
I.2 The Language and Grammar of Mathematics

1 Introduction

It is a remarkable phenomenon that children can learn to speak without ever being consciously aware of the sophisticated grammar they are using. Indeed, adults too can live a perfectly satisfactory life without ever thinking about ideas such as parts of speech, subjects, predicates, or subordinate clauses. Both children and

adults can easily recognize ungrammatical sentences, at least if the mistake is not too subtle, and to do this it is not necessary to be able to explain the rules that have been violated. Nevertheless, there is no doubt that one's understanding of language is hugely enhanced by a knowledge of basic grammar, and this understanding is essential for anybody who wants to do more with language than use it unreflectingly as a means to a nonlinguistic end.

The same is true of mathematical language. Up to a point, one can do and speak mathematics without knowing how to classify the different sorts of words one is using, but many of the sentences of advanced mathematics have a complicated structure that is much easier to understand if one knows a few basic terms of mathematical grammar. The object of this section is to explain the most important mathematical "parts of speech," some of which are similar to those of natural languages and others quite different. These are normally taught right at the beginning of a university course in mathematics. Much of *The Companion* can be understood without a precise knowledge of mathematical grammar, but a careful reading of this article will help the reader who wishes to follow some of the later, more advanced parts of the book.

The main reason for using mathematical grammar is that the statements of mathematics are supposed to be completely precise, and it is not possible to achieve complete precision unless the language one uses is free of many of the vaguenesses and ambiguities of ordinary speech. Mathematical sentences can also be highly complex: if the parts that made them up were not clear and simple, then the unclarities would rapidly accumulate and render the sentences unintelligible.

To illustrate the sort of clarity and simplicity that is needed in mathematical discourse, let us consider the famous mathematical sentence "Two plus two equals four" as a sentence of English rather than of mathematics, and try to analyze it grammatically. On the face of it, it contains three nouns ("two," "two," and "four"), a verb ("equals") and a conjunction ("plus"). However, looking more carefully we may begin to notice some oddities. For example, although the word "plus" resembles the word "and," the most obvious example of a conjunction, it does not behave in quite the same way, as is shown by the sentence "Mary and Peter love Paris." The verb in this sentence, "love," is plural, whereas the verb in the previous sentence, "equals," was singular. So the word "plus" seems to take two objects (which happen to be numbers) and produce out of them a new, single object,

while "and" conjoins "Mary" and "Peter" in a looser way, leaving them as distinct people.

Reflecting on the word "and" a bit more, one finds that it has two very different uses. One, as above, is to link two nouns, whereas the other is to join two whole sentences together, as in "Mary likes Paris and Peter likes New York." If we want the basics of our language to be absolutely clear, then it will be important to be aware of this distinction. (When mathematicians are at their most formal, they simply outlaw the noun-linking use of "and"—a sentence such as "3 and 5 are prime numbers" is then paraphrased as "3 is a prime number and 5 is a prime number.")

This is but one of many similar questions: anybody who has tried to classify all words into the standard eight parts of speech will know that the classification is hopelessly inadequate. What, for example, is the role of the word "six" in the sentence "This section has six subsections"? Unlike "two" and "four" earlier, it is certainly not a noun. Since it modifies the noun "subsection" it would traditionally be classified as an adjective, but it does not behave like most adjectives: the sentences "My car is not very fast" and "Look at that tall building" are perfectly grammatical, whereas the sentences "My car is not very six" and "Look at that six building" are not just nonsense but ungrammatical nonsense. So do we classify adjectives further into numerical adjectives and nonnumerical adjectives? Perhaps we do, but then our troubles will be only just beginning. For example, what about possessive adjectives such as "my" and "your"? In general, the more one tries to refine the classification of English words, the more one realizes how many different grammatical roles there are.

2 Four Basic Concepts

Another word that famously has three quite distinct meanings is "is." The three meanings are illustrated in the following three sentences.

- (1) 5 is the square root of 25.
- (2) 5 is less than 10.
- (3) 5 is a prime number.

In the first of these sentences, "is" could be replaced by "equals": it says that two objects, 5 and the square root of 25, are in fact one and the same object, just as it does in the English sentence "London is the capital of the United Kingdom." In the second sentence, "is" plays a completely different role. The words "less than 10" form an adjectival phrase, specifying a property that numbers may or may not have, and "is" in this sentence is like "is"

in the English sentence “Grass is green.” As for the third sentence, the word “is” there means “is an example of,” as it does in the English sentence “Mercury is a planet.”

