The Evolution of Logic

W. D. HART

CAMBRIDGE
UNIVERSITY PRESS

Contents

ONE

✦

Cantor's Paradise

Cantor's quarry was the infinite. The mathematics of number had always been about objects of which there are infinitely many, like natural numbers, or objects of which not only are there infinitely many but each is also itself infinite, like real numbers with endless decimal expansions. The infinities of geometry, like the infinity of points on a line or triangles in a plane, had always been there, but the applications of the calculus in geometry made its infinities more salient. The recognition of the infinity of its subject matter was always a reason not to test the conjectures of mathematics by checking the examples but rather to prefer proof. Aristotle urged that the infinite could only ever be potential, like a process with no fixed end, but that completed actual infinite wholes were ruled out. Such views look to countenance possibilities that could not be actual, which sounds contradictory, but even Gauss, the prince of mathematicians, had a *horror infiniti*. Cantor swam against the tide.

To work out a theory of the infinite per se, Cantor needed to figure out which things are classified as finite or infinite. That is one source of his interest in sets. For this purpose sets should be any old collections, whether unified by having something in common or not, like the Walrus's shoes and ships and cabbages and kings. Sets should be an utterly general sort, so whether there are infinitely many such and suches can always be re-asked as whether the set of such and suches is infinite. As horses are the kind that divides into stallions and mares, so sets are the kind that divides into finite and infinite.

Cantor distinguished between two sorts of infinity, one where order is front and center, and another where it is less obvious. (Order can often be taken as process finished and complete.) Since order is an extra, let us first put it aside. We want to articulate what it is for two sets to be the same in size. There are as many digits on your left hand as on your right. One way to check this is to count each and get the same answer, five, in both cases, but that procedure assumes number, something we also want to articulate. Another way is to match your digits one-to-one, so that each is matched with just one partner. This you can do without counting or numerical claims. Order does not matter. Palm to palm, you can match thumb to thumb, index finger to index finger, and so on. But you can also invert one hand and match thumb to pinkie, index to ring finger, and so on. And there are obviously many (120, in fact) ways to tie each digit to a unique digit on the other hand.

The general idea is that a set A has as many members as a set B exactly in case there is a way to match the members of A with the members of B one-to-one. The phrase "one-to-one" may make you worry that numbers like one are being smuggled in surreptitiously. The honest way to allay this worry is to lay out the set theoretic nuts and bolts of matching one-to-one so it is clear no numbers have snuck in. Laying out these nuts and bolts is also a way of illustrating how sets have become the arena in which logic, mathematics, and more are conducted. Sets are not just the natural kind of infinity; they are also a natural kind across logic, mathematics, and beyond. Frege's aim was to reduce the mathematics of number to logic. To do so, he treated extensions (of predicates, properties, or concepts) considerably more systematically than the comparatively casual use traditional logic had made of extensions for centuries before Frege. His treatment of extensions got into enough trouble that it is at least doubtful whether the mathematics of number is reducible to logic. But Frege's systematic treatment of extensions is an important stage in sets becoming the arena of mathematics and logic.

There are two primitive predicates in our exposition of basic naïve set theory. (We'll see later what the naïveté is.) We want a predicate for the relation of, say, a senator to the set of senators. This is called the membership, or elementhood, relation, and is usually written for short as similar to the small Greek letter epsilon, or $\in$. So if S is the set of states and A is the state of Alabama, then $A \in S$ says that Alabama is a member of the set of states. If our theory were to be a theory of nothing but sets, $\in$ could be our only primitive, and in that way set theory is the laws of the membership relation. But if we want to allow room for application to things like people and rocks that we don't think of as sets, so that we can have the set of people and the set of rocks, then we should also take identity as a primitive. We write this, as usual, as the double bar, =. To say that 7 + 5 = 12 means that the number 12 is the same thing as the sum of 7 and 5.

It is central to sets that they are identical when they have the same members. There aren't two totalities of all and only the shoes. The principle that sets with the same members are identical is called extensionality. Were we discussing nothing but sets, we could take membership as our only primitive predicate and use extensionality to introduce identity. But if we include things like the Rock of Gibraltar and Peter Abelard that presumably are not sets, then since only sets have members, Gibraltar and Abelard will have the same members, namely none, and yet not be identical. So when sets are applied, it is natural to assume identity as well as membership. Extensionality distinguishes sets from predicates and properties. Two predicates like "is directly over Big Ben" and "is directly above Big Ben" are true of all and only the same things, and yet are different predicates. Being spelled the same suffices for predicates to be identical. Properties can be had by the same things and yet differ. Easy instances are empty properties like being a centaur and being a griffin. But instantiated properties, like having a heart and having a liver, also seem to differ even if all and only the animals with hearts have livers. Some say that necessary coextensiveness suffices for property identity; others reply that necessity is unclear (without making it clear what clarity requires). Whatever the rights and wrongs of that dispute, there is more worked-out and settled lore about sets than about properties, so logicians and mathematicians favor sets over properties.

