of empirical evidence presupposes calculations, which are themselves arithmetical, or dependent on underlying claims that are. If this is right, then any claim that certain propositions are rendered probable by our evidence will presuppose arithmetic, which must already have been justified (if the probability claim is). Thus, Mill's claim that *arithmetic* is inductively justified (made probable) by experience itself requires that arithmetic be independently justified. Since this means that the ultimate grounding for arithmetic can't be empirical, it undermines his position.

In addition to this fundamental flaw in Mill's, and—Frege would say—any, empiricist view of arithmetic, he locates other shortcomings in Mill that any acceptable view must overcome. Although the passage from Mill adverts to specific arithmetical results, and talks about some laws being *deducible* from others, he doesn't offer any significant deductions, he does little to identify what numbers are, and he fails to show how their arithmetical properties could possibly be deduced from, or explicated by,

³⁰ Ibid., pp. 16–17, my emphasis; bracketed insert added by the translator, J. L. Austin.

empirical facts. One foray he does make into the identification of individual numbers with particular objects concerns the number three, which he says is defined by the equation ‘ $2 + 1 = 3$ ’. Though he takes this statement to be a definition, he also takes it—somehow—to state an empirical fact. The fact in question is that there exist groups of physical objects we perceive to be spatially arranged as follows:



They can be separated into two parts that look like this.³¹



Frege has a bit of fun with this, noting (i) that lots of things that aren’t, or can’t be, perceived can be numbered, (ii) that some groups of things are “nailed down,” in the sense that they don’t allow physical rearrangement, (iii) that how things appear to us, if they appear at all, has nothing to do with number, (iv) that very large numbers exist without our ever having encountered groups with the requisite number of discriminable parts, and (v) that it is hard to fit the number zero into Mill’s simplistic scheme.

To this we may add a further point, based on Frege’s insistence in other contexts, that ‘3’ and ‘the number three’ are singular terms that denote a unique object—the *one and only* number three. Mill slurs over this, adding to the first illustration above the idea that “we term all such parcels Threes.” Our question, however, is, “What is *the* number three?” Surely no one group of three things is a better candidate than any other for being this number, which, by definition, is *the* number that follows two. Is Mill telling us that there are many number threes? If not, what is he telling us? The same questions arise here.

The fact asserted in the definition of a number is a physical fact. *Each of the numbers* [he should have said “numerals”] two, three, four, &c., denotes physical *phenomena* [note the plural], and connotes a physical property of those phenomena. Two, for instance denotes *all pairs of things, and twelve all dozens of things*, connoting what makes them pairs or dozens; and that which makes them so is something physical.³²

In and of itself, this point may seem to be a small one. Instead of intimating something seemingly incoherent—that each collection of three things is *the* number three, and hence that the numeral, ‘3’, standing for that number, denotes each three-membered collection—couldn’t Mill have

³¹ Mill (1843), Book II, chapter 6, sec. 2, p. 169 (all references to the Longmans 1961 edition).

³² *Ibid.*, Book III, chapter 24, sec. 5, pp. 399–400.

said that the number three is the property of being a three-membered collection, or that it is the collection of all three-membered collections, either of which is something coherent and closer to what Frege has in mind? Perhaps. But the real issues here are larger. The features of Mill's position that he finds most attractive, and that are most distinctive to him—namely, the supposed inductive dependence of arithmetical truths on empirical evidence, and the avoidance of abstract objects (not existing in space or time) in favor of physical objects and properties—cause serious problems, while adding nothing to the explanation of what numbers really are, and how arithmetical truths are justified.

For Frege, the way to discover what numbers are, and how statements about them are justified, is first, to determine what we pre-theoretically know about them, and second, to frame definitions of each number, and of the class of natural numbers as a whole, in a way that allows the definitions to be combined with other background knowledge to deduce what we pre-theoretically know. How, for example, should 2, 3, 5, and addition be defined so that facts like (11) can be deduced from the definitions, plus our knowledge of logic and facts like (10)?

10. $\exists x \exists y (x \text{ is a black book on my desk} \ \& \ y \text{ is a black book on my desk} \ \& \ x \neq y \ \& \ \forall z (z \text{ is a black book on my desk} \rightarrow z = x \vee z = y)) \ \& \ \exists x \exists y \exists z (x \text{ is a blue book on my desk} \ \& \ y \text{ is a blue book on my desk} \ \& \ z \text{ is a blue book on my desk} \ \& \ x \neq y \ \& \ x \neq z \ \& \ y \neq z \ \& \ \forall z^* (z^* \text{ is a black book on my desk} \rightarrow z^* = x \vee z^* = y \vee z^* = z)) \ \& \ \forall x \forall y ((x \text{ is a black book} \ \& \ y \text{ is a blue book}) \rightarrow x \neq y)$

11a. The number of black books on my desk = 2 and the number of blue books on my desk = 3, so there are 5 books on my desk.

b. There are 2 black books on my desk, and 3 blue books on my desk, so there are 5 books on my desk.

More generally, how might a proper understanding of what natural numbers and arithmetical operations are be used to derive our purely arithmetical knowledge, plus its empirical applications, from the laws of logic, supplemented when necessary with relevant empirical facts? This, for Frege, is the most important question that a philosophical theory of number must answer.

His most fundamental objections to Mill—and to others he criticizes—are (i) that they generally don't even attempt to answer this fundamental question, and (ii) that what they do say only gets in the way of a proper answer. Mill's repeated emphasis on the *physical* properties of *physical* things being numbered is a case in point. This was attractive to him, and other naturalistically minded philosophers, as a way of avoiding commitment to non-spatiotemporal "logical" or "mathematical" objects, which can easily seem problematic. However, the wish to avoid potentially troubling commitments is no excuse for refusing to answer the fundamental Fregean question. Worse, even when the things being numbered are physical, as in

the example about the books on my desk, the fact that they are physical, and have physical properties, plays no role in getting from (10) to (11). To that extent, Mill offers no real theory of number at all. At best, his remarks about the relevance of our sense experience of physical things to statements about number are elaborate warm-ups for a pitch that never comes—except for cases in which the things numbered are not physical, or not experienced, and taking his remarks seriously *prevents* any pitch from being made.

This is the context in which Frege's remarks about the importance of the ordering of the natural numbers to our understanding of them, and the inability of induction from experience to shed light on what follows from this ordering, should be understood.

It is in their nature to be arranged in a fixed, definite order of precedence; and each one is formed in its own special way and has its own unique peculiarities, which are specially prominent in the cases of 0, 1, and 2. Elsewhere when we establish by induction a proposition about a species, we are ordinarily in possession already, merely from the definition of the concept of the species, of a whole series of its common properties. *But with the numbers we have difficulty in finding even a single common property which has not actually to be first proved common. . . .* The numbers are literally created, and determined *in their whole natures*, by the process of continually increasing by one. Now, this can only mean that from the way in which a number, say 8, is generated through increasing by one *all its properties can be deduced. But this is in principle to grant that the properties of numbers follow from their definitions, and to open up the possibility that we might prove the general laws of numbers from the method of generation which is common to them all, while deducing the special properties of the individual numbers from the special way in which, through the process of continually increasing by one, each one is formed.*³³

Here we have the ultimate critique of the view that arithmetic inductively depends on experience for justification. It doesn't because if we understand what the natural numbers are, and discover their proper definitions, we will see that the purely arithmetical propositions about them are logically derivable from those definitions, leaving nothing for induction to justify. Of course, at this stage of the *Foundations of Arithmetic*, Frege was not in a position to demonstrate this. However the burden of most of the rest of that work is to lay the foundations for just such a proof, to be executed in *The Basic Laws of Arithmetic*.