These differences are reflected in the fact that the sentences cease to resemble each other when they are written in a more symbolic way. An obvious way to write (1) is $5 = \sqrt{25}$. As for (2), it would usually be written $5 < 10$, where the symbol “ $<$ ” means “is less than.” The third sentence would normally not be written symbolically because the concept of a prime number is not quite basic enough to have universally recognized symbols associated with it. However, it is sometimes useful to do so, and then one must invent a suitable symbol. One way to do it would be to adopt the convention that if n is a positive integer, then $P(n)$ stands for the sentence “ n is prime.” Another way, which does not hide the word “is,” is to use the language of sets.

2.1 Sets

Broadly speaking, a *set* is a collection of objects, and in mathematical discourse these objects are mathematical ones such as numbers, points in space, or even other sets. If we wish to rewrite sentence (3) symbolically, another way to do it is to define P to be the collection, or set, of all prime numbers. Then (3) can be rewritten, “5 belongs to the set P .” This notion of belonging to a set is sufficiently basic to deserve its own symbol, and the symbol used is “ \in .” So a fully symbolic way of writing the sentence is $5 \in P$.

The members of a set are usually called its *elements*, and the symbol “ \in ” is usually read “is an element of.” So the “is” of sentence (3) is more like “ \in ” than “ $=$.” Although one cannot directly substitute the phrase “is an element of” for “is,” one can do so if one is prepared to modify the rest of the sentence a little.

There are three common ways to denote a specific set. One is to list its elements inside curly brackets: $\{2, 3, 5, 7, 11, 13, 17, 19\}$, for example, is the set whose elements are the eight numbers 2, 3, 5, 7, 11, 13, 17, and 19. The majority of sets considered by mathematicians are too large for this to be feasible—indeed, they are often infinite—so a second way to denote sets is to use dots to imply a list that is too long to write down: for example, the expressions $\{1, 2, 3, \dots, 100\}$ and $\{2, 4, 6, 8, \dots\}$ can be used to represent the set of all positive integers up to 100 and the set of all positive even numbers, respectively. A third way, and the way that is most important, is to define a set via a *property*: an example that shows how this is done is the expression $\{x : x \text{ is prime and } x < 20\}$. To read an expression such

as this, one first reads the opening curly bracket as “The set of.” Next, one reads the symbol that occurs before the colon. The colon itself one reads as “such that.” Finally, one reads what comes after the colon, which is the property that determines the elements of the set. In this instance, we end up saying, “The set of x such that x is prime and x is less than 20,” which is in fact equal to the set $\{2, 3, 5, 7, 11, 13, 17, 19\}$ considered earlier.

Many sentences of mathematics can be rewritten in set-theoretic terms. For example, sentence (2) earlier could be written as $5 \in \{n : n < 10\}$. Often there is no point in doing this (as here, where it is much easier to write $5 < 10$) but there are circumstances where it becomes extremely convenient. For example, one of the great advances in mathematics was the use of Cartesian coordinates to translate geometry into algebra and the way this was done was to define geometrical objects as sets of points, where points were themselves defined as pairs or triples of numbers. So, for example, the set $\{(x, y) : x^2 + y^2 = 1\}$ is (or represents) a circle of radius 1 with its center at the origin $(0, 0)$. That is because, by the Pythagorean theorem, the distance from $(0, 0)$ to (x, y) is $\sqrt{x^2 + y^2}$, so the sentence “ $x^2 + y^2 = 1$ ” can be reexpressed geometrically as “the distance from $(0, 0)$ to (x, y) is 1.” If all we ever cared about was which points were in the circle, then we could make do with sentences such as “ $x^2 + y^2 = 1$,” but in geometry one often wants to consider the entire circle as a single object (rather than as a multiplicity of points, or as a property that points might have), and then set-theoretic language is indispensable.

A second circumstance where it is usually hard to do without sets is when one is defining new mathematical objects. Very often such an object is a set together with a *mathematical structure* imposed on it, which takes the form of certain relationships among the elements of the set. For examples of this use of set-theoretic language, see sections 1 and 2, on number systems and algebraic structures, respectively, in SOME FUNDAMENTAL MATHEMATICAL DEFINITIONS [I.3].