There are two systematic ways to name sets. If a set is finite and we have names for its members, then curly brackets enclosing a list of those names separated by commas is a name for that set. So

{Mercury, Venus, Earth}

is the set of the three inmost planets of our solar system. Since Cantor's quarry was the infinite, such names would not have satisfied him. Suppose we have a predicate like "Ralph gave x a present." Here the variable "x" marks a blank that may be filled by singular terms (proper names, definite descriptions, demonstratives) denoting things that are or are not targets of Ralph's generosity. Abbreviate this predicate as "Px." Then

$$\{x \mid Px\}$$

is read "the set of all things, say x is one, such that Px" and in our case would be the set of all and only the recipients of Ralph's generosity. This set is called the extension of the predicate "Px," uniqueness here being justified by extensionality. The assumption that every predicate has a set that is its extension is called comprehension. Naïve set theory is the theory whose axioms are extensionality and comprehension, and as we shall see, comprehension is thought to be its naïveté.

The notation {x | Px} is called set abstraction. List terms can be replaced by abstracts on the model of

$$\{x \mid x = \text{Mercury or } x = \text{Venus or } x = \text{Earth}\},$$

so we can make do with abstraction if we wish to be economical. The abstraction notation was introduced by Giuseppe Peano. Like the definite description operator, it applies to predicates and yields singular terms. Such terms may occur in yet further predicates, whence intricate nesting may ensue. Abstraction and membership are like inverses of each other. When Pa, the predication factors into a being a member of the set of Ps; Quine calls this the principle of abstraction. When a is a member of the set of Ps, membership and abstraction cancel out, and so Pa; this Quine calls concretion.

Comprehension says there is a set of all those things not identical with themselves (or a set of all unicorns), and extensionality says it is unique. This set is called the empty set, and it is denoted by ∅, which is not the Greek letter phi, but similar to the Danish and Norwegian slashed O. Some people who think of sets as somehow constituted out of, or dependent for their existence on, their elements have metaphysical qualms about the empty set. But an empty set need be no more troubling than an empty glass. Extensionality says that a set's members suffice to fix its identity, but this is neither to say the set is constituted from its members nor to say it depends for its existence on them. Besides, the hypothesis that there is an empty set has proved its utility time and again, and confirmation need not be cowed by metaphysical intuitions.

For any objects a and b, there is a unique set {a, b} whose members are a and b. Since {a, b} and {b, a} have the same members, extensionality says they are identical. So {a, b} is called the unordered pair of a and b. When a is b, their unordered pair is the set whose sole member is a; this is written {a} and is called the unit set or singleton of a. If a is itself a set with none or many members, it will not have the same members as its singleton, so in general a should be distinguished from its singleton. (But in what might seem an excess of economic zeal, Quine favored identifying a non-set with its unit set, as he showed how to do consistently.)

The empty set and unordered pairs assure us some sets outright. There are also operations on sets that assure us their values given their arguments. The Boolean operations, named for George Boole, correspond to truth functions. Thus the union of a and b, written a ∪ b, is the set of all x such that x ∈ a or x ∈ b. (The notation "∪" is Peano's.) With ∅ and unit sets, repeated union gives us all finite sets. The intersection of a and b, written a ∩ b, is the set of all x such that x ∈ a and x ∈ b. (The notation "∩" is also Peano's.) The intersection of the set of all odd numbers and the set of all even numbers is the empty set. Such sets are called disjoint. Without the empty set, disjoint sets would have no intersection, and we could not form a ∩ b without checking that a and b meet; the convenience of always being able to form a ∩ b is an example of the utility of ∅. The complement of a, written variously whence we pick ā, is the set of all things not in a. Complements, as we shall see, are a mark of naïveté, and sophistication sometimes favors differences, written a – b and explained as the set of all x such that x ∈ a but x ∉ b. (The ∈ with a stroke is denial of membership.)

