chapter 7

The Foundational Systems of W. V. Quine

Starting in 1937, W. V. Quine elaborated several systems capable of reproducing a reasonably large portion of mathematics. The first system we shall study is called "New Foundations" or **NF** by abbreviation. Quine first introduced **NF** in a paper presented to the Mathematical Association in 1936, and it was the subject of an article by Quine in 1937 (see Quine [1]). The second system is that of Quine's *Mathematical Logic*, first published in 1940 and revised in 1951. We shall refer to this system as **ML**.

7.1 The System NF

The point of departure for **NF** is, to a great extent, the theory of types, particularly the system **ST**. Quine was strongly influenced by the work of Whitehead and Russell, and had long been seeking some way of liberalizing the type restriction to obtain a more satisfactory foundation. His answer was contained in the notion of *stratification*, which we shall presently define.

NF is a first-order theory with only one category of variables, the individual variables x_1, x_2, and so on. **NF** has only one primitive binary predicate letter which we shall denote by "$\in$". We shall also introduce an operator of abstraction identical with that of our system **F** of Chapter 3. In short, the wffs of **NF** are exactly the same as the wffs of Frege's system **F** of Chapter 3. The proper axioms of **NF** will, of course, be different from **F**. In form, **NF** is probably the closest of any system we study in this book to Frege's contradictory system.

We now turn to the problem of defining what is meant by the stratification of a wff. We begin by considering the wffs that do not contain any occurrence of the abstraction operator. Such wffs involve only quantifiers, variables, propositional connectives, and the primitive predicate "$\in$". We call such wffs of **NF** *simple*. A simple wff is *stratified* if and only if it is possible to replace each variable occurring in it by a whole number numeral in the following manner: We replace everywhere the same variable by the same numeral so that, for each occurrence of "$\in$", the numeral immediately following "$\in$" is the immediate successor of the numeral immediately preceding "$\in$".

A few examples will illustrate the idea. The wffs $(x_1 \in x_2) \supset (x_2 \in x_3)$ and $(x_1 \in x_2) \vee (x_1 \in x_3)$ are both stratified. In the first case, replace x_1 by "1", x_2 by "2", and x_3 by "3". In the second case, replace x_1 by "1", x_2 by "2", and x_3 by "2". Notice that we can replace different variables by the same numeral.

On the other hand, $(x_1 \in x_1)$, $\sim(x_1 \in x_1)$, and $(x_1 \in x_2) \wedge (x_2 \in x_1)$ are all unstratified. The reader will recognize the second of these wffs as the defining condition for the class used in deducing Russell's paradox.

We can formulate an algorithm to determine whether or not a given simple wff is stratified. First, pick some arbitrary variable of the wff and replace it everywhere by 0. Then, inductively: If some $n \in y$ occurs for some numeral n and some variable y, then replace everywhere y by $n + 1$. If some $y \in n$ occurs for a numeral n and a variable y, then replace everywhere y by $n - 1$. If neither of these two cases occurs, then select some arbitrary new variable x and replace it everywhere by 0. Continuing in this manner, we either come to violate the criterion of stratification (and the wff is unstratified), or we continue until all variables have been replaced by numerals and stratification is exhibited.

Clearly, testing for stratification is like restoring type indices. If we considered wffs of **ST** (and for the moment exclude those containing instances of the abstraction operator) as wffs of a first-order theory formed by suppressing type superscripts (replacing with new variables where necessary to avoid confusion), then clearly every such expression would be stratified. In fact, this is precisely the stratagem of "typical ambiguity" employed by Whitehead and Russell. In their work, type indices do not appear and the reader is to restore them in any way consistent with well-formedness. Since, in most cases, it is only relative types that matter in **ST**, it is usually unimportant what type indices (numerals) are actually assigned, as long as it is possible to make an assignment consistent with well-formedness.

However, stratification will not be a criterion of well-formedness in **NF**. This is the essential difference with **ST**. We have already defined the wffs of **NF** as the same as those for **F**, and clearly there are many simple wffs that are not stratified. Stratification will be used in **NF** to define the set of axioms rather than the set of wffs. Nevertheless, because of the obvious similarity with typical ambiguity, we make the following convention: For any wff A of **NF** (simple or not) and any assignment of numerals to the variables of A, the number n that is named by the numeral assigned to a variable x of A will be called the *type* of x under the assignment. Whenever we speak of an "assignment" of numerals, we always understand, as in the foregoing, that the same numeral is assigned to every occurrence of the same variable. Thus, every variable has a uniquely defined type for any assignment of numerals to variables.

We must now extend our definition of stratification to the whole set of wffs of **NF**, those involving abstracts as well as simple ones. For a wff A of **NF** and an assignment of numerals to the variables of A, the *type* of an abstract $\{x \mid B(x)\}$ that is a part of A, under the given assignment of numerals, is $n + 1$ where the type of the variable x is n under the given assignment. By an assignment of types to the *terms* of a wff A, we mean any assignment of numerals to the variables for which the type of every abstract is understood to be its type under the assignment of numerals, and the type of every variable is its type under the assignment of numerals. Finally, a wff A is said to be stratified if there is some assignment of types to the terms of A such that, for every occurrence of "$\in$" in A, the type of the term immediately following "$\in$" is the successor of (one more than) the type of the term immediately preceding "$\in$".

Notice that our way of assigning a type to an abstract is also in line with the abstraction operation in **ST**. When we circumflex a variable of type n in **ST**, we obtain an abstact of type $n + 1$. Of course, we allow the use of negative numbers as types for assignments of indices in **NF**, but this is obviously nonessential.

As examples of stratified wffs involving abstracts, we have

$$x_2 \in \{x_1 \mid (x_3)(x_3 \in x_1 \equiv x_3 \in x_2)\},$$

and $\{x_1 \mid x_1 \in x_2\} \in x_3$. In the first case, let x_3 be 0 and x_1 and x_2 be 1. Then the abstract has type 2 in this assignment and so we have stratification. In the second instance, let x_1 have type 0 and x_2 type 1. The abstract has type 1 and we can assign type 2 to x_3.

As examples of unstratified wffs involving abstracts, we have $x_2 \in \{x_1 \mid x_1 \in x_2\}$ and $\{x_1 \mid x_2 \in x_1\} \in x_1$. In the first case, let n be the type of x_1 under some assignment. Then $n + 1$ must be the type of the abstract and $n + 1$ must be the assignment to x_2 as well. Consequently, stratification is impossible. A similar argument shows the impossibility of stratification in the second case.

If a wff is stratified, then any assignment of types to terms that conforms to the pattern of stratification is called an *adequate* assignment of types to terms. We now turn to the task of stating the axioms of **NF**.

Definition 1. $x = y$ for $(z)(z \in x \equiv z \in y)$, where z is any variable that does not occur in the terms x and y.

Depending on the terms x and y, $x = y$ may or may not be stratified. Notice that in order for $x = y$ to be stratified, there must be some adequate assignment of types such that x and y have the same type, since both x and y must have the type $n + 1$ where z has type n under the assignment.

NF.1 $(x)(y)(x = y \supset (A(x, x) \equiv A(x, y)))$, where $A(x, y)$ is obtained from $A(x, x)$ by replacing y for x in zero, one, or more free occurrences of x in $A(x, x)$, and y is free for x in all of the occurrences of x that it replaces.

This is the same axiom of extensionality as that of Frege's system **F**.

NF.2 $(x)(x \in \{y \mid A(y)\} \equiv A(x))$, where $A(y)$ is stratified and contains y free, x is free for y in $A(y)$, and $A(x)$ results from $A(y)$ by replacing x for y in all the latter's free occurrences in $A(y)$.

This is the axiom of abstraction in **NF**. The distinctive feature of **NF**.2 is the requirement that the wff $A(y)$ be stratified. If we remove this requirement, we obtain immediately the contradictory axiom **F**.2 of Chapter 3. In fact, this requirement constitutes the only difference between **NF** and **F**.

As with Frege's system **F**, as well as with **ST**, we can formulate **NF** without a term operator of abstraction. When the only terms are variables, we could pose:

NF.2* $(Ey)(x)(x \in y \equiv A(x))$ where y does not occur in the stratified wff $A(x)$ and $A(x)$ contains the variable x free.

We avoid this formulation only to lighten the technical matters while working within the system. By the axiom of extensionality, we obtain essentially the same formulation as with our term operator. We have made similar observations with respect to **F**, **ST**, and **ZF**.

By the way we have defined the wffs of **NF**, we can formally apply an abstraction operator to any wff $A(x)$ that contains a free variable x, whether $A(x)$ is stratified or not. If $A(x)$ is stratified, we say that the term $\{x \mid A(x)\}$ is *stratified*. It is only to stratified terms that we can apply **NF**.2, and this is obviously our prime method for dealing with abstracts.