Sets are also very useful if one is trying to do *meta-mathematics*, that is, to prove statements not about mathematical objects but about the process of mathematical reasoning itself. For this it helps a lot if one can devise a very simple language—with a small vocabulary and an uncomplicated grammar—into which it is in principle possible to translate all mathematical arguments. Sets allow one to reduce greatly the number of parts of speech that one needs, turning almost all of them into nouns. For example, with the help of the membership

symbol “ \in ” one can do without adjectives, as the translation of “5 is a prime number” (where “prime” functions as an adjective) into “ $5 \in P$ ” has already suggested.¹ This is of course an artificial process—imagine replacing “roses are red” by “roses belong to the set R ”—but in this context it is not important for the formal language to be natural and easy to understand.

2.2 Functions

Let us now switch attention from the word “is” to some other parts of the sentences (1)–(3), focusing first on the phrase “the square root of” in sentence (1). If we wish to think about this phrase grammatically, then we should analyze what sort of role it plays in a sentence, and the analysis is simple: in virtually any mathematical sentence where the phrase appears, it is followed by the name of a number. If the number is n , then this produces the slightly longer phrase, “the square root of n ,” which is a noun phrase that denotes a number and plays the same grammatical role as a number (at least when the number is used as a noun rather than as an adjective). For instance, replacing “5” by “the square root of 25” in the sentence “5 is less than 7” yields a new sentence, “The square root of 25 is less than 7,” that is still grammatically correct (and true).

One of the most basic activities of mathematics is to take a mathematical object and transform it into another one, sometimes of the same kind and sometimes not. “The square root of” transforms numbers into numbers, as do “four plus,” “two times,” “the cosine of,” and “the logarithm of.” A nonnumerical example is “the center of gravity of,” which transforms geometrical shapes (provided they are not too exotic or complicated to have a center of gravity) into points—meaning that if S stands for a shape, then “the center of gravity of S ” stands for a point. A *function* is, roughly speaking, a mathematical transformation of this kind.

It is not easy to make this definition more precise. To ask, “What is a function?” is to suggest that the answer should be a *thing* of some sort, but functions seem to be more like processes. Moreover, when they appear in mathematical sentences they do not behave like nouns. (They are more like prepositions, though with a definite difference that will be discussed in the next subsection.) One might therefore think it inappropriate to ask what kind of object “the square root of” is. Should one not simply be satisfied with the grammatical analysis already given?

As it happens, no. Over and over again, throughout mathematics, it is useful to think of a mathematical phenomenon, which may be complex and very un-thinglike, as a single object. We have already seen a simple example: a collection of infinitely many points in the plane or space is sometimes better thought of as a single geometrical shape. Why should one wish to do this for functions? Here are two reasons. First, it is convenient to be able to say something like, “The derivative of \sin is \cos ,” or to speak in general terms about some functions being differentiable and others not. More generally, functions can have *properties*, and in order to discuss those properties one needs to think of functions as things. Second, many algebraic structures are most naturally thought of as sets of functions. (See, for example, the discussion of groups and symmetry in [I.3 §2.1]. See also HILBERT SPACES [III.37], FUNCTION SPACES [III.29], and VECTOR SPACES [I.3 §2.3].)

If f is a function, then the notation $f(x) = y$ means that f turns the object x into the object y . Once one starts to speak formally about functions, it becomes important to specify exactly which objects are to be subjected to the transformation in question, and what sort of objects they can be transformed into. One of the main reasons for this is that it makes it possible to discuss another notion that is central to mathematics, that of *inverting* a function. (See [I.4 §1] for a discussion of why it is central.) Roughly speaking, the inverse of a function is another function that undoes it, and that it undoes; for example, the function that takes a number n to $n - 4$ is the inverse of the function that takes n to $n + 4$, since if you add four and then subtract four, or vice versa, you get the number you started with.

Here is a function f that cannot be inverted. It takes each number and replaces it by the nearest multiple of 100, rounding up if the number ends in 50. Thus, $f(113) = 100$, $f(3879) = 3900$, and $f(1050) = 1100$. It is clear that there is no way of undoing this process with a function g . For example, in order to undo the effect of f on the number 113 we would need $g(100)$ to equal 113. But the same argument applies to every number that is at least as big as 50 and smaller than 150, and $g(100)$ cannot be more than one number at once.