Mill's inductivism is, of course, not the only form of naturalism to which Frege objected. Another variant of the view, psychologism, holds that the laws of logic are natural laws that describe human reasoning, while natural numbers are ideas in the minds of agents. Although the general lessons extracted from the critique of Mill apply with equal force against

³³ Frege (1950), pp. 15–16, my emphasis.

psychologism, Frege gives further objections to psychologism. The locus of his attack is the discussion in sections 26 and 27 of *The Foundations of Arithmetic*, and sections xiv–xxv of the preface to *The Basic Laws of Arithmetic*, where he makes the following points:

- (i) There may be natural laws describing the thinking processes of all human beings—past, present, and future—but we don’t, at present, know what they are. Whatever their content may turn out to be, though, they will be high-level generalizations subject to falsification by future experience, the justification of which must come from experience. By contrast, the laws of logic (and arithmetic) are readily identifiable, known to be true, and incapable of falsification by experience. Thus, it would, as Frege says, “be strange if the most exact of all the sciences had to seek support from psychology, which is still feeling its way none too surely.”³⁴
- (ii) The reason for the contrast between the (still unknown) natural laws of human psychology and the (well known) laws of logic is that they are laws in different senses. Whereas the former provide descriptive, perhaps causal, accounts of human reasoning, the latter are laws of truth preservation that underlie the way in which knowledge claims may be justified. The application of these laws to human reasoning is derivative and indirect. To the extent that one aims at truth, validly deriving a logical consequence from a set of premises is correct—in the sense of being guaranteed to produce a truth (if one’s premises are jointly true), and of transferring one’s justification to one’s conclusion (if the conjunction of ones premises is justified).³⁵
- (iii) If numbers were mental constructs, they would be ideas in the minds of agents. But, Frege thinks, one’s ideas are private to oneself. From this, he draws three conclusions, each intended as a *reductio ad absurdum* of the claim that numbers are ideas in the mind. First, if numbers were ideas, then we could no longer speak of *the* number two, for example. Instead, “we should then have it might be many millions of twos on our hands. We should have to speak of my two and your two.” Second, although there might be millions of twos, large numbers might not exist, since it would “be doubtful whether there existed the infinite number of numbers that we ordinarily suppose. 10^{10} , perhaps, might be only an empty symbol, and there might exist no idea at all, in any being whatever, to answer to that name.” Finally, since (Frege thinks) the ideas

³⁴ Ibid., sec. 17, p. 38.

³⁵ This statement of what some call “the normativity of logic” is a bit weaker, and more explicitly qualified, than Frege’s own statement of the view in sec. 15, p. 202, of *The Basic Laws of Arithmetic*: “That the logical laws should be guiding principles for thought in the attainment of truth is generally admitted at the outset; but it is only too easily forgotten. The ambiguity of the word ‘law’ is fatal here. In one sense it states what is, in the other it prescribes what should be. Only in the latter sense can the logical laws be called laws of thought, in laying down how one should think.”

in one mind are inaccessible to others, for all we know, the numbers and arithmetical properties in one mind might be different from those in other minds, in which case arithmetic would be entirely subjective, rather than objective.³⁶

The upshot of these remarks is that, for Frege, psychologism was less a coherent theory of the nature of logic and mathematics than a source of confusion that obscured, and didn't take seriously, the fundamental questions with which any genuine philosophical theory of these sciences must be concerned.

The same is true of his critique of the crude versions of formalism of his time that identified numbers with physical marks on paper—such as inscriptions of numerals.³⁷ Pointing out that this is a confusion—in fact, the same confusion as the all-too-common identification of functions from numbers to numbers with expressions (often formulas containing variables)—was an important prerequisite to clarifying the mathematical use of the notion of a function, and extending it to logic, the philosophy of mathematics, and philosophical semantics, where, thanks to Frege, it now plays central roles. This is the purpose of his 1891 article “On Function and Concept,” the first few pages of which are devoted to dispelling crude formalist confusions of symbols with things signified. He says, in criticism,

Thus, e.g., the expression ‘ $2x^3 + x$ ’ would be a function of x , and ‘ $2 \times 2^3 + 2$ ’ would be a function of 2. This answer cannot satisfy us, for here no distinction is made between form and content, sign and thing signified. . . . Now what is the content of ‘ $2 \times 2^3 + 2$ ’? The same thing as ‘18’ or ‘ 3×6 ’. What is expressed in the equation ‘ $2 \times 2^3 + 2 = 18$ ’ is that the right-hand complex of signs has the same reference as the left-hand one. I must here combat the view that, e.g., $2 + 5$ and $3 + 4$ are equal but not the same. . . . Difference of sign cannot by itself be a sufficient ground for difference of the thing signified. The only reason why in our case the matter is less obvious is that the reference of the numeral 7 is not anything perceptible to the senses. . . . [T]his leads . . . to numerals being taken to be numbers. . . . [S]uch a conception . . . is untenable, for we cannot speak of any arithmetical properties of numbers whatsoever without going back to what the signs stand for. For example, *the property belonging to 1, of being the result of multiplying itself by itself, would be a mere myth; for no microscopical or chemical investigation . . . could ever detect this property in the possession of the innocent character we call . . . [numeral] one. . . . The characters we call numerals have . . . physical and chemical properties depending on*

³⁶ All these points, and the quoted passages, come from *The Foundations of Arithmetic* (Frege 1950), sec. 27, pp. 37–38.

³⁷ These views, attacked by Frege, are not to be confused with the much more sophisticated views of David Hilbert, for example, which were presented and attained a following toward the end of Frege's life, long after the publication of *The Foundations of Arithmetic* and the two volumes of *The Basic Laws of Arithmetic*.

the writing material. One could imagine the introduction some day of quite new numerals. . . . Nobody is seriously going to suppose that in this way we should get quite new numbers.³⁸

As before, we see the primacy of Frege's central insight. The account of what numbers are should be such that from it (plus logic and background facts needed for empirical applications) we can deduce the properties they have that constitute our pre-philosophical knowledge of them. No matter what other difficulties or absurdities may plague physical, psychological, or formalistic conceptions of number—of which there are many—it is their failure to satisfy this central criterion that disqualifies them as viable philosophical theories.

5.2. Critique of Kant

Unlike the views criticized above, Kant's views of mathematics were taken quite seriously by Frege. He agreed with Kant that there is an important distinction between the analytic and the synthetic that cuts across the distinction between a priori truths (the justifications of which aren't dependent on experience) and a posteriori truths (which can be justified only by experience). According to Kant the "judgment" expressed by $[A \text{ is } B]$ is analytic iff the concept/property expressed by the predicate B is contained in the concept/property expressed by subject A, whereas a synthetic judgment is one in which the concept/property expressed by the predicate is not so contained, in which case the judgment adds something to the subject concept/property.³⁹ We may take Kant's "judgments" to correspond to Frege's thoughts, and his "concepts/properties" to roughly correspond to Frege's senses. A key point for Kant, and for Frege, is that judgments that count as analytic *in Kant's sense* are epistemologically trivial. Since they don't provide new information, they don't represent substantial increases in our knowledge. Thus, it is not surprising that Kant regarded both geometry and arithmetic to be examples of *the synthetic a priori*.