In line with our practice in dealing with other systems, we refer to the terms of **NF** as *sets*. **NF**.2 poses the existence of sets satisfying stratified conditions. This restriction in **NF**.2 is a positive one because it directly poses the existence of certain sets while not directly excluding the existence of other sets that may not be definable by stratified wffs. Thus, Russell's paradox is not directly deducible, since the contradictory Russell condition $\sim(x \in x)$ is not stratified and thus **NF**.2 cannot be applied to it. But this is not a guarantee that a set satisfying some such contradictory condition is not, by some other indirect means, definable in **NF**.

Stratification is a formal, linguistic device. We have not attempted to give an intuitive model for **NF**, nor have we tried to give an informal characterization of sets definable by stratified conditions. The reason is that, as of the present writing (July 1967), no model for **NF** has ever been discovered! This is in direct contrast to both **ST** and **ZF** for which we have intuitive (albeit highly nonconstructive) models.

Now, one theorem of logic deducible within any set theory such as

ZF is that a first-order theory is consistent if and only if it has a model. In the light of this theorem, our failure to find a model for NF after thirty years might be viewed as augmenting the possibility of the inconsistency of NF. As we shall see later, there are other anomalies of NF which seem to undermine one's confidence in the consistency of the system. In particular, the axiom of choice can be disproved within the system, and Cantor's paradox is narrowly avoided.

Still, no contradiction has ever been deduced in NF, and several logicians have made a concerted effort to find such a contradiction. Moreover, we know from Chapter 6 that there are models of ZF without ZF.10 in which the axiom of choice is false. Thus, the failure of the axiom of choice in NF cannot, in itself, exclude the possibility of finding a model of NF in ZF without ZF.10. Also, as we shall see, there is even a way of providing for most uses of the axiom of choice in NF by applying certain limiting conditions.

Of course, failure to deduce a contradiction in NF is no proof that the system is consistent. All this should leave the reader in the same position with respect to the consistency of NF as the author: Doubt and uncertainty. Each of the points mentioned in the above paragraph will be amplified in our later discussion, and we now proceed to the development of mathematics in NF.

Our method of developing mathematics in NF follows Frege rather than von Neumann. Since Frege's constructions were possible in type theory, they will certainly be possible in NF, which is an obvious liberalization of ST. Moreover, von Neumann's natural numbers, which we used in ZF, will not work here. To see this, observe that the set $x \cup \{x\}$, which is the successor of x in ZF, is defined as $\{y \mid y = x \vee y \in x\}$. But this term is unstratified, since y and x must be of the same type (in any adequate assignment of types) in order for $y = x$ to be stratified. Thus, we cannot use NF.2 in dealing with $x \cup \{x\}$.

Definition 2. V for $\{x_1 \mid x_1 = x_1\}$

Theorem 1. $\vdash (x_1)(x_1 \in V)$

PROOF. The condition $x_1 = x_1$ is stratified, and so the proof is the same as the corresponding theorem in Frege's system.

Definition 2 and Theorem 1 show that we have a universal class in NF. NF is the first system we have considered since Frege's contradictory system in which a class containing everything in our universe is definable.

Corollary. $\vdash V \in V$

PROOF. Immediate from Theorem 1.

The proof of Theorem 1 illustrates an important relationship between NF and F. If the defining conditions of the relevant terms are stratified, the proofs of theorems in **NF** will proceed in exactly the same manner as in **F**. This is immediately obvious, since the worry of stratification is the only difference between the two systems. We shall not attempt to maintain any sameness of numberings between theorems and definitions of this Chapter and of Chapter 3, but we shall often refer to a theorem in this chapter as "the same as the corresponding theorem" in Chapter 3. The "corresponding theorem" is the one that asserts the same statement.

Definition 3. $x \subset y$ for $(z)(z \in x \supset z \in y)$ where z is a variable not appearing in the terms x and y.

Definition 4. $P(y)$ for $\{x \mid x \subset y\}$ where x is any variable not appearing in the term y.

Exercise. Prove that $P(y)$ is stratified if y is.

Theorem 2. $\vdash (x_1)(x_1 \subset V)$

PROOF. The reader will prove Theorem 2 as an exercise.

Theorem 3. $P(V) = V$

PROOF. Immediate from Theorems 1 and 2 and the relevant definitions.

Here we have a set in **NF** which is the same as its power set. If Cantor's theorem is provable in **NF**, we have an immediate contradiction, the Cantor paradox. Cantor's theorem (see Chapter 3, Section 3.5) says that the power set of every set has greater cardinality than the set. Since every set has the same cardinality as itself, the existence of a universal set V with the properties of the V of **NF** is inconsistent with Cantor's theorem.

However, Cantor's theorem is apparently not provable in **NF**. In any case, the usual proof of the theorem does not succeed because certain relevant conditions are not stratified. We shall discuss this in detail in the next section of this chapter.

Definition 5. Λ for $\{x_1 \mid x_1 \neq x_1\}$

Theorem 4. $\vdash (x_1)(x_1 \notin \Lambda)$

PROOF. Same as the corresponding theorem in **F**, since Λ is stratified.

In the foregoing theorem and definition, we have used the usual abbreviations "$\neq$" and "$\notin$" for applying negation to formulas involving these predicate symbols.

Definition 6. $\{x\}$ for $\{y \mid y = x\}$ where x is any term and y is any variable not occurring in x.

Definition 7. $\{x, y\}$ for $\{z \mid z = x \vee z = y\}$ where z does not occur in x or y.

$\{x\}$ is stratified if x is, and $\{x, y\}$ is stratified if and only if there is some adequate assignment of types for which x and y have the same type.

Definition 8. $\{y_1, \ldots, y_n\}$ for $\{x \mid x = y_1 \vee, \ldots, \vee x = y_n\}$ where x is some variable not occurring in any of the y_i.

Such a finite set of n elements will be stratified if and only if there is an adequate assignment of types in which each element has the same type.

Definition 9. 0 for $\{\Lambda\}$

0 is stratified.

Definition 10. $(x \cup y)$ for $\{z \mid z \in x \vee z \in y\}$ where z does not occur in either of the terms x or y.

Definition 11. $(x \cap y)$ for $\{z \mid z \in x \wedge z \in y\}$ where z does not occur in either of the terms x or y.

Definition 12. $\bar{x}$ for $\{y \mid y \notin x\}$ where y does not occur in the term x.

$(x \cup y)$ and $(x \cap y)$ are stratified if and only if there is some adequate assignment of types in which x and y have the same type. $\bar{x}$ is stratified if x is and it has the same type as x under any adequate assignment of types.

The reader can begin to see one major drawback of **NF**. It is that stratification can be tested directly only when a wff or term is in an unabbreviated form. This is a tremendous practical hindrance in working within **NF**, for we depend heavily on definition to lighten an otherwise unwieldy symbolism in any set theory. We do not intend to retranslate into primitive symbolism in every instance, but we are obliged to accompany definitions with statements about stratification in order to know whether or not (or under what conditions) our axiom of abstraction is applicable to defined terms.

With these general definitions at hand, we can proceed with the Fregean definition of the natural numbers.

Definition 13. $S(x)$ for $\{z \mid (Ey)(y \in z \wedge z \cap \overline{\{y\}} \in x\}$ where y and z are distinct variables not occurring in the term x.

Exercise. Prove that $S(x)$ is stratified if x is, and that $S(x)$ has the same type as x under any adequate assignment of types.

Definition 14. N for $\{x_1 \mid (x_2)(0 \in x_2 \wedge (x_3)(x_3 \in x_2 \supset S(x_3) \in x_2) \supset x_1 \in x_2)\}$
N is stratified.

Definition 15. Fin for $\{x_1 \mid (Ex_2)(x_2 \in N \wedge x_1 \in x_2)\}$

Definition 16. Inf for $\overline{\text{Fin}}$
Fin is the set of all finite sets. (A set is finite if, for some n, it is an element of a natural number n regarded as a class of all n-element sets.) Inf is the set of all infinite sets. Both of these terms are stratified.

Theorem 5. $\vdash 0 \in N$

PROOF. The reader will prove Theorem 5 as an exercise. (*Hint:* same as corresponding theorem in **F**.)

Theorem 6. $\vdash (x_1)(x_1 \in N \supset S(x_1) \in N)$

PROOF. The reader will prove Theorem 6 as an exercise. (*Hint:* same as corresponding theorem in **F**.)

Theorem 7. $\vdash (x_1)(0 \neq S(x_1))$

PROOF. The reader will prove Theorem 7 as an exercise. (*Hint:* same as corresponding theorem in **F**.)

Theorem 8.

$$\vdash (x_1)(0 \in x_1 \wedge (x_2)(x_2 \in x_1 \wedge x_2 \in N \supset S(x_2) \in x_1) \supset N \subset x_1)$$

PROOF. The reader will prove Theorem 8 as an exercise.
The last is the set-theoretic form of the principle of mathematical induction. Our procedure in dealing with systems in which Frege's natural numbers are used is to obtain the metatheorem of mathematical induction from the set-theoretic form by use of the abstraction operator. What we obtain in **NF** is the following:

Theorem 9.

$$\vdash A(0) \wedge (x)(x \in N \wedge A(x) \supset A(S(x))) \supset (x)(x \in N \supset A(x))$$

where $A(x)$ is any *stratified* wff which contains x free. $A(0)$ and $A(S(x))$ result from $A(x)$ by the indicated substitutions.