Now let us consider the function that doubles a number. Can this be inverted? Yes it can, one might say: just divide the number by two again. And much of the time this would be a perfectly sensible response, but not, for example, if it was clear from the context that the numbers being talked about were positive integers. Then one might be focusing on the difference between even and

1. For another discussion of adjectives see ARITHMETIC GEOMETRY [IV.6 §3.1].

odd numbers, and this difference could be encapsulated by saying that odd numbers are precisely those numbers n for which the equation $2x = n$ does *not* have a solution. (Notice that one can undo the doubling process by halving. The problem here is that the relationship is not symmetrical: there is no function that can be undone by doubling, since you could never get back to an odd number.)

To specify a function, therefore, one must be careful to specify two sets as well: the *domain*, which is the set of objects to be transformed, and the *range*, which is the set of objects they are allowed to be transformed into. A function f from a set A to a set B is a rule that specifies, for each element x of A , an element $y = f(x)$ of B . (Not every element of the range needs to be used: consider once again the example of “two times” when the domain and range are both the set of all positive integers.)

The following symbolic notation is used. The expression $f : A \rightarrow B$ means that f is a function with domain A and range B . If we then write $f(x) = y$, we know that x must be an element of A and y must be an element of B . Another way of writing $f(x) = y$ that is sometimes more convenient is $f : x \mapsto y$. (The bar on the arrow is to distinguish it from the arrow in $f : A \rightarrow B$, which has a very different meaning.)

If we want to undo the effect of a function $f : A \rightarrow B$, then we can, as long as we avoid the problem that occurred with the approximating function discussed earlier. That is, we can do it if $f(x)$ and $f(x')$ are different whenever x and x' are different elements of A . If this condition holds, then f is called an *injection*. On the other hand, if we want to find a function g that is undone by f , then we can do so as long as we avoid the problem of the integer-doubling function. That is, we can do it if every element y of B is equal to $f(x)$ for some element x of A (so that we have the option of setting $g(y) = x$). If this condition holds, then f is called a *surjection*. If f is both an injection and a surjection, then f is called a *bijection*. Bijections are precisely the functions that have inverses.

It is important to realize that not all functions have tidy definitions. Here, for example, is the specification of a function from the positive integers to the positive integers: $f(n) = n$ if n is a prime number, $f(n) = k$ if n is of the form 2^k for an integer k greater than 1, and $f(n) = 13$ for all other positive integers n . This function has an unpleasant, arbitrary definition but it is nevertheless a perfectly legitimate function. Indeed, “most” functions, though not most functions that one actually uses, are so arbitrary that they cannot be defined. (Such functions may not be useful as individual objects, but they

are needed so that the set of all functions from one set to another has an interesting mathematical structure.)

2.3 Relations

Let us now think about the grammar of the phrase “less than” in sentence (2). As with “the square root of,” it must always be followed by a mathematical object (in this case a number again). Once we have done this we obtain a phrase such as “less than n ,” which is importantly different from “the square root of n ” because it behaves like an adjective rather than a noun, and refers to a property rather than an object. This is just how prepositions behave in English: look, for example, at the word “under” in the sentence “The cat is under the table.”

At a slightly higher level of formality, mathematicians like to avoid too many parts of speech, as we have already seen for adjectives. So there is no symbol for “less than”: instead, it is combined with the previous word “is” to make the phrase “is less than,” which is denoted by the symbol “ $<$.” The grammatical rules for this symbol are once again simple. To use “ $<$ ” in a sentence, one should precede it by a noun and follow it by a noun. For the resulting grammatically correct sentence to make sense, the nouns should refer to numbers (or perhaps to more general objects that can be put in order). A mathematical “object” that behaves like this is called a *relation*, though it might be more accurate to call it a potential relationship. “Equals” and “is an element of” are two other examples of relations.

As with functions, it is important, when specifying a relation, to be careful about which objects are to be related. Usually a relation comes with a set A of objects that may or may not be related to each other. For example, the relation “ $<$ ” might be defined on the set of all positive integers, or alternatively on the set of all real numbers; strictly speaking these are different relations. Sometimes relations are defined with reference to two sets A and B . For example, if the relation is “ \in ,” then A might be the set of all positive integers and B the set of all sets of positive integers.