Although Frege agreed that the distinction between analytic and synthetic truths is both important and not coextensive with the distinction between the a priori and the a posteriori, he was critical of the assumption—implicit in Kant's formulation of the analytic/synthetic distinction—that all judgments are of subject-predicate form.⁴⁰ This was a natural assumption for Kant to make during a time in which the traditional Aristotelian logic of generality still held sway, while being clearly unacceptable to the man who showed us the logical and philosophical benefits to be gained from abandoning it. Once the assumption was abandoned, it was clear to

³⁸ "On Function and Concept," pp. 21–23 in Geach and Black (1970), my emphasis.

³⁹ See Immanuel Kant (1781, 1787), A6–7/B10.

⁴⁰ See Frege (1950), sec. 88, p. 100.

Frege that Kant's division of judgments into analytic and synthetic simply left out those that are not of subject-predicate form. Hence, the distinction had to be redrawn.

For Frege, an analytic truth is either a logical truth or a consequence of a set of logical truths plus one or more correct definitions of nonlogical notions. A synthetic truth is any truth that isn't analytic. Since he took logical truths to be a priori, while implicitly assuming that correct definitions are themselves a priori (and that logical consequences of a priori truths are always themselves a priori), Frege naturally regarded all analytic truths as a priori.⁴¹ However, since he also realized that logical consequences of truths, each of which we already know, may be decidedly nontrivial, he recognized that analytic truths can sometimes be highly informative extensions of our knowledge. Thus, the informativeness of arithmetic and other branches of mathematics was, for Frege, no bar to their being analytic.

This was the backdrop for his view that arithmetic, and much else, is analytic. His agreement with Kant, whom he held in high regard, that Euclidean geometry is synthetic a priori is more surprising.⁴²

There is a remarkable difference between geometry and arithmetic in the manner in which they justify their principles. The elements of all geometrical constructions are intuitions, and geometry points to intuition as the source of all its axioms. Since the axioms of arithmetic have no intuitiveness, its principles cannot be derived from intuition.⁴³

To understand this passage, one must understand what Frege, following Kant, means by "intuition." When he claims that the axioms of arithmetic "have no intuitiveness," he is not saying that they aren't self-evidently obvious. He is saying that their correctness isn't guaranteed by Kantian "intuitions," which are presentations of objects in perception, introspection, and imagination. For Kant space and time are categories imposed

⁴¹Though these assumptions are correct within the sphere in which Frege uses them, they are connected to more contentious and complex issues about the justification of definitions that will be examined in chapter 2.

⁴²*Foundations of Arithmetic*, sec. 89 (Frege 1950, pp. 101–2):

I have no wish to incur the reproach of picking petty quarrels with a genius to whom we must all look up with grateful awe; I feel bound, therefore, to call attention also to the extent of my agreement with him, which far exceeds any disagreement. . . . I consider Kant did great service in drawing the distinction between synthetic and analytic judgments. In calling the truths of geometry synthetic and a priori, he revealed their true nature. . . . If Kant was wrong about arithmetic, that does not seriously detract, in my opinion, from the value of his work. His point was, that there are such things as synthetic judgments a priori; whether they are to be found in geometry only, or in arithmetic as well, is of less importance.

⁴³Frege (1874), at p. 50 of Angelelli (1967).

on our experience as forms in which things are presented to us. It is in the nature of our minds to perceive events as temporally ordered, and objects as arranged in Euclidean space. It is because no other experience of objects is possible, or even perceptually imaginable, that the truth of Euclidean geometry is guaranteed, a priori, to be true of the world *as we experience it*. Frege seems to have accepted this. But he didn't accept Kant's view that arithmetic is similarly guaranteed to be true by our temporal "intuitions"—i.e., by the way our minds must, of necessity, perceive events.

Two factors seem to have played a role in his divergence from Kant on this point.⁴⁴ First, whereas geometry applies only to things in space, our conception of which may be constituted by our "spatial intuitions," arithmetic is not limited to the temporal or the spatial, since everything whatsoever, including purely abstract objects, can be numbered.⁴⁵ Second, whereas in geometry Frege thinks we are able to see individual figures as representing all other perceptually identical figures, and hence to draw conclusions about the whole class from what we see to be true of the representing instance, the inherent peculiarities of each individual number don't allow us to use one to represent all the others.⁴⁶ For these reasons, he concludes, arithmetic can't be given the kind of Kantian grounding that Euclidean geometry can. Thus, a different grounding, in logic, is required.

Looking back on Frege today, the most difficult aspect of his view for us to fathom is not his idea that arithmetic is a priori in virtue of being analytic, but that geometry is also a priori, despite its being synthetic. Non-Euclidean geometries had been around for more than 50 years at the time of the *Begriffsschrift*. For decades some had speculated that they were more accurate depictions of physical space than Euclidean geometry, and in 1868 models of non-Euclidean systems were produced, demonstrating their consistency. Why, we now ask, did Frege think that Euclidean geometry is a priori? In section 14 of *The Foundations of Arithmetic*, he hints at an answer.

Empirical propositions hold good of what is physically or psychologically actual, the truths of geometry govern *all that is spatially intuitable, whether actual or product of our fancy*. The wildest visions of delirium, the boldest inventions of legend and poetry . . . all these remain, so long as they remain *intuitable*, still subject to the axioms of geometry. *Conceptual thought alone can after a fashion shake off this yoke*, when it assumes, say, a space of four dimensions or

⁴⁴ See Currie (1982), pp. 34–37, for an illuminating discussion.

⁴⁵ "[T]he ideal as well as the real, concepts as well as objects, temporal as well as spatial entities, events as well as bodies, methods as well as theorems; even numbers can themselves be counted. . . . [T]he basic propositions on which arithmetic is based cannot apply merely to a limited area whose peculiarities they express in the way in which the axioms of geometry express the peculiarities of what is spatial; rather these basic propositions must extend to everything that can be thought." Frege (1885), pp. 141–42.

⁴⁶ See Frege (1950), sec. 13, pp. 19–20.

positive curvature. To study such conceptions is not useless by any means; but it is to leave the ground of *intuition* entirely behind. . . . For purposes of conceptual thought we can always assume the contrary of some one or other of the geometrical axioms, without involving oneself in any self-contradictions when we proceed to our deductions, *despite the conflict between our assumptions and our intuition*. The fact that this is possible shows that the axioms of geometry are independent of one another and the primitive laws of logic, and consequently are synthetic.⁴⁷

Frege recognizes that non-Euclidean geometries are logically consistent, and hence that Euclidean geometry is synthetic, but he denies that our knowledge of it is empirical, and so neither certain nor a priori. Euclidean geometry can't be empirical because it applies to all genuinely conceivable—i.e., imaginable—space, as opposed merely to actual space. Like Kant, he took Euclidean geometry to be grounded in our spatial intuitions, which provide a priori certainty about space *as we experience it*. On this view, the physical space we experience *must be* as we imagine it to be. Euclidean geometry provides knowledge of this space. Although he grants that competing conceptions of space are conceptually coherent (in the narrow sense of being logically consistent) and so can be abstractly investigated, he doesn't take the idea that they could be true of space as it really is very seriously, presumably because he finds it hard to credit the thought that *space as it really is* could differ in fundamental respects from *space as we do, and must, experience it*.⁴⁸

Our main concern is, of course, not with Frege's questionable views of geometry, but with his contrast between it and arithmetic, which is

⁴⁷ Ibid., pp. 20–21, my emphasis.