In addition to Boolean operations, we also have the subset, or inclusion, relation. A set a is a subset of a set b, written a ⊆ b (like a softened less-or-equal sign), just in case all members of a are members of b. If b is also a subset of a, then they have the same members and so are identical. Note that when a ∈ b, then every member of {a} is a member of b, so {a} ⊆ b. Thereby may, but need not, hang a tale. Some people picture a layered world. On the ground floor, or layer 0, are the non-sets, the shoes and ships and so on. On layer 1 are the sets of things on layer 0. On layer 2 are either the sets of things on layer 1 (if, like Russell, you like your layers exclusive) or the things on layer 0 or layer 1 (if you like your layers cumulative). And so on for longer than you might expect. On this picture, Plato and everybody else is on layer 0, while the set of people and Plato's unit set are on layer 1. Then ∈ relates *across* layers (Plato is a member of his unit set and the set of people), while ⊆ relates *within* layers (the singleton of Plato is a subset of the set of people). It took a long, long time for us to learn to distinguish between ∈ and ⊆. The distinction was drawn clearly and driven home only in the nineteenth century. The premisses and conclusions of traditional syllogisms were either universal (All men are mortal) or particular (Some dogs are terriers). Singular premisses (Socrates is a man) were recognized, as in the old textbook inference from our universal and singular premisses to a singular conclusion (Socrates is mortal), but the effort to assimilate the singular to the universal or particular encouraged a confusion between $\in$ and $\subseteq$, as if Socrates were a tiny species.

The picture of layers helps distinguish between $\in$ and $\subseteq$, which is a virtue of it. Some people think it is the only right, or possible or coherent, way to picture the world. Maybe, but that view carries substantial commitments, so be wary of buying into it thoughtlessly. We will see larger issues later, but here is a smaller one. Consider propositions and self-reference. Russell (and probably Leibniz) thought of propositions as extensions of sentences as sets are extensions of predicates and as its denotation is the extension of a name. For example, the proposition that Socrates is bald would be the ordered pair <s, B> whose first member, s, is Socrates and whose second member, B, is the set of bald people. (We will get to ordered pairs very soon, but for now the important thing is that when a is different from b, the ordered pair <a, b> with a first and b second is a different thing from <b, a> with b first and a second.) This proposition <s, B> is true just in case $s \in B$, which opens a natural story about truth. Now consider a self-referential proposition like

This proposition can be expressed in eight words.

Let E be the set of propositions expressible in eight words, and let p be the proposition we are now considering. On Russell's conception, p is <p, E>, the doubling being the self-reference. We will soon construe ordered pairs as sets, and on the layered picture, an ordered pair will lie two layers above its members. On a layered picture, a set lies on a layer higher than its members, which would forbid self-referential propositions. But proposition p seems in order, indeed true, and we will later see more systematic reasons for reluctance to give up self-reference. It would not be shrewd to commit fully to the layered picture unreflectively, even if it is the conventional wisdom.

The set of tigers is the extension of the predicate "is a tiger." This predicate is unary (Latin) or monadic (Greek), both of which mean that it has one blank or empty space that on being filled with a singular term (like "Tony") yields a sentence. Each predicate has a number of blanks, filling all of which with singular terms yields a sentence. This number is called the predicate's polyadicity (Greek) or, much more rarely, its arity (Latin). The predicate "love" is binary (or dyadic) since it has two blanks for names, as in "Regina loved Søren," and "give" is ternary (or triadic) since it has three blanks, as in "The president gave the contract to his brother-in-law.” The Greek and Latin of logicians give out and they speak instead of 5-adic (or 5-ary) predicates. (Some predicates, as in “Andrew united Bob, Curt, David, and Ed in a conspiracy,” seem to lack a unique polyadicity, but they are rare.) In reckoning the polyadicity of a predicate in a sentence, one may count as many of the occurrences of singular terms as one wishes. For example, in

Richard gave the diamond to Elizabeth

one may count three singular terms filling the blanks in a ternary predicate, but one may count any two filling blanks in a more complex binary predicate, and one may count any one filling the blank in a yet more complex unary predicate. The logician is prescinding from grammatical roles (like direct or indirect object) and, as it were, counting several singular terms all as several subjects of a polyadic predicate.