PROOF. Immediate form Theorem 8 by $e\forall$.
Theorem 9 is a limited form of induction, since we can induct only on stratified conditions. In **ST** we could induct on any condition, but then

all conditions in **ST** are stratified. Nevertheless, we had no such limitation on mathematical induction in **ZF**; we could induct on all conditions expressible in that system. This limitation of **NF** will not be overcome, and is a permanent feature of the system.

The limitation resulting from our inability to induct on all formulas will be considerably lessened if we can deduce the remaining Peano postulate, i.e., $S(x) = S(y) \supset x = y$ where $x \in N$ and $y \in N$. We have already seen that the von Neumann natural numbers are not definable in **NF** because $x \cup \{x\}$ is unstratified for all terms x. Since we used the von Neumann natural numbers to prove the remaining Peano postulate in our system **F** of Chapter 3, it follows that we cannot use the same method to prove the theorem here. Let us review our line of reasoning in Chapter 3 in order to see the role played by the von Neumann natural numbers in proving the remaining Peano postulate.

Theorem 30 of Chapter 3,

$$\vdash (x_1)(x_2)(x_1 \in N \wedge x_2 \in N \wedge S(x_1) = S(x_2) \supset x_1 = x_2),$$

was an immediate consequence of Theorem 29,

$$\vdash (x_1)(x_2)(S(x_1) \subset S(x_2) \wedge x_1 \in N \wedge x_2 \in N \supset x_1 \subset x_2).$$

This theorem was proved by assuming the hypotheses and letting x_3 be some element of x_1. Since x_3 is an element of a natural number x_1, there is some a_1 not in x_3. (This last fact is the Corollary to Theorem 28. Theorem 28 asserts that the universal set V is not finite; that is, V is not an element of a natural number.) Since a_1 is not in x_3, $(x_3 \cup \{a_1\}) \cap \overline{\{a_1\}} = x_3$. But, as follows from the very definition of the successor, $x_3 \cup \{a_1\} \in S(x_1)$. By hypothesis this yields $x_3 \cup \{a_1\} \in S(x_2)$. Finally, again applying the definition of successor and Theorem 13 of Chapter 3, we obtain that $x_3 = (x_3 \cup \{a_1\}) \cap \overline{\{a_1\}}$ is an element of x_2. Since x_3 was any element of x_1, the theorem is proved.

The whole method of the proof of Theorem 29 hinges on the existence of an a_1 not in x_3, and this fact is turn depends on the nonfiniteness of V, as we have just indicated. The nonfiniteness of V, Theorem 28, was deduced from Theorem 26, which asserts $\vdash \Lambda \notin N$, i.e., that no natural number is empty. Finally, the fact that no natural number is empty was proved (through induction) by showing that each natural number had at least one element, namely, one of the von Neumann natural numbers.

The sole use of the von Neumann natural numbers in the foregoing line of reasoning is to show that no natural number is empty. From this fact it follows that V is infinite, and from this conclusion it follows that there is something that is not in any given natural number (the Corollary to Theorem 28, Chapter 3). Thus, if we can establish any of the succeeding links of this chain independently of the preceding links, we can complete the argument and prove the remaining Peano postulate without appeal to the von Neumann natural numbers. In particular, if we can prove in some independent manner that, for every natural number n, there is something that is not an element of n (Corollary to Theorem 28), then our remaining postulate obviously follows by reasoning identical to that of Theorem 29 and Theorem 30 of Chapter 3. In Section 7.3 of this chapter we show how to obtain the statement of the Corollary to Theorem 28 as a theorem of **NF**.

Notice, there is no obvious way of doing this. In each of our other systems such as **ST** and **ZF**, we have had to pose an axiom of infinity, though we were able, in **ZF**, to develop a theory of natural numbers without it. The question of whether or not a theorem of infinity is provable in **NF** was an open question for many years. However, in 1953, E. Specker proved that the axiom of choice does not hold in **NF**. But the axiom of choice does hold for finite sets. From this, it follows that there are infinite sets in **NF**. Thus, by this devious route the remaining Peano postulate is actually provable in **NF**. Notice that this shows the greater strength of **NF** over **ST**, since the remaining Peano postulate is not provable in **ST** without **ST**.3. In a later section we discuss the question of the axiom of choice and the theorem of infinity in more detail.

To facilitate this discussion, we need to develop the notion of cardinal number in **NF**. This we now proceed to do.

Definition 17. $\langle x, y\rangle$ for $\{\{x\}, \{x, y\}\}$ where x and y are any terms. This is our usual definition of the notion of ordered pair.

Theorem 10. $\vdash \langle x_1, x_2\rangle = \langle x_3, x_4\rangle \equiv x_1 = x_3 \wedge x_2 = x_4$

PROOF. The reader will prove Theorem 10 as an exercise.

The ordered pair $\langle x, y\rangle$ will be stratified only if there is some adequate assignment of types to x and y in which x and y have the same type.

Definition 18. R for $\{x_1 \mid (x_2)(x_2 \in x_1 \supset (Ex_3)(Ex_4)(x_2 = \langle x_3, x_4\rangle))\}$
R is the class of all relations. An element of R is a class of ordered pairs. R is stratified.

Definition 19. F for

$$\{x_1 \mid x_1 \in R \wedge (x_2)(x_3)(x_4)(\langle x_2, x_3\rangle \in x_1 \wedge \langle x_2, x_4\rangle \in x_1 \supset x_3 = x_4)\}$$

F is the class of all functional relations or functions. F is stratified.
In systems such as **ST** and **ZF**, the large classes such as R and F do not exist. However, such classes do exist in **NBG** or **MKM**.

Definition 20. F^1 for

$$\{x_1 \mid x_1 \in F \wedge (x_2)(x_3)(x_4)(\langle x_2, x_3\rangle \in x_1 \wedge \langle x_4, x_3\rangle \in x_1 \supset x_2 = x_4)\}$$

F^1 is the class of all 1–1 functions. F^1 is stratified.

Definition 21. $D(x)$ for $\{y \mid (Ez)(\langle y, z\rangle \in x)\}$ where x is any term and the distinct variables y and z do not occur in x.

This is the domain of the relational part of x.

Definition 22. $I(x)$ for $\{y \mid (Ez)(\langle z, y\rangle \in x)\}$ where x is any term, and the distinct variables y and z do not occur in x.

This is the range of a functional relation, the set of images of a given relation.

We now define the relation of cardinal similarity.

Definition 23. $x \operatorname{Sm} y$ for $(Ez)(z \in F^1 \wedge D(z) = x \wedge I(z) = y)$ where z does not occur in the terms x and y.

Theorem 11. $\vdash (x_1)(x_1 \operatorname{Sm} x_1)$

PROOF. The set $\{x_2 \mid (Ex_3)(\langle x_3, x_3\rangle = x_2 \wedge x_3 \in x_1)\}$ is clearly a 1–1 function whose domain and range is x_1.

Theorem 12. $\vdash (x_1)(x_2)(x_1 \operatorname{Sm} x_2 \supset x_2 \operatorname{Sm} x_1)$

PROOF. The reader will prove Theorem 12 as an exercise.

Theorem 13. $\vdash (x_1)(x_2)(x_3)(x_1 \operatorname{Sm} x_2 \wedge x_2 \operatorname{Sm} x_3 \supset x_1 \operatorname{Sm} x_3)$

PROOF. The reader will prove Theorem 13 as an exercise.

The last three theorems show that the relation of cardinal similarity is an equivalence relation on sets. This permits us to define the notion of the cardinal number of a set as the set of all similar sets. In Zermelo set theory, we introduced cardinals as certain ordinals. Of course we cannot construct the ordinals in **NF** as we did in **ZF** either. Instead we use the method of defining them as sets of ordinally similar sets, which is the intuitive conception of an ordinal number in intuitive set theory. In short, in **NF** we define cardinals and ordinals as the class of all classes of the given cardinal (ordinal) type, whereas in **ZF** we proceed by choosing a canonical representative from each cardinal (ordinal) class.

Definition 24. $\operatorname{Nc}(x)$ for $\{y \mid y \operatorname{Sm} x\}$ where y does not occur in x.

Definition 25. $(x + y)$ for

$$\{z \mid (Ew)(Er)(z = (w \cup r) \wedge w \in x \wedge r \in y \wedge w \cap r = \Lambda)\},$$

where the variables w, r, and z are all distinct and do not occur in the terms x and y.

Exercise. Prove that $(x + y)$ is stratified if and only if there is an adequate assignment of types to x and y in which they have the same type.

If x and y are cardinal numbers, then $x + y$ is the cardinal sum of x and y.