⁴⁸ Frege's attachment to Euclidean geometry is illustrated by the following passage from p. 169 of "On Euclidean Geometry," written between 1899 and 1906, published in Frege (1979).

If Euclidean geometry is true, then non-Euclidean geometry is false, and if non-Euclidean geometry is true then Euclidean geometry is false. . . . People at one time believed they practiced a science, which went by the name of alchemy; but when it was discovered that this alleged science was riddled with error, it was banished from among the sciences. . . . The question at the present time is whether Euclidean or non-Euclidean geometry should be struck off the role of the sciences and made to line up as a museum piece alongside alchemy and astrology as mummies. If one is content to have only phantoms hovering around one, there is no need to take the matter so seriously, but in science we are subject to the necessity of seeking after truth. There it is a case of either in or out! Well, is it Euclidean or non-Euclidean geometry that should get the sack? That is the question. Does one dare to treat Euclid's elements, which have exercised unquestioned sway for over 2000 years, as we have treated astrology? It is only if we do not dare to do this that we can put Euclid's axioms forward as propositions that are neither false nor doubtful. In that case, non-Euclidean geometry will have to be counted amongst the pseudo-sciences, to the study of which we still attach some slight importance, but only as historical curiosities.

emphasized in the lines immediately following, and completing, those of the passage just quoted.

Can the same be said of the fundamental propositions of the science of number? *Here we have only to try denying any one of them, and complete confusion ensues. Even to think at all seems no longer possible.* The basis of arithmetic lies deeper, it seems, than that of any of the empirical sciences, and even than that of geometry. The truths of arithmetic govern all that is numerable. *This is the widest domain of all; for to it belongs not only the actual, not only the intuitable, but everything thinkable.* Should not the laws of number, then, be connected very intimately with the laws of thought?⁴⁹

This is an echo of something we saw above in section 4, where we discussed Frege's idea that the fundamental laws of logic are both self-evident (and so not in need of or susceptible to any justification) and also essential norms of thought, which we are incapable of rejecting "if we are not to throw our thought to confusion and in the end renounce judgment altogether."⁵⁰ This is connected to their absolute generality, applying, as he takes them to do, not just to all that is plus all that could be visualized, but also to all that can be conceived or thought. For Frege, the reason that arithmetic shares this status with logic is that it is derivable from the combination of pure logic with the logical definitions of arithmetical terms. It is time to make good on this idea.

5.3. Frege's Definition of Number

5.3.1. NUMERICAL STATEMENTS ARE ABOUT CONCEPTS

In determining the sense and referents of numerical expressions, we can't appeal to mental images of numbers, or to objects we have in mind when we use such expressions. It is unlikely that we have definite images of any number, or group of objects, in mind when we use any numeral, and it is obvious that we don't have them in mind for every numerical expression we use. Thus, Frege concluded, we can't first determine what the numbers are, and then use this identifying knowledge to analyze numerical statements. But we do have a good sense of what we use such statements to say, or assert. Thus, he reasoned, the best strategy is to abstract a definition of number from the contents of these statements.

He starts down this road in section 46 of *The Foundations of Arithmetic* when he says:

It should throw some light on the matter to consider number *in the context of a judgment* which brings out its basic use. While looking at one and the same external phenomenon, *I can say with equal truth* both "It is a copse"

⁴⁹ Frege (1950), sec. 17, p. 21.

⁵⁰ *The Basic Laws of Arithmetic*, vol. 1, Preface, sec. 17, in Frege (1997), p. 204.

and “It is five trees,” or both “Here are four companies” and “Here are 500 men.” Now what changes here from one judgment to the other is neither any individual object, nor the whole, the agglomeration of them, but rather *my terminology*. *But that is itself only a sign that one concept has been substituted for another. This suggests . . . that the content of a statement of number is an assertion about a concept.*⁵¹

By ‘concept’ Frege means neither a word, nor an agent’s subjective idea or mental content associated with a word. Rather it is an objective entity denoted by a predicate—a function from objects to truth values. The idea, in a nutshell, is that the content of the statement we make when we say, e.g., that Jupiter has four moons, is that given by (12).

12. The number of entities falling under the concept *moon of Jupiter* = 4.

5.3.2. BUT NUMBERS ARE OBJECTS

Example (12) is an identity statement flanked by a pair of singular terms. Since the referents of such terms are objects, for Frege, this means that if the statement is true, then the terms ‘the number of entities falling under the concept *moon of Jupiter*’ and ‘4’ must refer to the same object. What is this object? We haven’t yet been told. After emphasizing this in section 56 of *The Foundations of Arithmetic*, in section 58 Frege considers the objection that, because nothing we can picture, or imagine, seems to be an apt candidate for this number, it is a mistake to suppose that numbers are objects at all—which he rejects on the grounds that there are many meaningful words that are not uniformly connected by speakers with specific images, or definite (subjective) ideas. The lesson he draws from this is what has come to be called his “Context Principle,” stated in section 60.

That we can form no idea of its [a number term’s] content is therefore no reason for denying all meaning to a word, or for excluding it from our vocabulary. We are indeed only imposed on by the opposite view because we will, when asking for the meaning of a word, consider it in isolation, which leads us to accept an idea as the meaning. Accordingly, any word for which we can find no corresponding mental picture appears to have no content. But we ought always to keep before our eyes a complete proposition [sentence]. *Only in a proposition have the words really a meaning.* It may be that mental pictures float before us all the while, but these need not correspond to the logical elements in the judgment. *It is enough if the proposition taken as a whole has a sense; it is this that confers on its parts also its contents.*⁵²

The Context Principle, expressed by the two italicized sentences, seems to be that (i) the notion of a sentence having a meaning (which Frege identifies with the claim it is used to assert or express) is explanatorily

⁵¹ Frege (1950), p. 59, my emphasis.

⁵² Ibid., p. 71, my emphasis.

primary, while (ii) what it is for a word or phrase to have a meaning is to be explained in terms of what it contributes to the meanings of sentences containing it. Somehow, in a manner not further clarified by Frege, this is supposed to disarm the objection that number words (including numerals) can't designate objects, because we can't—simply by understanding them—identify which objects they are supposed to designate.

But how, precisely, is the Context Principle supposed to do this? Perhaps the idea is that, given fixed meanings for sentences, we may employ one or more ways of abstracting word meanings from sentence meanings so as to see the latter as constituted by the former. In the case of number, this would mean (i) determining what numerical sentences mean (as used both in mathematics and ordinary life), and hence what senses/thoughts they express, (ii) constructing definitions of 'zero', 'successor', 'natural number', and other arithmetical notions, irrespective whether, when we consider the definitions in isolation, we can recognize the referents presented by the *definienda* as what we pre-theoretically had in mind, (iii) articulating principles for constructing senses of arithmetical sentences out of the senses assigned to their parts, and (iv) showing that the senses yielded by (ii) and (iii) are, in fact, those identified in (i). Of course, Frege doesn't exactly do this, in part because at the time he wrote *The Foundations of Arithmetic*, he hadn't yet developed his theory of the senses of expressions and sentences. However, that isn't the whole story, since—even given his later theory—it is far from clear that anyone could do the job.