There are many situations in mathematics where one wishes to regard different objects as “essentially the same,” and to help us make this idea precise there is a very important class of relations known as *equivalence relations*. Here are two examples. First, in elementary geometry one sometimes cares about shapes but not about sizes. Two shapes are said to be *similar* if one can be transformed into the other by a combination of reflections, rotations, translations, and enlargements (see figure 1); the relation “is similar to” is an

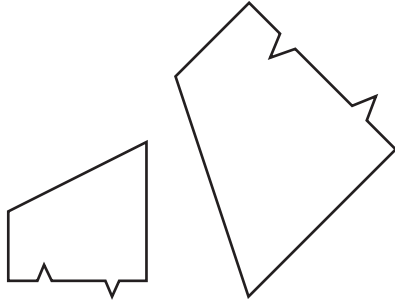


Figure 1 Similar shapes.

equivalence relation. Second, when doing ARITHMETIC MODULO m [III.61], one does not wish to distinguish between two whole numbers that differ by a multiple of m : in this case one says that the numbers are *congruent* (mod m); the relation “is congruent (mod m) to” is another equivalence relation.

What exactly is it that these two relations have in common? The answer is that they both take a set (in the first case the set of all geometrical shapes, and in the second the set of all whole numbers) and split it into parts, called *equivalence classes*, where each part consists of objects that one wishes to regard as essentially the same. In the first example, a typical equivalence class is the set of all shapes that are similar to some given shape; in the second, it is the set of all integers that leave a given remainder when you divide by m (for example, if $m = 7$ then one of the equivalence classes is the set $\{\dots, -16, -9, -2, 5, 12, 19, \dots\}$).

An alternative definition of what it means for a relation \sim , defined on a set A , to be an equivalence relation is that it has the following three properties. First, it is *reflexive*, which means that $x \sim x$ for every x in A . Second, it is *symmetric*, which means that if x and y are elements of A and $x \sim y$, then it must also be the case that $y \sim x$. Third, it is *transitive*, meaning that if x , y , and z are elements of A such that $x \sim y$ and $y \sim z$, then it must be the case that $x \sim z$. (To get a feel for these properties, it may help if you satisfy yourself that the relations “is similar to” and “is congruent (mod m) to” both have all three properties, while the relation “ $<$,” defined on the positive integers, is transitive but neither reflexive nor symmetric.)

One of the main uses of equivalence relations is to make precise the notion of QUOTIENT [I.3 §3.3] constructions.

2.4 Binary Operations

Let us return to one of our earlier examples, the sentence “Two plus two equals four.” We have analyzed the word “equals” as a relation, an expression that sits between the noun phrases “two plus two” and “four” and makes a sentence out of them. But what about “plus”? That also sits between two nouns. However, the result, “two plus two,” is not a sentence but a noun phrase. That pattern is characteristic of *binary operations*. Some familiar examples of binary operations are “plus,” “minus,” “times,” “divided by,” and “raised to the power.”

As with functions, it is customary, and convenient, to be careful about the set to which a binary operation is applied. From a more formal point of view, a binary operation on a set A is a function that takes pairs of elements of A and produces further elements of A from them. To be more formal still, it is a function with the set of all pairs (x, y) of elements of A as its domain and with A as its range. This way of looking at it is not reflected in the notation, however, since the symbol for the operation comes between x and y rather than before them: we write $x + y$ rather than $+(x, y)$.

There are four properties that a binary operation may have that are very useful if one wants to manipulate sentences in which it appears. Let us use the symbol $*$ to denote an arbitrary binary operation on some set A . The operation $*$ is said to be *commutative* if $x * y$ is always equal to $y * x$, and *associative* if $x * (y * z)$ is always equal to $(x * y) * z$. For example, the operations “plus” and “times” are commutative and associative, whereas “minus,” “divided by,” and “raised to the power” are neither (for instance, $9 - (5 - 3) = 7$ while $(9 - 5) - 3 = 1$). These last two operations raise another issue: unless the set A is chosen carefully, they may not always be defined. For example, if one restricts one’s attention to the positive integers, then the expression $3 - 5$ has no meaning. There are two conventions one could imagine adopting in response to this. One might decide not to insist that a binary operation should be defined for every pair of elements of A , and to regard it as a desirable extra property of an operation if it *is* defined everywhere. But the convention actually in force is that binary operations *do* have to be defined everywhere, so that “minus,” though a perfectly good binary operation on the set of all integers, is not a binary operation on the set of all positive integers.

An element e of A is called an *identity* for $*$ if $e * x = x * e = x$ for every element x of A . The two most obvious examples are 0 and 1, which are identities for “plus” and “times,” respectively. Finally, if $*$ has an identity e and

x belongs to A , then an *inverse* for x is an element y such that $x * y = y * x = e$. For example, if $*$ is “plus” then the inverse of x is $-x$, while if $*$ is “times” then the inverse is $1/x$.