Definition 26. 1 for $S(0)$

Definition 27. $\mathrm{Pu}(x)$ for $\{y \mid (Ez)(y = \{z\} \wedge z \in x)\}$, where y and z are distinct variables that do not occur in the term x.

$\mathrm{Pu}(x)$ is the set of all unit (one-element) subsets of x. $\mathrm{Pu}(x)$ is stratified if x is. We use this notion to connect cardinal addition with our operations as already defined for the natural numbers.

Theorem 14. $\vdash 1 = \mathrm{Pu}(V)$

PROOF. The reader will prove Theorem 14 as an exercise.

Theorem 15. $\vdash (x_1)(S(x_1) = (x_1 + 1))$

PROOF. The reader will prove Theorem 15 as an exercise.

With these general notions now defined in **NF**, we are ready to proceed to a closer discussion of Cantor's theorem in **NF**.

7.2 Cantor's Theorem in NF

In intuitive set theory, Cantor's theorem asserts that no set is cardinally similar to its power set. In our language, this is expressed by: $(x_1)(\sim x_1 \text{ Sm } P(x_1))$. Now, by Theorem 3, we have $\vdash V = P(V)$ and by Theorem 11 we have that $\vdash (x_1)(x_1 \text{ Sm } x_1)$. Applying $e\forall$ to Theorem 11 we obtain $\vdash V \text{ Sm } V$. Using this together with Theorem 3 and **NF**.1, we immediately obtain $\vdash V \text{ Sm } P(V)$. If Cantor's theorem is provable, we can apply $e\forall$ and obtain $\sim V \text{ Sm } P(V)$, and we have a formal contradiction in **NF**. However, as we have indicated in previous discussion, the usual derivation of Cantor's theorem fails in **NF**. Let us see why.

In intuitive set theory, the argument for Cantor's theorem runs as follows: Let x be a set and assume that x is similar to its power set $\mathscr{P}(x)$. Let f be a function that gives the 1–1 correspondence between x and $\mathscr{P}(x)$. Now, for all $y \in x$, $\langle y, z\rangle \in f$, either $y \in z$ or $y \notin z$ holds. Let

$$w = \{y \mid y \in x \wedge (z)(\langle y, z\rangle \in f \supset y \notin z)\}.$$

Clearly, $w \subset x$, since every element of w is an element of x. Thus $w \in \mathscr{P}(x)$. Since f is a 1–1 function whose domain is x and whose range is $\mathscr{P}(x)$, there is some one element $a \in x$ such that $\langle a, w\rangle \in f$. Now, if $a \in w$, then, by the principle of abstraction, a satisfies the defining condition of w and $(z)(\langle a, z\rangle \in f \supset a \notin z)$. Applying $e\forall$ we obtain $\langle a, w\rangle \in f \supset a \notin w$. But $\langle a, w\rangle \in f$ and so, by MP, we obtain $a \notin w$. We have deduced that $a \in w$ implies $a \notin w$ which is equivalent to $a \notin w$. On the other hand, if $a \notin w$, then, since $\langle a, z\rangle \in f \supset z = w$ (f is 1–1) we have that

$$(z)(\langle a, z\rangle \in f \supset a \notin z),$$

and a satisfies the defined condition of w. By the abstraction principle, we then obtain $a \in w$. This contradicts $a \notin w$ and establishes the falsity of the assumption that x and $\mathscr{P}(x)$ are similar (that is, that there is a 1–1 correspondence between them).

The only thing preventing this line of reasoning in **NF** is the fact that the term w is not stratified. Remember that an ordered pair is stratified only if there is some adequate assignment of types in which both members of the pair are of the same type. Thus, any adequate assignment of types to w will require that y and z have the same type. But the wff $y \notin z$ is also part of w and this means that the type of z must be one higher than the type of y in any adequate assignment of types to w. Since y and z cannot have both the same type and different types, no adequate assignment of types to w is possible. This prevents the application of the principle of abstraction and thus avoids the direct deduction of the Cantor theorem (and thus the Cantor paradox) in **NF**.

Notice, however, that if we could define w as

$$\{y \mid y \in x \wedge (z)(\langle\{y\}, z\rangle \in f \wedge y \notin z)\},$$

then we would obtain a stratified condition. Taking the unit set $\{y\}$ as the first member of the pair means that $\{y\}$ and z must have the same type, which will happen precisely when the type of z is one higher than that of y. We obtain a stratified condition, but, in this case, the domain of f is no longer x but rather $\mathrm{Pu}(x)$, the set of all unit (one-element) subsets of x. We can thus prove:

Theorem 16. $\vdash (x_1)(\sim \mathrm{Pu}(x_1) \mathrm{\ Sm\ } P(x_1))$

PROOF. Assume $\mathrm{Pu}(x_1) \mathrm{\ Sm\ } P(x_1)$. This means that there is some 1–1 function, let us call it a_1, $a_1 \in F^1$, such that $D(a_1) = \mathrm{Pu}(x_1)$ and

$$I(a_1) = P(x_1).$$

We let w stand for $\{x_2 \mid x_2 \in x_1 \wedge (x_3)(\langle\{x_2\}, x_3\rangle \in a_1 \supset x_2 \notin x_3))\}$. The term w is stratified, and so by **NF**.2 we have

$$\vdash x_2 \in w \equiv x_2 \in x_1 \wedge (x_3)(\langle\{x_2\}, x_3\rangle \in a_1 \supset x_2 \notin x_3).$$

We easily obtain $(x_2)(x_2 \in w \supset x_2 \in x_1)$ and so $w \subset x_1$. Thus

$$w \in P(x_1) = I(a_1).$$

There is an element of $D(a_1)$, let us call it a_2, such that $\langle a_2, w\rangle \in a_1$. But $a_2 \in \mathrm{Pu}(x_1)$, and so there is an element of x_1, let us call it a_3, such that $a_2 = \{a_3\}$. Thus, $\langle\{a_3\}, w\rangle \in a_1$.

If we assume $a_3 \in w$, then, by applying **NF**.2 and MP, we obtain $(x_3)(\langle\{a_3\}, x_3\rangle \in a_1 \supset a_3 \notin x_3)$. Applying $e\forall$ we obtain

$$\langle\{a_3\}, w\rangle \in a_1 \supset a_3 \notin w.$$

Since $\langle\{a_3\}, w\rangle \in a_1$ holds from our foregoing hypotheses, we have $a_3 \notin w$ by MP. Thus, $a_3 \in w$ implies $a_3 \notin w$ which yields $a_3 \notin w$.

On the other hand, $a_1 \in F^1$ means that $(x_3)(\langle\{a_3\}, x_3\rangle \in a_1 \supset x_3 = w)$ holds. Thus, $(x_3)(\langle\{a_3\}, x_3\rangle \in a_1 \supset a_3 \notin x_3)$ since $a_3 \notin w$ holds. Since $a_3 \in x_1$ also holds, a_3 satisfies the defining condition of w and so $a_3 \in w$ by **NF**.2. This contradiction establishes the falsity of the assumption that

$$\mathrm{Pu}(x_1)\ \mathrm{Sm}\ P(x_1),$$

and so our theorem is established.

Theorem 16 establishes that there is no 1–1 correspondence between the one-element subsets of a given set and the set of all of its subsets. Since $\mathrm{Pu}(x)$ is a subset of $P(x)$, it easily follows that $\mathrm{Nc}(P(x))$ is greater than $\mathrm{Nc}(\mathrm{Pu}(x))$ for any set x; that is, it easily follows when the usual way of defining the "greater than" relation between cardinal numbers is introduced in the system. ($\mathrm{Nc}(x)$ is greater than $\mathrm{Nc}(y)$ will mean that y is similar to a subset of x but not conversely.)

However, Theorem 16 would seem to give rise to the possibility of a paradox in the following manner: We know that $\mathrm{Pu}(x)$ and $P(x)$ are not similar. But there is an obvious similarity between any set x and the set $\mathrm{Pu}(x)$. Just let each element $y \in x$ correspond to the set $\{y\}$. Since $\mathrm{Pu}(x)$ is simply the set of all such unit subsets of x, we have a 1–1 correspondence f between any two sets x and $\mathrm{Pu}(x)$. We obtain $\mathrm{Pu}(V)\ \mathrm{Sm}\ V$ and $V\ \mathrm{Sm}\ P(V)$, yielding $\mathrm{Pu}(V)\ \mathrm{Sm}\ P(V)$ which contradicts Theorem 16. This intuitively obvious reasoning, valid in intuitive set theory, is blocked by the fact that our "obvious" 1–1 correspondence f between x and $\mathrm{Pu}(x)$ is not stratified. Formally, f would be defined as $\{z \mid (Ey)(y \in x \wedge \langle y, \{y\}\rangle = z)\}$. But f is not stratified, since the ordered pair $\langle y, \{y\}\rangle$ is not stratified. In any adequate assignment of types, y and $\{y\}$, the two components of the pair, must have the same type; a clear impossibility. It is thus impossible to prove in this manner that a set x is similar to its set of unit subsets $\mathrm{Pu}(x)$ in **NF**.