Frege does, of course, provide the definitions alluded to in (ii), and use them to derive the axioms of arithmetic from (what he takes to be) axioms of logic. In so doing, he also provides convincing reasons to believe that his definitional counterparts of ordinary numerical sentences are (outside of hyper-intensional contexts like those provided by belief and other propositional attitude ascriptions) both necessarily and a priori equivalent in truth value to numerical statements we all, pre-theoretically, understand. Whether or not this is enough to justify the claim that he has *correctly and accurately defined* the relevant numerical notions is, perhaps, the most important question to ask about Frege's philosophical conception of his project. *If* we take the Context Principle as implicitly embodying the claim that such equivalence *is* sufficient to justify the correctness and accuracy of his analysis, then, as Currie points out, this enhanced understanding of the Principle will not only disarm the objection that prompted it, but also express the governing idea behind both Frege's own philosophical understanding of his project, and his criticism of other accounts of the nature of arithmetical claims.⁵³ Examining the contentious criterion of analytic success embodied in this enhanced understanding of the Context Principle is a matter for chapter 2. At this stage, nothing so far-reaching is needed. Since, by the time the Context Principle is stated

⁵³ See pp. 151–56 of Currie (1982).

in *The Foundations of Arithmetic*, Frege has already convincingly shown that many words we properly take to designate objects are not uniformly connected by speakers with specific images or definite subjective ideas, he is free to proceed with his definitions of number.

5.3.3. OBJECTS AND IDENTITY

Having disposed of the objection to taking numbers to be objects, even though we have no antecedent idea what they are, Frege takes up another objection in section 62.

How, then, are numbers to be given to us, if we cannot have any ideas or [Kantian] intuitions [perceptions/apprehensions] of them? Since it is only in the context of a proposition [sentence] that words have any meaning, our problem becomes this: To define the sense of a proposition in which a number word occurs. That, obviously, leaves us still a very wide choice. But we have already settled that number words are to be understood as standing for self-subsistent objects [which can't be referents of predicates—p. 72]. *And that is enough to give us a class of propositions which must have a sense, namely those which express our recognition of a number as the same again. If we are to use the symbol a to signify an object, we must have a criterion for deciding in all cases whether b is the same as a, even if it is not always in our power to apply the criterion.* In our present case, we have to define the sense of the proposition 'the number which belongs to the concept F is the same as that which belongs to the concept G.'⁵⁴

Here, Frege takes a good idea too far. Having used the occurrence of numerical expressions in identity statements as the basis for his claim that their referents—numbers—are objects, he can defend this position by providing a criterion for assigning truth values to identities like 'The number of entities falling under the concept F = the number falling under the concept G' the results of which we recognize to be correct. This is precisely what he does in the sections immediately following the quoted passage in his discussion of a principle he attributes to David Hume. However, his *general* comments about the need for a criterion of identity *whenever* we use singular terms to refer to objects are highly contentious, and go far beyond the Humean principle he needs. In both empirical science and everyday life we often use singular terms for things without having any nontrivial criterion for distinguishing those things from all others. Think for example of oceans or mountain ranges. Surely there are such things, and singular terms referring to them—despite the fact that there is no non-question-begging criterion known to anyone that correctly and precisely specifies where one ocean (or mountain range) begins and another ends. By the same token, there is no known criterion the application of which (even by God) would distinguish all true identity statements involving oceans or mountain ranges from all false ones. The same is, of course, true

⁵⁴ Frege (1950), p. 73, my emphasis.

of many other things we use singular terms to designate. We will return to this point in our discussion in chapter 2 of Frege's notion of a logically perfect language. Here it is enough to note that his definition of number need not be made to depend on any such over-the-top *no entity without a criterion of identity* claim.

Hume's Principle, discussed in sections 64–66 of *The Foundations of Arithmetic*, is a way of understanding, and assigning intuitively correct truth values to, sentences 'the number of F's = the number of G's'—without presupposing any prior understanding of numerical terms.⁵⁵ According to the principle, for all concepts F and G, the number belonging to F is identical with the number belonging to G iff the extension of F (the class of things falling under F) can be put in one-to-one correspondence with the extension of G. This simple and commonsensical idea gives correct results for all finite classes, while being extendable to infinite classes in a natural way. The number belonging to F is identical with the number belonging to G iff one can exhaust the class of things falling under F and the class of things falling under G by forming pairs the first of which is a member of the class of things falling under F and the second of which is a member of the class of things falling under G, where no member of either class occurs in more than one pair.

Lest one be concerned that numerical notions have been illicitly presupposed in this explanation, one-to-one correspondence may be defined as follows.

One-to-One Correspondence

For all concepts F and G, the extension of F (the class of things falling under F) is in one-to-one correspondence with the extension of G (the class of things falling under G) iff for some relation R, (i) for every object x such that Fx, there is an object y such that Gy & Rxy, and for every object z if Gz & Rxz, then z = y, and (ii) for every object y such that Gy, there is an object x such that Fx & Rxy, and for every object z if Fz & Rzy, then z = x.

Since this definition neither contains nor presupposes any arithmetical terms, the relation of one-to-one correspondence it defines can be used in Hume's Principle without presupposing the notion of number Frege aims to explicate. Some terminology may be helpful. When two extensions (classes) are in one-to-one correspondence, we will say they are *equinumerous*, and when the extensions of two concepts are equinumerous the

⁵⁵ In this sentence 'F' and 'G' are used as metalinguistic variables over predicates/formulas; in the rest of the paragraph, they are used as second-order variables ranging over Fregean concepts.

concepts will be called *equal*. With this understanding, Hume's Principle tells us that for all concepts F and G , the number belonging to F is identical with the number belonging to G iff F *equals* G iff the extension of F is *equinumerous* with the extension of G . So, when \mathcal{F} and \mathcal{G} are predicates (formulas) designating concepts F and G , respectively, \lceil the number of \mathcal{F} s = the number of \mathcal{G} s \rceil is true iff the extension of F is equinumerous with the extension of G .

Since this criterion assigns correct truth conditions to the target class of sentences, \lceil the number of \mathcal{F} s = the number of \mathcal{G} s \rceil , without presupposing any antecedent understanding of number, it suits Frege's purposes. However, even though it correctly accounts for the truth of the sentence 'the number of fingers on my right hand = the number of fingers on my left hand', for example, it does not identify the object to which the singular terms flanking the identity sign refer. By the same token, it tells us nothing about the numeral '5', or sentences containing it, and leaves us in the dark about what the number 5, or any other number, is. These shortcomings are remedied by Frege's definition of number.

5.3.4. THE NUMBER OF F'S, ZERO, SUCCESSOR, AND THE NUMERALS

Frege's definition of number, from section 68, is: For any concept F , the number that belongs to F is the extension of the concept *equal to F*.⁵⁶ Let's unpack this. The extension of a concept is the class of entities that fall under it. For any concept F , the entities that fall under the concept *equal to F* are those the extensions of which are equinumerous with the extension of F . So, the number that belongs to F is the class of all and only those concepts the extensions of which are classes equinumerous with the class of things of which F is true. So, the number of fingers on my right hand is the class of all those concepts the extensions of which are classes equinumerous with the class of fingers on my right hand—it is the class of concepts that apply to five things, or, if we ignore the difference between concepts and the classes of things that fall under them, it is the class of all five-membered classes. Although we can see from the definition that this is so, we are, of course, not yet allowed to use a numeral like 'five', since no definitions have yet been given for the numerals. However, this is easily remedied.