These basic properties of binary operations are fundamental to the structures of abstract algebra. See FOUR IMPORTANT ALGEBRAIC STRUCTURES [I.3 §2] for further details.

3 Some Elementary Logic

3.1 Logical Connectives

A *logical connective* is the mathematical equivalent of a conjunction. That is, it is a word (or symbol) that joins two sentences to produce a new one. We have already discussed an example, namely “and” in its sentence-linking meaning, which is sometimes written by the symbol “ \wedge ,” particularly in more formal or abstract mathematical discourse. If P and Q are statements (note here the mathematical habit of representing not just numbers but any objects whatsoever by single letters), then $P \wedge Q$ is the statement that is true if and only if both P and Q are true.

Another connective is the word “or,” a word that has a more specific meaning for mathematicians than it has for normal speakers of the English language. The mathematical use is illustrated by the tiresome joke of responding, “Yes please,” to a question such as, “Would you like your coffee with or without sugar?” The symbol for “or,” if one wishes to use a symbol, is “ \vee ,” and the statement $P \vee Q$ is true if and only if P is true or Q is true. This is taken to include the case when they are both true, so “or,” for mathematicians, is always the so-called *inclusive* version of the word.

A third important connective is “implies,” which is usually written “ \Rightarrow .” The statement $P \Rightarrow Q$ means, roughly speaking, that Q is a consequence of P , and is sometimes read as “if P then Q .” However, as with “or,” this does not mean quite what it would in English. To get a feel for the difference, consider the following even more extreme example of mathematical pedantry. At the supper table, my young daughter once said, “Put your hand up if you are a girl.” One of my sons, to tease her, put his hand up on the grounds that, since she had not added, “and keep it down if you are a boy,” his doing so was compatible with her command.

Something like this attitude is taken by mathematicians to the word “implies,” or to sentences containing the word “if.” The statement $P \Rightarrow Q$ is considered to be true under all circumstances except one: it is not true if P is true and Q is false. This is the *definition* of “implies.” It

can be confusing because in English the word “implies” suggests some sort of connection between P and Q , that P in some way causes Q or is at least relevant to it. If P causes Q then certainly P cannot be true without Q being true, but all a mathematician cares about is this logical consequence and not whether there is any reason for it. Thus, if you want to prove that $P \Rightarrow Q$, all you have to do is rule out the possibility that P could be true and Q false at the same time. To give an example: if n is a positive integer, then the statement “ n is a perfect square with final digit 7” implies the statement “ n is a prime number,” not because there is any connection between the two but because no perfect square ends in a 7. Of course, implications of this kind are less interesting mathematically than more genuine-seeming ones, but the reward for accepting them is that, once again, one avoids being confused by some of the ambiguities and subtle nuances of ordinary language.

3.2 Quantifiers

Yet another ambiguity in the English language is exploited by the following old joke that suggests that our priorities need to be radically rethought.

- (4) Nothing is better than lifelong happiness.
- (5) But a cheese sandwich is better than nothing.
- (6) Therefore, a cheese sandwich is better than lifelong happiness.

Let us try to be precise about how this play on words works (a good way to ruin any joke, but not a tragedy in this case). It hinges on the word “nothing,” which is used in two different ways. The first sentence means “There is no single thing that is better than lifelong happiness,” whereas the second means “It is better to have a cheese sandwich than to have nothing at all.” In other words, in the second sentence, “nothing” stands for what one might call the null option, the option of having nothing, whereas in the first it does not (to have nothing is not better than to have lifelong happiness).

Words like “all,” “some,” “any,” “every,” and “nothing” are called *quantifiers*, and in the English language they are highly prone to this kind of ambiguity. Mathematicians therefore make do with just two quantifiers, and the rules for their use are much stricter. They tend to come at the beginning of sentences, and can be read as “for all” (or “for every”) and “there exists” (or “for some”). A rewriting of sentence (4) that renders it unambiguous (and much less like a real English sentence) is

- (4') For all x , lifelong happiness is better than x .

The second sentence cannot be rewritten in these terms because the word “nothing” is not playing the role of a quantifier. (Its nearest mathematical equivalent is something like the *empty set*, that is, the set with no elements.)

Armed with “for all” and “there exists,” we can be clear about the difference between the beginnings of the following sentences.

- (7) Everybody likes at least one drink, namely water.
- (8) Everybody likes at least one drink; I myself go for red wine.