Of course, the impossibility of proving in this manner the similarity of x and $\mathrm{Pu}(x)$ does not, in itself, yield that there are actually sets that are not similar to their set of unit subsets. Yet, we can establish just this in **NF**. We first define:

Definition 28. $\mathrm{Can}(x)$ for $x \,\mathrm{Sm}\, \mathrm{Pu}(x)$ where x is any term.

A set x is "Cantorian" if it is similar to its set of unit subsets. We now prove that there are non-Cantorian sets in **NF**, in particular V.

Theorem 17. $\vdash \sim \mathrm{Can}(V)$

PROOF. Suppose $\mathrm{Can}(V)$, that is, $V \,\mathrm{Sm}\, \mathrm{Pu}(V)$. By Theorem 3, $V = P(V)$, and so $\vdash V \,\mathrm{Sm}\, P(V)$ by **NF**.1. By the transitivity of Sm, we deduce $\mathrm{Pu}(V) \,\mathrm{Sm}\, P(V)$ which contradicts Theorem 16. Thus, our assumption is false and $\vdash \sim \mathrm{Can}(V)$.

We now have before us the paradoxical conclusion that V is a set with the same cardinality as its power set (the set of *all* its subsets). Yet V does not have the same cardinality as the set of all of its one-element subsets $\mathrm{Pu}(V)$. The cardinality of $\mathrm{Pu}(V)$ is less than that of V, since the cardinality of $\mathrm{Pu}(V)$ is less than that of $P(V) = V$. It is only the fact that stratification fails for ordered pairs of the form $\langle y, \{y\}\rangle$ that prevents a direct proof of contradiction in **NF**.

Of course, we would expect some type of paradoxical result to be forthcoming with the existence of a universal set V, as well as the possibility of proving Theorem 16. We finally obtain such a paradox in the form of the existence of non-Cantorian sets. We call such sets "non-Cantorian", since each intuitively conceived Cantorian set can certainly be put into 1–1 correspondence with its set of unit subsets, and Cantor would have surely preferred to dispense with the universal set rather than discard the classic form of the theorem bearing his name. This, of course, is what **ZF** does.

The existence of non-Cantorian sets in **NF** shows in a new way why finding a model for **NF** is so difficult. A model would have to be some set-theoretic structure, and one that contained non-Cantorian sets. Of course, such a model might be "nonstandard". It might be some abstract structure in which the elements were not sets in the usual sense at all. The recent result of Dana Scott, in which nonstandard models for **ZF** in which the continuum hypothesis fails are exhibited, shows that such a model of **NF** might exist. One might ultimately reject **NF** as a foundation for purely practical reasons, e.g., because the system is too cumbersome. Nevertheless, the resolution of the question of a model for a system with such unusual properties is of genuine interest.

We now turn to the problem of the axiom of choice in **NF**.

7.3 The Axiom of Choice in NF and the Theorem of Infinity

In Chapter 6, we discussed the fact that the axiom of choice is independent with respect to the other axioms of **ZF**. In particular, then, the axiom of choice can be consistently added as a hypothesis to **ZF** (as we added it in Chapter 5). Quine (see Quine [3], p. 164), Rosser (see Rosser [2], p. 512 and p. 517), and other logicians had generally supposed that the axiom of choice was probably independent in **NF** (or at least "as independent" as it is in **ZF**, since the independence with respect to **ZF** had not yet been proved). However, in 1953, E. P. Specker published a proof that the axiom of choice is contradictory if added to **NF**. Specker's proof served to heighten concern for the possibility of inconsistency in **NF**, especially since Gödel had proved in 1940 that the axiom of choice is consistent with respect to the other axioms of **ZF**.

If we reflect on the meaning of the axiom of choice and the structure of **NF** as we have explored it, we can begin to see why indeed the axiom of choice might fail in **NF**. For a disjoint collection of nonempty sets, the axiom of choice allows us to pick a choice set having exactly one element from each set in the collection. Now clearly, in intuitive set theory, such a choice set is in 1–1 correspondence with the disjoint collection, since we have only to allow each element of the choice set to correspond to the set of the collection from which it is chosen. However, for any nonempty set x, $\mathrm{Pu}(x)$ is a disjoint collection of nonempty sets, and clearly the choice set of $\mathrm{Pu}(x)$ is just x itself. Yet, there are sets in **NF** for which x and $\mathrm{Pu}(x)$ are not similar, as we have just shown. Again, it is the non-Cantorian sets that deviate from our intuitive conception.

Of course, our argument in the foregoing paragraph is only heuristic, and does not itself constitute a disproof of the axiom of choice. This is so because once again our construction of the "obvious" 1–1 correspondence between the choice set and the original disjoint collection involves an unstratified definition. Specker's proof involves the axiom of choice in the form which states that the cardinal numbers are well ordered (see our statement of different forms of the axiom of choice in Chapter 5). Let us get a closer look at what is involved.

We define the relation $x \leq y$ between sets by "there is a 1–1 correspondence between x and a subset of y". Formally, we have:

Definition 29. $x \leq y$ for $(Ez)(z \subset y \wedge z \mathrm{\ Sm\ } x)$ where z does not occur in x or y.

We have already proved that $x \leq x$, since we know that $\vdash x \mathrm{\ Sm\ } x$. We obviously have transitivity as well. Another property of the relation $\leq$ is antisymmetry: $x \leq y \wedge y \leq x \supset x \mathrm{\ Sm\ } y$. This is the famous theorem of Schröder-Bernstein, which states that if x is similar to a subset of y and y is similar to a subset of x, then $x \mathrm{\ Sm\ } y$. This theorem can be proved without aid of the axiom of choice (see Rosser [2], p. 353).

We can now define the relation of cardinal dominance between sets:

Definition 30. $x < y$ for $x \leq y \wedge \sim x \mathrm{\ Sm\ } y$ where x and y are any terms.

From the foregoing properties of $\leq$, we have immediately that $<$ is a partial order (see our discussion in Chapter 5, Section 5.1). That is, $\vdash(x_1)(\sim(x_1 < x_1))$ and $\vdash(x_1)(x_2)(x_1 < x_2 \supset \sim(x_2 < x_1))$ hold. Furthermore, the transitivity of $<$ holds. Given any such partial order, we obtain a total order if we also have the property of connectivity, that is

$$(x_1)(x_2)(x_1 < x_2 \vee x_2 < x_1 \vee x_1 \text{ Sm } x_2).$$

This says that, for any two sets, one must be similar to a subset of the other. This can be proved only with the aid of the axiom of choice. In conjunction with other facts of the ordinal numbers, it is equivalent to the axiom of choice.

The addition of the property of connectivity guarantees that the relation "$<$" is a total ordering on sets. A cardinal number is a set of all the sets of a given similarity type, and so cardinal numbers are ordered by comparing their representative elements. That is, given two cardinal numbers $\mathfrak{a}$ and $\mathfrak{b}$, we take $x \in \mathfrak{a}$, $y \in \mathfrak{b}$. The cardinals are ordered as are the sets x and y. This is obviously independent of the choice of the representatives x and y.

To see that this total ordering on cardinals is, in fact, a well ordering, recall that the axiom of choice is equivalent to the assertion that any set can be well ordered. This means that, for any given cardinal $\mathfrak{a}$, every element of $\mathfrak{a}$ can be put into 1–1 correspondence with some representative of an ordinal number α. An ordinal number in **NF**, we recall, is the set of all well-ordered sets of a given order type. This means that, for every cardinal number $\mathfrak{a}$, there is at least one ordinal α whose elements are similar to the elements of $\mathfrak{a}$. The class of all ordinals whose elements are similar to the elements of a given cardinal is thus nonempty, and therefore it has a least element in the ordering of the ordinals. In this way, we associate with each cardinal $\mathfrak{a}$ the smallest ordinal α (smallest in the ordering of the ordinals) whose elements are similar to the elements of $\mathfrak{a}$. Of course, the elements of any ordinal β smaller than α will have cardinality less than $\mathfrak{a}$, since α is the smallest ordinal whose elements have the cardinal type $\mathfrak{a}$. Finally, given a nonempty set of cardinal numbers, we associate with each cardinal in the set its unique smallest ordinal as previously defined. The corresponding set of ordinals is well ordered and has a smallest element, and the cardinal associated with this ordinal will be smallest in the ordering of the cardinals.

Thus, the axiom of choice yields a well ordering of the cardinal numbers. If we can show that the cardinals are not well ordered in **NF**, it will follow that the axiom of choice fails in **NF**. Specker succeeds in proving that the cardinals are not well ordered in **NF**. We shall not include the details of Specker's proof. The interested reader can consult Specker [1] and also the discussion in Quine [4], p. 294–295.

Now, once we have established that the cardinals are not well ordered, we can observe that the set of all finite cardinals is well ordered.