We begin with Frege's definition of zero as the number that belongs to the concept *not identical with itself*. So, zero is the class of all and only those concepts the extensions of which can be put into one-to-one correspondence with the class of things that are not identical with themselves. Since there are no objects that are not identical with themselves, this means that zero is the class of concepts the extensions of which can be put into one-to-one correspondence with the empty class—i.e., the class with no members. So, zero is the class of concepts that don't apply to anything.

⁵⁶ In this section, I return to using 'F', 'G', etc. as variables. Here, 'F' is a variable over Fregean concepts.

Next we define the notion of *n directly following* (i.e., succeeding) *m*. According to Frege, *n* directly follows (succeeds) *m* iff for some concept *F*, and some object *x* falling under *F*, *n* is the number belonging to *F*, and *m* is the number belonging to the concept *falling under F but not identical with x*. When *m* is zero, this means that *n* directly follows (succeeds) zero iff for some concept *F* and object *x* falling under *F*, *n* is the number belonging to *F*, and zero—namely the class of concepts that don't apply to anything—is the number belonging to the concept *falling under F but not identical with x*. Since this concept doesn't apply to anything, *n* is a class of concepts each of which applies to some object *x*, and to nothing else. Assuming something I will say more about in section 6—namely that there is one and only one such class—we have identified *the unique object* that directly follows (succeeds) zero. In general, we define the successor of *m* as the unique object that directly follows (succeeds) *m*. As for numerals, '1' designates the successor of 0, '2' designates the successor of 1, etc. So, 0 is the class of concepts under which nothing falls, 1 is the class of concepts under which, for some *x*, *x*, and only *x*, falls, 2 is the class of concepts under which some nonidentical objects *x* and *y*, and only those objects, fall, and so on. From this it is transparent that 0 is the class of concepts that apply to nothing, 1 is the class that apply to exactly one thing, 2 is the class that apply to exactly two things, and—in general—*n* is the class of concepts that apply to exactly *n* things. Note, we do not presuppose an antecedent understanding of the numerals in deriving this result. Rather, the result is an automatic consequence of the Fregean definitions.

Two features speak to the naturalness of what Frege has done.

- (i) Just as redness isn't identical with any red thing, but rather is something all red things have in common, so a particular number *n* is not identical with any collection of *n* things (or with any concept under which just those *n* things fall); rather it is something all concepts applying to *n* things have in common—membership in the number *n*.
- (ii) Just as counting any collection of things consists in putting them in one-to-one correspondence with the numerals, starting with '1', used in the count, so such correspondence is the crucial notion in defining each of the numbers.

5.3.5. THE NATURAL NUMBERS

Having defined '0' and 'successor', plus individual numerals, the next step is to define what we mean by the "natural numbers," which constitute the domain of arithmetic. The simplest thought is that a natural number is something that is a member of every class that contains zero and is closed under successor (i.e., that contains the successor of *x* whenever it contains *x*). We can put this in terms of concepts by first defining *an inductive concept* as one that is true of zero and closed under successor (i.e., is true of the successor of *x* whenever it is true of *x*). A *natural number* is then one

that falls under all inductive concepts. Although this is not identical to the definition Frege gave, it is equivalent to it. The virtue of the definition (and its Fregean equivalent) is that it makes it easy to prove *mathematical induction*. Since this is a crucial axiom of arithmetic, proving it from Frege's system of logic was a major step in his program. To see that mathematical induction must be true on the suggested definition, let F be any concept true of zero and closed under successor. Since F satisfies the definition of an inductive concept, and since, by definition, the natural numbers fall under every such concept, F is true of every natural number. Q.E.D.

Frege's own definition of natural number (and proof of mathematical induction) is given in the *Begriffsschrift*, published five years before *The Foundations of Arithmetic*. Though equivalent to the one just mentioned, it is based on related ideas that are independently worth knowing. We begin by noting the connection between the relations *parent* and *ancestor*: any parent of x is an ancestor of x , as is any parent of any ancestor of x . Because this condition holds for any x , the *ancestor* relation is called the *transitive closure* of the *parent* relation. We can put this more generally, using T to define what it is for a relation to be transitive and TC to define the transitive closure R_{TC} for arbitrary R .

$$T. \forall x \forall y \forall z ((Rxy \ \& \ Ryz) \supset Rxz)$$

$$TC. \forall x \forall y (R_{TC}xy \text{ iff } \forall P [(\forall z (Rxz \supset Pz) \ \& \ \forall u \forall v ((Pu \ \& \ Ruv) \supset Pv)) \supset Py])$$

Applying these to *parent* and *ancestor*, we see that the former isn't transitive, while the latter is, and that when Rxy is the relation *y is a parent of x*, $R_{TC}xy$ is the relation *y is an ancestor of x*. For (going left to right in TC) if y is an ancestor of x , then, for any concept P , y will fall under P , provided that (i) every parent of x falls under P , and (ii) any parent of someone who falls under P also falls under P ; and (right to left in TC) if for any concept P whatsoever, y will fall under P provided that (i) every parent of x falls under P , and (ii) any parent of someone who falls under P also falls under P , then—given all this— y must be an ancestor of x . For this reason, the transitive closure of an intransitive relation R is often called *the ancestral of R*. It is this notion of the ancestral of a relation that Frege used in the *Begriffsschrift* to establish mathematical induction.

Let Rxy be the relation that holds between zero and its successor, and in general between any extension of a concept and its successor—using Frege's definition of successor (i.e., *directly follows*) given above. The ancestral of this relation is the relation *y follows x (in a series under successor)*, or, in other words, *y is greater than x*. We can then define the natural numbers as those objects (extensions of concepts) that are greater than or equal to zero. With this definition in mind, we return to mathematical induction, which states that if zero falls under a concept P that is closed under successor, then every natural number falls under P . To prove this we assume the antecedent—that zero falls under P and P is closed under successor—and show that every natural number must fall under P . Using the notation in TC , we express the antecedent as $P(0) \ \& \ \forall u \forall v ((Pu \ \& \ Ruv) \supset Pv)$, with Rxy as the successor

relation y directly follows x . Since *greater than* is the ancestral, R_{TC} , of successor, R , TC tells us that everything greater than zero falls under P . Since, by definition, the natural numbers are zero plus everything greater than zero, it follows that every natural number falls under P . This is the import of Frege's theorem 81 in the *Begriffsschrift*. It is also the import of his discussion in sections 79–81 of *The Foundations of Arithmetic*, where (i) (in section 79) he repeats his definition of the ancestral— y following in the ϕ -series after x —of a relation ϕ , (ii) (in section 81) he takes the natural numbers to be zero plus those that follow in the ϕ -series after zero (where ϕ is the relation directly following/successor), and (iii) (in section 80) he comments that “Only by means of this definition of following in a series [i.e., the ancestral of the successor relation] is it possible to reduce the argument from n to $(n + 1)$, which is peculiar to mathematics, to the general laws of logic.”⁵⁷

6. THE LOGICIST REDUCTION

The derivations of the axioms of arithmetic from what Frege took to be the basic laws of logic are carried out in daunting and meticulous detail in *The Basic Laws of Arithmetic*. No such exhaustive treatment will be given here. Instead, the leading ideas will be presented informally, but in enough detail to give the reader an idea of the strategies used to derive the crucial results.