The first sentence makes the point (not necessarily correctly) that there is one drink that everybody likes, whereas the second claims merely that we all have something we like to drink, even if that something varies from person to person. The precise formulations that capture the difference are as follows.

- (7') There exists a drink D such that, for every person P , P likes D .
- (8') For every person P there exists a drink D such that P likes D .

This illustrates an important general principle: if you take a sentence that begins “for every x there exists y such that ...” and interchange the two parts so that it now begins “there exists y such that, for every x , ...,” then you obtain a much stronger statement, since y is no longer allowed to depend on x . If the second statement is still true—that is, if you really can choose a y that works for all the x at once—then the first statement is said to hold *uniformly*.

The symbols \forall and \exists are often used to stand for “for all” and “there exists,” respectively. This allows us to write quite complicated mathematical sentences in a highly symbolic form if we want to. For example, suppose we let P be the set of all primes, as we did earlier. Then the following symbols make the claim that there are infinitely many primes, or rather a slightly different claim that is equivalent to it.

$$(9) \forall n \exists m \quad (m > n) \wedge (m \in P).$$

In words, this says that for every n we can find some m that is both bigger than n and a prime. If we wish to unpack sentence (6) further, we could replace the part $m \in P$ by

$$(10) \forall a, b \quad ab = m \Rightarrow ((a = 1) \vee (b = 1)).$$

There is one final important remark to make about the quantifiers “ \forall ” and “ \exists .” I have presented them as if they

were freestanding, but actually a quantifier is always associated with a set (one says that it *quantifies over* that set). For example, sentence (10) would not be a translation of the sentence “ m is prime” if a and b were allowed to be fractions: if $a = 3$ and $b = \frac{7}{3}$ then $ab = 7$ without either a or b equaling 1, but this does not show that 7 is not a prime. Implicit in the opening symbols $\forall a, b$ is the idea that a and b are intended to be *positive integers*. If this had not been clear from the context, then we could have used the symbol \mathbb{N} (which stands for the set of all positive integers) and started sentence (10) with $\forall a, b \in \mathbb{N}$ instead.

3.3 Negation

The basic idea of negation in mathematics is very simple: there is a symbol, “ \neg ,” which means “not,” and if P is any mathematical statement, then $\neg P$ stands for the statement that is true if and only if P is not true. However, this is another example of a word that has a slightly more restricted meaning to mathematicians than it has in ordinary speech.

To illustrate this phenomenon once again, let us take A to be a set of positive integers and ask ourselves what the negation is of the sentence “Every number in the set A is odd.” Many people when asked this question will suggest, “Every number in the set A is even.” However, this is wrong: if one thinks carefully about what exactly would have to happen for the first sentence to be false, one realizes that all that is needed is that *at least one* number in A should be even. So in fact the negation of the sentence is, “There exists a number in A that is even.”

What explains the temptation to give the first, incorrect answer? One possibility emerges when one writes the sentence more formally, thus:

$$(11) \forall n \in A \quad n \text{ is odd.}$$

The first answer is obtained if one negates just the last part of this sentence, “ n is odd”; but what is asked for is the negation of the *whole sentence*. That is, what is wanted is not

$$(12) \forall n \in A \quad \neg(n \text{ is odd}),$$

but rather

$$(13) \neg(\forall n \in A \quad n \text{ is odd}),$$

which is equivalent to

$$(14) \exists n \in A \quad n \text{ is even.}$$

A second possible explanation is that one is inclined (for psycholinguistic reasons) to think of the phrase “every element of A ” as denoting something like a single, typical element of A . If that comes to have the feel of a particular number n , then we may feel that the negation of “ n is odd” is “ n is even.” The remedy is not to think of the phrase “every element of A ” on its own: it should always be part of the longer phrase, “for every element of A .”

3.4 Free and Bound Variables

Suppose we say something like, “At time t the speed of the projectile is v .” The letters t and v stand for real numbers, and they are called *variables*, because in the back of our mind is the idea that they are changing. More generally, a variable is any letter used to stand for a mathematical object, whether or not one thinks of that object as changing through time. Let us look once again at the formal sentence that said that a positive integer m is prime:

$$(10) \quad \forall a, b \quad ab = m \Rightarrow ((a = 1) \vee (b = 1)).$$

In this sentence, there are three variables, a , b , and m , but there is a very important grammatical and semantic difference between the first two and the third. Here are two results of that difference. First, the sentence does not really make sense unless we already know what m is from the context, whereas it is important that a and b do *not* have any prior meaning. Second, while it makes perfect sense to ask, “For which values of m is sentence (10) true?” it makes no sense at all to ask, “For which values of a is sentence (10) true?” The letter m in sentence (10) stands for a fixed number, not specified in this sentence, while the letters a and b , because of the initial $\forall a, b$, do not *stand for* numbers—rather, in some way they search through all pairs of positive integers, trying to find a pair that multiply together to give m . Another sign of the difference is that you can ask, “What number is m ?” but not, “What number is a ?” A fourth sign is that the meaning of sentence (10) is completely unaffected if one uses different letters for a and b , as in the reformulation

$$(10') \quad \forall c, d \quad cd = m \Rightarrow ((c = 1) \vee (d = 1)).$$

One cannot, however, change m to n without establishing first that n denotes the same integer as m . A variable such as m , which denotes a specific object, is called a *free* variable. It sort of hovers there, free to take any value. A variable like a and b , of the kind that does not denote a specific object, is called a *bound* variable, or sometimes a *dummy* variable. (The word “bound”

is used mainly when the variable appears just after a quantifier, as in sentence (10).)

Yet another indication that a variable is a dummy variable is when the sentence in which it occurs can be rewritten without it. For example, the notation $\sum_{n=1}^{100} f(n)$ is shorthand for $f(1) + f(2) + \cdots + f(100)$, and the second way of writing it does not involve the letter n , so n was not really standing for anything in the first way. Sometimes, actual elimination is not possible, but one feels it could be done in principle. For instance, the sentence “For every real number x , x is either positive, negative, or zero” is a bit like putting together infinitely many sentences such as “ t is either positive, negative, or zero,” one for each real number t , none of which involve a variable.

4 Levels of Formality

It is a surprising fact that a small number of set-theoretic concepts and logical terms can be used to provide a precise language that is versatile enough to express all the statements of ordinary mathematics. There are some technicalities to sort out, but even these can often be avoided if one allows not just sets but also numbers as basic objects. However, if you look at a well-written mathematics paper, then much of it will be written not in symbolic language peppered with symbols such as \forall and \exists , but in what appears to be ordinary English. (Some papers are written in other languages, particularly French, but English has established itself as the international language of mathematics.) How can mathematicians be confident that this ordinary English does not lead to confusion, ambiguity, and even incorrectness?

The answer is that the language typically used is a careful compromise between fully colloquial English, which would indeed run the risk of being unacceptably imprecise, and fully formal symbolism, which would be a nightmare to read. The ideal is to write in as friendly and approachable a way as possible, while making sure that the reader (who, one assumes, has plenty of experience and training in how to read mathematics) can see easily how what one writes could be made more formal if it became important to do so. And sometimes it does become important: when an argument is difficult to grasp it may be that the only way to convince oneself that it is correct is to rewrite it more formally.

Consider, for example, the following reformulation of the principle of mathematical induction, which underlies many proofs:

- (15) Every nonempty set of positive integers has a least element.

If we wish to translate this into a more formal language we need to strip it of words and phrases such as “nonempty” and “has.” But this is easily done. To say that a set A of positive integers is nonempty is simply to say that there is a positive integer that belongs to A . This can be stated symbolically:

$$(16) \exists n \in \mathbb{N} \quad n \in A.$$

What does it mean to say that A has a least element? It means that there exists an element x of A such that every element y of A is either greater than x or equal to x itself. This formulation is again ready to be translated into symbols:

$$(17) \exists x \in A \quad \forall y \in A \quad (y > x) \vee (y = x).$$

Statement (15) says that (16) implies (17) for every set A of positive integers. Thus, it can be written symbolically as follows:

$$(18) \forall A \subset \mathbb{N} \\ [(\exists n \in \mathbb{N} \quad n \in A) \\ \Rightarrow (\exists x \in A \quad \forall y \in A \quad (y > x) \vee (y = x))].$$

Here we have two very different modes of presentation of the same mathematical fact. Obviously (15) is much easier to understand than (18). But if, for example, one is concerned with the foundations of mathematics, or wishes to write a computer program that checks the correctness of proofs, then it is better to work with a greatly pared-down grammar and vocabulary, and then (18) has the advantage. In practice, there are many different levels of formality, and mathematicians are adept at switching between them. It is this that makes it possible to feel completely confident in the correctness of a mathematical argument even when it is *not* presented in the manner of (18)—though it is also this that allows mistakes to slip through the net from time to time.