A cardinal in **NF** is finite or infinite according to whether the sets which make up the cardinal are finite or infinite. By Definition 15, a set is finite precisely if it belongs to some natural number. No cardinal is empty, since $x \in \mathrm{Nc}(x)$ for all x. A finite cardinal is thus a nonempty natural number. It is not too difficult to prove that the finite cardinals are well ordered. Since the set of all cardinals is not well ordered, it follows that there is an infinite cardinal. The elements of such an infinite cardinal cannot be finite sets.

Thus, let x by any infinite cardinal and y any element of x. Since y is infinite and does not belong to any natural number, we obtain the following theorem: For any natural number n, there is a set (namely y) that is not an element of n. Formally stated, we have

$$\vdash (x_1)(x_2)(x_1 \in x_2 \wedge x_2 \in N \supset (Ex_3)(x_3 \notin x_1).$$

But this theorem is precisely the corollary to Theorem 28 of Chapter 3, and we have already seen in our previous discussion that this is all that is needed to prove the remaining Peano postulate in **NF**. We thus obtain the last remaining Peano postulate in **NF** as a corollary of the failure of the axiom of choice!†

Any of the foregoing derivative results of the disproof of the axiom of choice can be considered an axiom of infinity in **NF**. The existence of an infinite cardinal, or just the proof of the remaining Peano postulate will imply that V is infinite; that is, V belongs to no natural number.

In Rosser [2], which is based on Quine's **NF**, the remaining Peano postulate was assumed as an additional axiom, since the foregoing devious proof of a theorem of infinity in **NF** was not known at the time. The foregoing proof of the theorem of infinity clearly shows how failure of stratification of certain terms in **NF** does not exclude the possibility of deducing the desired result by other means. The von Neumann natural numbers, used in **F** to prove the last Peano postulate, were not stratified in **NF**, thus blocking our direct method of proof in **NF**. Nevertheless, the remaining postulate is provable in **NF**. In the same way, we cannot be certain that Cantor's paradox is not deducible in **NF** simply because the usual proof of Cantor's theorem is blocked by an unstratified term.

In **ST** we encountered essentially the same difficulty as that in **NF**, i.e., that the von Neumann natural numbers are not definable in type theory. But in **ST** we knew that we had to pose an axiom of infinity since we could exhibit a model of **ST** in a finite simple type hierarchy. By a model-theoretic analysis of **ST** we could see that our axioms did not necessitate an infinite model. Since no model for **NF** is known, the

† The author would like to thank Mr. Nicholas Goodman for making available the details of his formal proof of the axiom of infinity in **NF** based on Specker's disproof of the axiom of choice. It should be noted that Specker [1] notes that the axiom of infinity is provable in **NF** as a corollary of the failure of the axiom of choice.

possibility that a theorem of infinity could be proved was, until its proof was discovered, an open question.

Returning for a moment to the axiom of choice in **NF**, we observe that the generalized continuum hypothesis also fails in **NF**, since the axiom of choice is a consequence of the generalized continuum hypothesis. The generalized continuum hypothesis is consistent with **ZF**, and so we have another deviation of **NF** from **ZF**. Since mathematicians sometimes like to assume the continuum hypothesis in order to simplify certain questions of cardinality, and since they very often assume the axiom of choice, it would seem that we could exclude **NF** as a foundation simply by virtue of the fact that these two useful assumptions, consistent with **ZF** and other set theories, are not consistent with **NF**.

Again there is a possibility of responding to this criticism of **NF**. Clearly the non-Cantorian sets are the cause of the failure of the axiom of choice. Thus, we can still apply the axiom of choice to Cantorian sets in **NF**. Since Cantorian sets are the only ones that mathematicians ever work with anyway, the non-Cantorian ones being an oddity of **NF**, no known or useful application of the axiom of choice is excluded by restricting it to Cantorian sets. Of course, we have no proof that such a restricted use of the axiom of choice is consistent with **NF**. But there is no reason to suppose that this restricted use of the axiom leads to contradiction when added to **NF** if **NF** is consistent to begin with.

These considerations seem to show that we cannot reject **NF** as a foundation simply because the axiom of choice generally fails, for we can save the usual applications of the axiom by restricting the sets involved to Cantorian sets. Of course, the constant necessity of introducing this supplementary condition is irritating. Clearly, though, the single most important question for **NF** is consistency. In another important paper, Specker has cast some light on this question. This paper furnishes the basis for much of our discussion in the following section.

7.4 NF and ST; Typical Ambiguity

We have already noted, while defining the notion of stratification, that stratification and typical ambiguity are related. Indeed, stratification is one way to deal with typical ambiguity. If we restrict the wffs of **NF** to stratified formulas, we obtain a system much like **ST**. Yet, it is not immediately clear just how strong such a system is, since typical ambiguity with respect to **ST** is more of a relaxation of formal rigor than a precise generalization of the system. In this section, we show how a rigorous and careful analysis of the notion of typical ambiguity leads to a clarification of the precise relationship between **NF** and **ST**. Our ideas follow closely the paper Specker [2].

We consider now the system **ST**, but without the axiom of infinity **ST**.3. Our first meaning of typical ambiguity is contained in the following proof-theoretic observation about **ST**, without infinity: given any wff A of **ST** without **ST**.3, let A^+ be the wff obtained from A by raising the type superscript on every term of A by exactly one. We call A^+ the *type lift* of A. Then a wff of **ST**, without axiom of infinity, is provable only if its type lift is provable. That is, if $\vdash A$ then $\vdash A^+$. To see that this is true, we observe that the type lift of any axiom **ST**.1 or **ST**.2 is an axiom of the same kind. Furthermore, the operation of type lifting preserves our rules of inference. Thus, a proof of A can be immediately translated into a proof of A^+.

Exercise. Prove rigorously, by mathematical induction on the length of the proof of A, the foregoing metatheorem (if $\vdash A$, then $\vdash A^+$) for **ST** without **ST**.3.

For the remainder of this section, **ST** *will always mean* **ST** *without* **ST**.3.

We now have at least one precise notion of typical ambiguity in **ST**. It is that provability is preserved by type lift. It is only natural to ask whether or not provability is also preserved by a lowering of types. That is, is it true that $\vdash A^+$ only if $\vdash A$? The answer is no. Without the axiom of infinity, any finite simplified type hierarchy is a model for **ST**. Thus, the domain of individuals T_0 may have only one element. In that case, $T_1 = \mathscr{P}(T_0)$ has two elements, $T_2 = \mathscr{P}(T_1)$ has four elements, and, generally speaking, T_n has 2^n elements. Moreover, we can prove within **ST** that there are at least 2^n different sets of type n. Let us see this briefly.

We define V^1 as $\{x_1^0 \mid x_1^0 = x_1^0\}$, and Λ^1 as $\{x_1^0 \mid x_1^0 \neq x_1^0\}$. We have, as usual, the theorems $\vdash(\forall x_1^0)(x_1^0 \in V^1)$ and $\vdash(\forall x_1^0)(x_1^0 \notin \Lambda^1)$. From the first of these we have $\vdash(Ex_1^0)(x_1^0 \in V^1)$ and from the second we obtain

$$\vdash \sim (Ex_1^0)(x_1^0 \in \Lambda^1).$$

Thus, by the axiom of extensionality, we easily prove $\vdash V^1 \neq \Lambda^1$, and thereby $\vdash(Ex_1^1)(Ex_2^1)(x_1^1 \neq x_2^1)$. This proves that there are at least $2^1 = 2$ things of type 1. We can now iterate this process, using our usual definitions of union, intersection, and the like. We can form the terms $\{V^1\}$, $\{V^1, \Lambda^1\}$, $\{\Lambda^1\}$, Λ^2. These sets will all be different terms of type 2, and so we can prove that there are at least $2^2 = 4$ different sets of type 2. We leave as an exercise to the reader the completion of our inductive argument.

Now, the wff $(Ex_1^1)(Ex_2^1)(x_1^1 \neq x_2^1)$ is the type lift of $(Ex_1^0)(Ex_2^0)(x_1^0 \neq x_2^0)$. If our converse rule is to hold, then this latter formula must be a theorem since the first one is. But it is easy to see that this latter formula cannot be a theorem, since we have a model for **ST** in which there is only one individual in T_0. There is thus a model in which $(Ex_1^0)(Ex_2^0)(x_1^0 \neq x_2^0)$ is false, and so it cannot be a theorem.