6.1. The Axioms of Logic and Arithmetic

We begin with the system of logic Frege used. The first part of the system consists of axioms and inference rules for proving standard logical truths in the sense, recognized today, of formulas that come out true on all interpretations of their nonlogical symbols, and all choices of domains of quantification. As explained in section 4, Frege's system for proving such truths was as effective as any we now have. Since I won't here be constructing explicit formal proofs, there is no need to go into details of this aspect of either his, or equivalent, systems. However, something must be said of other “logical principles,” implicit or explicit in Frege's system.

Since meaningful predicates/formulas denote concepts that determine which objects they are true of, the comprehension principle for concepts is taken for granted.

Concept Comprehension

For every stateable condition ϕ on things, there exists a concept C that is true of all and only those things that satisfy the condition $\exists C \forall y (Cy \leftrightarrow \phi y)$.

⁵⁷ Frege (1950), p. 93.

A further principle explicates Frege’s conception of what concepts are—namely, assignments of truth values to objects. As we saw in section 3, he took this to be all there is to concepts, and so took concepts that assign the same values to the same arguments to be identical—which is what *Concept Extensionality* tells us.

Concept Extensionality

Concepts P and Q are identical iff everything that falls under one falls under the other: $\forall P\forall Q (P = Q \text{ iff } \forall x (Px \rightarrow Qx))$.

Being inherent in his understanding of his symbolism, these two principles about concepts didn’t require separate statement. But when it came to the derivation of the axioms of arithmetic, he did require a special “logical” axiom—numbered V in his system—that guarantees extensionality and comprehension for *extensions of concepts*—i.e., classes.

Axiom V

For all (first-level) concepts P and Q, the extension of P (the class of things falling under P) = the extension of Q (the class of things falling under Q) iff $\forall x (Px \leftrightarrow Qx)$.

This gives us *comprehension for classes* since ‘ $\forall P\forall x (Px \leftrightarrow Px)$ ’ will always be true, which, by Axiom V, means that the class of things falling under P is identical with the class of things falling under P. Although this doesn’t require the class to be nonempty, it does require there to be a class of all and only those things of which P is true. Of course, Axiom V also guarantees *extensionality for classes*—identifying classes with the same members. This special axiom is ubiquitous in Frege’s proof of the axioms of arithmetic from those of logic.

Prior to Frege, a complete set of axioms for arithmetic had been given by the mathematician Richard Dedekind, who appears to have come up with it in 1888. Nevertheless, the resulting system is usually called *Peano arithmetic*, after Giuseppe Peano, who published an influential work (with a footnote to Dedekind) in 1889. The axioms can be given different formulations, depending on whether *successor* is stated as a totally defined function, or simply a two-place relation. In order to be as explicit as possible in indicating what must be proved from the logic, I have selected the latter option, which requires the presence of Axiom A2, which, together with A5, establishes that successor is a function defined on all natural numbers. With this proviso, the axioms of Peano arithmetic are:

- A1 Zero isn’t a successor of anything. $\sim \exists x Sx0$
- A2 Nothing has more than one successor. $\forall x\forall y\forall z ((Sxy \ \& \ Sxz) \supset y = z)$

- A3 No two things have the same successor: $\forall x \forall y \forall z ((Sxy \ \& \ Szy) \supset x = z)$
 A4 Zero is a natural number: $NN0$
 A5 Every natural number has a successor: $\forall x (NNx \supset \exists y Sxy)$
 A6 A successor of a natural number is a natural number: $\forall x \forall y ((NNx \ \& \ Sxy) \supset NNy)$
 A7 Mathematical Induction: If zero falls under a concept, and a successor of something that falls under a concept always falls under the concept, then every natural number falls under the concept. $\forall P [(P0 \ \& \ \forall x \forall y ((Px \ \& \ Sxy) \supset Py)) \supset \forall x (NNx \supset Px)]$

6.2. Informal Proofs of the Arithmetical Axioms

We start with A1, which, given the definition of *successor*, states that there is no concept F and object x falling under F such that (i) zero is the number belonging to F, and (ii) there is a y that is the number belonging to the concept *falling under F but not identical with x*. The proof is trivial. Since zero is, by definition, the class (the existence and uniqueness of which is guaranteed by Axiom V) the only member of which is the concept under which nothing falls, the extension of that concept (the empty class) can't be put into one-to-one correspondence with the extension of any concept F under which something falls. So, zero can't be the number belonging to F.

Next consider A2. In order for x to have nonidentical successors y and z, there would have to be concepts F and G such that (i) y = the number belonging to F, z = the number belonging to G, and $y \neq z$, and (ii) for some object o_F falling under F and some object o_G falling under G, x = the number belonging to the concept *falling under F but not identical with o_F* and x = the number belonging to the concept *falling under G but not identical with o_G* . This could be true only if the extensions of the concepts F and G could not be put into one-to-one correspondence, but the results of removing a single item from each could be put into such correspondence. Since this is impossible, A2 must be true. This shows that the successor relation is indeed a function.

A3 says that no two different things have the same successor. In order for two such things x and y to have the same successor z there must be concepts F and G and objects o_F and o_G such that (i) z = the number belonging to F = the number belonging to G, (ii) x = the number belonging to the concept *falling under F but not identical with o_F* and y = the number belonging to the concept *falling under G but not identical with o_G* , and (iii) $x \neq y$. (i) tells us that the extensions of the concepts F and G can be put into one-to-one correspondence, while (ii) and (iii) tell us the extensions that result from removing a single item from each can't be put into such correspondence. Since this is impossible, A3 is true. Given A2, A3 tells us that successor is a 1-1 function.

A4 says that zero is a natural number, which, by Frege's official definition, means that zero falls under the concept *equals zero or follows zero*

in a series under successor. Hence A4 is true by definition (plus Axiom V). Although A5 is more complicated, we can still see why it must be true. We start with zero, which is the number belonging to the concept *not identical with itself*. From Axiom V, we know that there is a class of all and only those things that are not identical with themselves. Since this class is empty, zero is the class of concepts the extensions of which are equinumerous with the empty class—i.e., the class of concepts under which nothing falls. Given his conception of concepts, this is the class the only member of which is the concept that assigns the False to every argument. Next, the successor of zero is, by definition, the number belonging to a concept F under which an object x falls, such that the concept *falling under F but not identical with x* is a member of zero. Can we be sure, on the basis of Frege's logic alone, that there is such a concept F and object x? Yes. We already have zero, and we know there is a concept *being identical with zero*. This plus Axiom V guarantees that there is a class that is the number belonging to this concept, and hence that there is a class of concepts the extensions of which are equinumerous with the class the only member of which is zero. This class of concepts under which exactly one thing falls is the successor of zero—i.e., the number 1. Since 1 follows zero in the series under successor, it satisfies the definition of being a natural number. Similar reasoning—this time using the concept *being identical with either with zero or 1*—establishes that 1 has a successor—the number 2—which is the class of concepts under which pairs of nonidentical things, and nothing else, fall. Since 2 also follows zero in the series under successor, it is also a natural number. Similar results can, in this way, be established for each n. So, each instance of A5 is derivable using Frege's logic plus definitions. Although this isn't itself a proof of the universal generalization A5, Frege found a way of turning it into one. With this, we are assured that the successor function is totally defined on the natural numbers.

This brings us to A6, which, in the presence of A5, is trivial. For suppose that x is a natural number. Then, x is either zero or something that follows zero in the series under successor. A5 tells us that x has a successor y. But then y must follow zero in the series under successor—which, by the definition of *natural number*, guarantees that y is a natural number. As for A7, mathematical induction, we have already seen how Frege's definition of *natural number* guarantees its truth.

6.3. Arithmetical Operations

Frege's achievement was to show how to derive the axioms of Peano arithmetic from his system of logic. However, Peano arithmetic consists of more than a set of axioms characterizing the natural numbers. In addition, it also defines the arithmetical operations *addition* and *multiplication*. The definition of addition is as follows.

Definition of Addition

For any natural numbers x and y , the result of adding zero to x is x ; the result of adding the successor of y to x is the successor of the result of adding y to x .

In formulating this definition in symbols we use '\$' to stand for the function that assigns a number its unique successor. This gives us

$$\forall x \forall y [(NNx \ \& \ NNy) \rightarrow (x + 0) = x \ \& \ (x + \$(y)) = \$(x + y)]$$

This kind of definition is called a *recursive*, or *inductive*, definition. It works by first specifying what it is to add zero to an arbitrary number x , and then specifying what it is to add the successor of a number y to a number x . So, we first determine the sum of zero and x to be x . Then we determine the sum of x and the successor of zero (namely 1) to be the successor of x . Applying the definition again, we determine the sum of x and the successor of 1 (namely 2) to be the successor of the successor of x . The process can be repeated to determine, for each number y , the result of adding y to x . Since x can be any number, the definition completely determines the sum of every pair of numbers, even though it does not have the familiar form of an explicit definition.

The way in which particular results are derived in Peano arithmetic is illustrated by the example: $3 + 2 = 5$.

- (i) $\$ (\$ (\$ (0))) + \$ (\$ (0)) = \$ [\$ (\$ (\$ (0))) + \$ (0)]$
from the definition of '+' together with A4 and A6, which guarantee that $\$ (\$ (\$ (0)))$ and $\$ (\$ (0))$ are natural numbers
- (ii) $\$ (\$ (\$ (0))) + \$ (0) = \$ [\$ (\$ (\$ (0))) + 0]$
from the definition of '+', A4, and A6
- (iii) $\$ (\$ (\$ (0))) + \$ (\$ (0)) = \$ (\$ [\$ (\$ (\$ (0))) + 0])$
from substitution in (i) of equals for equals on the basis of (ii)
- (iv) $\$ (\$ (\$ (0))) + 0 = \$ (\$ (\$ (0)))$
from the definition of '+'
- (v) $\$ (\$ (\$ (0))) + \$ (\$ (0)) = \$ (\$ (\$ (\$ (\$ (0)))))$
from substitution in (iii) on the basis of (iv)

That is how the arithmetical system to be reduced to Fregean logic derives particular arithmetical results. To show that Frege's logical system allows the derivation of results such as these, involving addition, one must show that his logical axioms guarantee there is a (unique) function f that satisfies the pair of equations defining addition: $f(x, 0) = x$ and $f(x, \$(y)) = \$(f(x, y))$ —or, what comes to the same thing, that there is a (unique) three-place relation (concept) $Rxyz$ (intuitively, z is a sum of x and y) that satisfies the pair of formulas $Rx0z \leftrightarrow z = x$ and $Rx\$(y)z \leftrightarrow \exists v(Rxyv \ \& \ z = \$(v))$. Given the strength of the comprehension and extensionality principles

generated by Frege's assumptions about concepts and their extensions, including Axiom V, this is not problematic.

As pointed out in Burgess (2005), once we have this, we can use mathematical induction to establish the associative law for addition— $(x + y) + z = x + (y + z)$ —by (i) establishing associativity when $z = 0$, using the clause defining addition of zero, which gives us $(x + y) + 0 = x + y$, and $(y + 0) = y$, which, together, give us $(x + y) + 0 = x + (y + 0)$, and (ii) showing that associativity holds for the successor of z , provided that it holds for z , using the other clause of the definition of addition. Once this result is established, the commutative law of addition—that $x + y = y + x$ —can be similarly proved, using mathematical induction.

Analogous results hold for multiplication, which is defined in a similar fashion (using '×' as the symbol for the multiplication function).

Definition of Multiplication

For any natural numbers x and y , the result of multiplying x times zero is zero, and the result of multiplying x times the successor of y is the sum of x and the result of multiplying x times y .

$$\forall x \forall y [(NNx \ \& \ NNy) \rightarrow ((x \times 0) = 0 \ \& \ x \times S(y) = (x \times y) + x)]$$

This works the same way the definition for addition does. Thus, multiplication is defined in terms of repeated addition, which in turn is defined in terms of repeated application of the successor function. As before, there is no difficulty deriving the arithmetic results involving multiplication from Frege's logical system.

6.4. Further Issues

This completes the overview of Frege's reduction of arithmetic to logic, which was, of course, only one step in the grander project of reducing higher mathematics to logic. The next step was to reduce these advanced theories to arithmetic/logic by deriving their axioms from the axioms and definitions of arithmetic/logic plus definitions of the numbers—integers (including negative and positive), real (including rational and irrational), complex, and infinite—required by such theories. Although Frege had much to say about this, others contributed valuable and important work on this as well.

In 1903, when Frege was about to publish volume 2 of *The Basic Laws of Arithmetic*, the prospects for the success of the ambitious logicist project of showing mathematics to be an elaboration of pure logic seemed bright. However, daunting difficulties were to come. In the next chapter we will see that despite Frege's undeniable achievement, all was not

well with the reduction of arithmetic to logic. As Bertrand Russell was to show—informing Frege in a letter in 1902, shortly before the publication of volume 2—there was a paradox lurking in Frege’s crucial, and heavily utilized, Axiom V. Although Frege inserted a hastily constructed fix into the volume, which even Russell initially thought might allow him to avoid contradiction, it soon became clear that it wouldn’t. This proved to be a heavy blow, altering the course of Frege’s future work. However, it wasn’t the end of the story. As we shall see in chapters 7 and 10, Russell drew his own conclusions from the paradox, using a conception of number much like Frege’s to reduce arithmetic to a system of logic, complete with a way—albeit at significant cost—of avoiding the paradox. Much more recently, considerable work has been done within an avowedly Fregean framework to construct a paradox-free version of his derivation of arithmetic, and, in general, to provide a foundation sufficient for much or all of classical mathematics. The best, most exhaustive, and nuanced review and assessment of this work is found in Burgess (2005).⁵⁸

⁵⁸The basic idea behind the paradox-free treatment is to scrap Frege’s Axiom V, with its overly strong comprehension principle for extensions of concepts (sets), and to make due with Hume’s Principle (thought of as an implicit definition of natural number). Doing so changes the proofs needed for the derivation of arithmetic from logic, which still seems to go through. The question then becomes how much mathematics can be reduced to this primitive “logical base.” Burgess’s verdict is, “not enough.”