Suppose we now consider the theory obtained by adding to **ST** the general rule $\vdash A^+$ only if $\vdash A$.† Our argument from the preceding paragraph shows that we obtain a strictly stronger theory. In fact, finite type hierarchies will be excluded as models with the addition of such a rule. To see this, suppose that a simple type hierarchy with T_0 of finite cardinality n is a model of **ST** with our added rule. Now, as we have seen, we can prove in **ST** that there are 2^n different sets of type n. More particularly, we can prove

$$(Ex_1^n)(Ex_2^n)\cdots(Ex_{2^n}^n)(x_1^n \neq x_2^n \wedge \cdots x_1^n \neq x_{2^n}^n \wedge x_2^n \neq x_3^n \wedge \cdots \wedge \cdots \wedge x_{2^n-1}^n \neq x_{2^n}^n).$$

Call this formula A^n. A^n involves only variables of type n, and it asserts the existence of 2^n different sets of type n. With our added rule, we can therefore prove $\vdash A^{n-1}$, obtained by lowering the types in A^n by one, and this asserts the existence of 2^n sets of type n–1. Proceeding inductively, that is, by iterated application of our new rule, we can prove $\vdash A^0$, which is obtained by n iterations of our rule starting with $\vdash A^n$, and which asserts the existence of $2^n > n$ objects of type 0. Since A^0 is provable if we use our added rule, it is true in every type hierarchy that is a model of **ST** with our added rule. Thus, the assumption that T_0 has only n elements is false. Consequently, no finite hierarchy is a model for **ST** with the added rule.

We speak of **ST** with our added rule as **ST** *with typical ambiguity*. Our added rule of typical ambiguity is an axiom of infinity as our foregoing argument has just shown. We can give a (nonconstructive) model-theoretic proof of the consistency of **ST** with typical ambiguity relative to **ST**.

Let X be the set of all wffs that are theorems of **ST** with typical ambiguity according to *one application* of our new rule. Precisely, a wff A is in X if and only if A^+ is a theorem of **ST**. We wish to show that every finite subset of X is consistent relative to **ST**. Since only a finite number of wffs can be used in any proof, only a finite number of the new theorems added by one application of our rule can appear in any proof. It follows that the whole set X will be consistent relative to **ST**. Let $A_1, \ldots, A_m$ be m wffs of X. This finite set of wffs has a model if and only if the conjunction $(A_1 \wedge A_2 \wedge \cdots \wedge A_m)$ has a model. Now, $(A_1 \wedge A_2 \wedge \cdots \wedge A_m)^+$ is the same as $(A_1^+ \wedge A_2^+ \wedge \cdots \wedge A_m^+)$. Since each A_i in our finite set is in X, the wff A_i^+ is a theorem of **ST** for each i. That is, $\vdash_{\mathbf{ST}} A_i^+$ for $1 \leq i \leq m$. Thus, $\vdash_{\mathbf{ST}}(A_1^+ \wedge A_2^+ \wedge \cdots \wedge A_m^+)$, and thus $\vdash_{\mathbf{ST}}(A_1 \wedge A_2 \wedge \cdots \wedge A_m)^+$. The type lift of the conjunction of our wffs is a theorem of **ST**, and is thus true in every model of **ST**.

Since we have excluded **ST**.3 from **ST** in this discussion, any simple type hierarchy $T = \{T_0, T_1, \ldots\}$ will be a model for **ST**. Given T, let

† For any given wff of **ST**, let A^n represent the wff $A^{++\cdots+}$ (to n occurrences of $+$) obtained from A by exactly n applications of type lift. Then our new rule can be stated: If $\vdash_{\mathbf{ST}} A^n$, then $\vdash A$.

$U = \{U_0, U_1, \ldots\}$ be the denumerable collection defined by the condition that $U_i = T_{i+1}$ for all natural numbers i. Thus, $U_0 = T_1$, $U_1 = T_2$, and so on. U is also a simple type hierarchy, for the relationship between the sets T_i is given by $T_{i+1} = \mathcal{P}(T_i)$, and this relationship obviously carries over to U. Thus, U is also a model for **ST**. Now, the wff

$$(A_1 \wedge A_2 \wedge \cdots \wedge A_m),$$

because of its formal relation to its type lift $(A_1 \wedge A_2 \wedge \cdots \wedge A_m)^+$, has exactly the same truth set in U as its type lift has in T. Since the type lift is true in T (because it is a theorem and thus true in every model), the wff $(A_1 \wedge A_2 \wedge \cdots \wedge A_m)$ must be true in U. Consequently, this wff is consistent, since it holds in some model U of **ST**. The finite set $A_1, \ldots, A_m$ thus has a model, and is therefore consistent. Since this is an arbitrary finite subset of X, the whole set X is consistent and no contradiction can ensue by a single application of our new rule.

To complete the demonstration, it suffices to observe that the proof of any wff A as a theorem of **ST** with typical ambiguity can involve only a finite number of applications of our new rule. As we observed in the footnote on page 255, we can state our new rule in the form: If $A^{++\cdots+}$, to any finite number n of type lifts, *is a theorem of* **ST**, then A is a theorem in our enlarged system. An inductive application of the foregoing proof yields that **ST** with typical ambiguity is consistent relative to **ST**.

The consistency of **ST** with typical ambiguity raises the question of whether we obtain a consistent system if we add to **ST** the stronger principle $\vdash A \equiv A^+$ for any closed wff A of **ST**. We call the system obtained from **ST** by adding this stronger principle, **ST** *with complete typical ambiguity*. **ST** with complete typical ambiguity obviously contains **ST** with typical ambiguity as a subsystem.

ST with typical ambiguity roughly corresponds to the system obtained by restricting the wffs of **NF** to stratified formulas. Clearly we have typical ambiguity in such a system. But a more thorough model-theoretic analysis is necessary to show that **ST** with typical ambiguity is essentially the same as **NF** restricted to stratified formulas. What Specker has shown is the more interesting result that **ST** with *complete* typical ambiguity is essentially equivalent to **NF** itself. More precisely, Specker proves that there is a model of **ST** with complete typical ambiguity if and only if there is a model for **NF**. Since we have already seen that **NF** has certain anomalies, this result shows that the added principle $\vdash A \equiv A^+$, A is a sentence, is very strong indeed. Intuitively, it also seems unnatural that **ST** with complete typical ambiguity is consistent, since our added principle says that everything true is one type must be true in any other type. This generally is not true of our simplified type hierarchies where each set T_{n+1} is the power set of T_n. It may be that no model T of **ST** with complete typical ambiguity will be a simplified type hierarchy. T may be some nonstandard model of **ST** in which the sets T_i are related in a more "unnatural" way than the relatively simple relation of power set of a simplified type hierarchy. Yet, we have no proof that **ST** with complete typical ambiguity is contradictory. Specker's result throws light on the question of the consistency of **NF** by showing the exact relation of **NF** to **ST**, but the question of the consistency of **NF** remains open.

We close this section by briefly presenting the idea of Specker's proof that **ST** with complete typical ambiguity is consistent if and only if **NF** is. For this purpose, we suppose, for the remainder of this section, that **NF** and **ST** are formalized without the use of a primitive term operator of abstraction. This makes it easier to visualize models of the systems, since the interpretation of the axiom of abstraction in any model is now determined by our assignment of a relation to the predicate letter "$\in$" without further specification of how to interpret the abstraction operator. We still refer to these systems as "**NF**" and "**ST**" rather than adding the asterisk to write "**NF***" and "**ST***". Moreover, since we speak not just of simplified type hierarchies, but of models of **ST** in a general sense, we should also recall the fact that we have a first-order formulation of **ST**. At this point it might be helpful for the reader to review briefly the discussion of these points in the last sections of Chapter 4.

We remarked informally that our added principle $\vdash A \equiv A^+$, A a sentence, amounts to saying that everything true in one type is true in another. Precisely, let $T = \{T_i\}$, $i \in N$, be any model of **ST** with complete typical ambiguity, and let U be the structure defined by $U_i = T_{i+1}$. (We naturally suppose that the interpretation of "$\in$" remains unchanged; that is, $a \in_U b$ holds if and only if $a \in_T b$ holds when a and b are any two elements of the domain of the interpretation U. The original model T may, of course, be any model, and not necessarily a standard one.) Then U_i is also a model of **ST** with complete typical ambiguity. In fact, U and T are elementarily equivalent as models of **ST**. To see this, let A be any sentence of our system **ST** with complete typical ambiguity. Because of our added rule, A is true in T if and only if A^+ is true in T. But A is true in U if and only if A^+ is true in T as follows from the formal relationship between A and A^+ and from the relationship between our two structures. Thus A is true in T if and only if A is true in U, and so the two structures are elementarily equivalent. Since T is a model of **ST** with complete typical ambiguity and consequently satisfies all the theorems of that system, U is also a model.

Conversely, let T be any model of **ST** which is elementarily equivalent to the structure U defined by $U_i = T_{i+1}$. Then T is a model of **ST** with complete typical ambiguity. This follows immediately from the fact that all biconditionals of the form $A \equiv A^+$, A a sentence, will hold in such a model. This, in turn, follows from the fact that A is true in U if and only if A^+ is true in T, and A is true in U if and only if A is true in T, since U and T are assumed to be elementarily equivalent.

Suppose now that we have a model T of **ST** that is not only elementarily equivalent to U as just defined, but isomorphic to it. This would mean that there is a 1–1 mapping f from the union of all the sets T_i to the union of all the sets $U_i = T_{i+1}$ that satisfies the following conditions: For all x, x is an element of T_i if and only if $f(x)$ is an element of $U_i = T_{i+1}$, and $x\ E\ y$ holds if and only if $f(x)\ E f(y)$ holds, where E is the relation that interprets the predicate letter "$\in$" of **ST** in the model T. We say informally that "there is an isomorphism between T_i and T_{i+1} for all i." Now any model T admitting an isomorphism between T_i and T_{i+1} for all i is obviously elementarily equivalent to U as previously, for the notion of isomorphism is strictly stronger than the notion of elementary equivalence. Thus, any model T admitting such an isomorphism between T_i and T_{i+1} for all i is a model for **ST** with complete typical ambiguity as follows from our previous results. What Specker succeeds in proving is that if **ST** with complete typical ambiguity is consistent, then there is some model T of **ST** admitting an isomorphism between T_i and T_{i+1} for all i (see Specker [2], p. 119). In general, elementarily equivalent models of a given theory will not be isomorphic, and so the proof of this result depends essentially on the particular structure of **ST** with complete typical ambiguity. The fact that the existence of a model for **ST** with complete typical ambiguity implies the existence of a model T with T_i isomorphic to T_{i+1} for all i shows just how strong the added rule $\vdash A \equiv A^+$, A a sentence, really is.

We can now use this basic theorem to show that there is a model of **ST** with complete typical ambiguity if and only if there is a model of **NF**. Let T be a model of **ST** with complete typical ambiguity, and let T admit an isomorphism f from T_i onto T_{i+1}. We construct a model $\langle D, g\rangle$ for **NF** by letting the domain of the model be the set T_0 of individuals. g assigns to the predicate letter "$\in$" of **NF** the set E^* of all ordered pairs $\langle a, b\rangle$, such that $\langle a, b\rangle$ is in E^* if and only if $\langle a, f(b)\rangle$ is in the relation E, which is the interpretation of "$\in$" of **ST** in the model T. We have an axiom of extensionality and an axiom of abstraction in both theories, and the nature of our isomorphism f assures us that $\langle D, g\rangle$ will be a model for **NF**.

Given now a model $\langle D, g\rangle$ for **NF**, where E^* is the relation assigned to "$\in$" of **NF** by g, we construct a model for **ST** with complete typical ambiguity. The domain of our model for **ST** with complete typical ambiguity will be the set of all ordered pairs $\langle a, n\rangle$ where a is in D and n is a natural number, i.e., the set $D \times N$. For each n, the set T_n of objects of type n will be the ordered pairs $\langle a, n\rangle$ whose second component is n. The relation E which is the interpretation of "$\in$" in **ST** with complete typical ambiguity is defined by $\langle a, n\rangle\ E\ \langle b, m\rangle$ if and only if $m = n + 1$ and $\langle a, b\rangle$ is in the relation E^*. From the axioms of extensionality and abstraction of **NF** we easily obtain that this structure T is a model for **ST**. Furthermore, the mapping f defined by $f(\langle a, n\rangle) = \langle a, n + 1\rangle$, for $\langle a, n\rangle$ an element of T_n, is a 1–1 mapping from T_n onto T_{n+1}. It is easy to check that f is an isomorphism, and so it follows that our model is a model for

ST with complete ambiguity. Thus, **NF** is consistent if and only if **ST** with complete typical ambiguity is.

Exercise. Write a short essay entitled "**NF** as a foundation for mathematics." Bring in carefully each of the positive and negative points raised in this chapter. What is the crucial open problem for **NF**?

7.5 Quine's System ML

In 1940, Quine published his work *Mathematical Logic* in which he exponded a system **ML** which is related to **NF** in much the way **NBG** and **ZF** are related. That is, Quine considers the addition of proper classes to **NF**, just as we add proper classes to **ZF** to obtain **NBG**. All the sets of **NF** are classes, but there are classes of **ML** that cannot be the element of anything else. These are proper classes. Actually, the relation between **ML** and **NF** is more like the relation between **MKM** and **ZF** than that between **NBG** and **ZF**, since Quine does not restrict his axiom of abstraction to wffs predicative over the sets of **NF**.

Quine's original version of **ML** was proved inconsistent by R. Lyndon and J. B. Rosser (independently), but the inconsistency was basically attributable to a fault in exposition as Wang later showed. Wang corrected the fault and proved that the correct version of **ML** was consistent relative to **NF**. Our presentation of **ML** is based partly on the revised edition of Quine [3] and partly on Wang [1].

We formulate **ML** without use of a term operator of abstraction. **ML** is a first-order theory with one primitive predicate letter "$\in$" of degree 2. As usual, we write "$(x \in y)$" instead of "$\in(x, y)$". Our only terms are variables and consequently our wffs are all the first-order wffs definable by applying quantification and sentential connectives to our prime wffs. We begin by defining the notion "x is a set":

Definition 1. $M(x)$ for $(Ey)(x \in y)$, where y is different from x.

This is the same definition we gave for the system **NBG**. A set is something that is an element of something else.

Again, as with the system **NBG**, we define the restriction of a quantifier to sets: $(x)(M(x) \supset A(x))$ is the restriction of the quantifier (x) to sets and $(Ex)(M(x) \wedge A(x))$ is the restriction of (Ex) to sets. A wff in which every occurrence of every quantifier is restricted to sets is called *predicative*. We now state the axioms of **ML**.

ML.1. Same as **NF**.1. This is the axiom of extensionality.

ML.2. Where $A(x)$ is any wff of **ML** containing x free, and where y does not occur in $A(x)$, then $(Ey)(x)(x \in y \equiv M(x) \wedge A(x))$ is an axiom. This is the axiom of class existence. It is clearly the same as the axiom of class existence for **MKM** of Chapter 5. The usual deduction of Russell's paradox is avoided in the same way as in that system. **ML**.2 poses the existence of classes whose elements are sets, but we need an axiom of set existence. We pose:

ML.3. Let $A(x)$ be a wff of **ML** that is stratified and that is obtained from a stratified wff by restricting all bound variables to sets. $A(x)$ is a stratified, predicative wff containing x free. Let all the free variables of $A(x)$ occur in the list $x, y_1, \ldots, y_n$. Then, the following is an axiom: $M(y_1) \wedge M(y_2) \wedge \cdots \wedge M(y_n) \supset (Ew)(M(w) \wedge (x)(x \in w \equiv M(x) \wedge A(x)))$ where w does not occur in $A(x)$.

This completes the axioms of **ML**.

ML.3 says, in effect, that the sets of **NF** are the sets of **ML**. It says intuitively that, for any condition given by a predicative, stratified, wff $A(x)$ whose free variables are sets, there exists a *set* w whose elements are precisely those sets satisfying the condition. Because of **ML**.2, we also have the possibility of other classes that are not sets.

Exercise. Prove that there is a model for the axioms **ML**.1 and **ML**.2 in a domain with only one element.

The advantage of operating with **ML** is that certain difficulties of **NF** are avoided. For example, the principle of mathematical induction can be proved to hold for any wff, stratified or not. A theorem of infinity is also provable in **ML**. **ML** is a very liberal theory and liable to be dangerously close to the paradoxes because of its liberality. However, Wang's relative consistency proof of **ML** with respect to **NF** shows that **ML** is no more risky than **NF**. Because of this, logicians who work on Quine's systems focus their attention on **NF**, since its form is simpler and metatheorems are easier to prove for it than for **ML**. However, **ML** is more flexible as a foundation.

Exercise. Define the universal class V in **ML** and prove that $V \in V$. This differs from the universal class of **NBG**. That class is not an element of itself. In **ML**, the class of all sets is itself a set.

For a detailed development of arithmetic in **ML** (which nevertheless is quite close to our development in **NF**), consult Quine [3].

7.6 Conclusions

Quine originally propounded his systems as a basis for the logistic thesis that mathematics reduces to logic (See Quine [1] and [3]). To do this is to argue that **NF** (or **ML**) is a "naturally true" or "universally valid" system. Since some results of **NF** contradict our Cantorian intuition of sets, most mathematicians would reject such a claim. However, the paradoxes of set theory show that our original intuition is itself suspect. Consequently, something of the old intuition will probably be lost in any reformulation of set theory. In any case, the motivation of Quine's approach is clear. Recall that the notable failure of type theory in defending the logistic thesis was that the axiom of infinity, necessary for mathematics, was not a universally valid principle of type theory (it is not true in every model of the other axioms of **ST**). However, a theorem of infinity is provable in **NF**, and so any support for the view that **NF** is a "natural" system would tend to support the logistic thesis.

In recent years, Quine seems to have changed his position, at least to some degree. In any case, he clearly adopts a different philosophic attitude in Quine [4], in which he even refers to the unnatural character of **NF** (see Quine [4], p. 299). An important question for **NF** remains over the relative consistency of **NF** to **ZF**.