The set of tigers is the extension of the monadic predicate “is a tiger.” We would also like extensions for polyadic predicates. As tigers one by one fill out the extension of “is a tiger,” we expect pairs to fill out the extension of a dynadic predicate like “loves.” But we notice immediately that order matters. Regina seems to have been a normal person and to have loved Søren, but we owe Kierkegaard’s works at least in part to his inability to make up his mind that he loved Regina. Unrequited love shows that the members of the extension of “loves” cannot be unordered pairs. We write ordered pairs with angle brackets, so the ordered pair whose first member is Regina and whose second is Søren is <r, s>. This pair is in the extension of “loves,” but <s, r> is not, so it had better turn out that <s, r> ≠ <r, s>. This illustrates a central aspect of order: when a ≠ b, <a, b> ≠ <b, a>; order alone suffices to distinguish ordered pairs. More generally, <a, b> = <c, d> if and only if a = c and b = d (while, by contrast, if a = d and b = c, then {a, b} = {c, d}). This principle articulates what Tarski in the 1920s will call a material adequacy condition, that is, a condition an account of something (in Tarski’s case truth, in ours, order) should meet to be adequate. In the 1910s, Norbert Wiener and Kazimierz Kuratowski each showed a way to explain the ordered pair in the primitive terms of set theory so as to satisfy the adequacy condition. (Quine said this work is a philosophical paradigm.)

We mostly follow Kuratowski, whose later account explains <a, b> as {{a}, {a, b}}. It would be a mistake to stare at this hoping for insight into order. Such insight as there is to be had was already articulated in the adequacy condition. Kuratowski’s account is adequate (as is Wiener’s different one) if it proves to satisfy the adequacy condition.

There is no enlightenment to be found in the proof that Kuratowski's account works, but students always ask to see a proof, so here goes. Suppose <a, b> = <c, d>. Then, by Kuratowski's account, {{a}, {a, b}} = {{c}, {c, d}}. Since {a} is a member of the set on the left, it is in the one on the right, so it is {c} or {c, d}. In the first case, a is c, while in the second both c and d are a. So in any case, a = c. Next we distinguish two cases. For the first, suppose a = b. Since {c, d} is in the second set, it is in the first, so it is {a} or {a, b}; if it is {a}, then d is a, which is b; while if it is {a, b}, then d is a or b, which are identical, so d is again b. Hence if a = b, b = d. So for our second case, suppose a ≠ b. If b = c, then since a = c, a = b, so since we're supposing a ≠ b, b ≠ c. Then b ∈ {c, d}, so since b ≠ c, b = d. Hence, in any case, b = d, as we were to show. This argument is a welter of unmemorable cases, so don't worry if your attention glazed over; what matters is that it works. Russell called Kuratowski's (and Wiener's) construction a trick.

Once we have ordered pairs, we may take an ordered triple <a, b, c> as <<a, b>, c>, an ordered pair whose first member is an ordered pair. An ordered quadruple <a, b, c, d> is <<a, b, c>, d>, and so on through all the ordered n-tuples. Then we may take the extension of an n-adic predicate to be a set of ordered n-tuples. We should work an example to fix ideas. The extension of the binary predicate "a is n years old at noon today" (where the blanks in the predicate are marked with the variables "a" and "n") is the set of ordered pairs <a, n> such that a is n years old at noon today. This one could also think of as (the noon today time slice of) the age relation. Let us focus on people: let P be the set of people (alive at noon today) and let N = {0, 1,…} be the set of all natural (i.e., non-negative, whole) numbers. The set of all ordered pairs <a, b> whose first member a is an element of P and whose second member b is an element of N is called the Cartesian or cross product of P and N. It is written P × N. It is called Cartesian in memory of rectangular Cartesian coordinates for the points on a Euclidian plane; it is called cross because if A has n members and B has k members, then A × B has n times k members (which hints at reconstructing arithmetic in set theory). A binary relation between people (alive at noon today) and natural numbers is any old subset of P × N. A relation between members of A and members of B is a subset of A × B. Age is a relation between people and numbers; age (at noon today) is

{<p, n> | p ∈ P and n ∈ N and p is n years old at noon today}, which is a subset of P × N. An n-ary relation among members of n sets $A_1, A_2, \ldots, A_n$ is a subset of $A_1 \times A_2 \times \ldots \times A_n$, the set of ordered n-tuples whose first member is in A_1, whose second is in $A_2, \ldots$, and whose n^{th} is in A_n.

Of course, a relation may hold between some members of a set like P and others. Such a relation is a subset of P × P. For example, parenthood (at noon today) is a subset of the set of all ordered pairs <a, b> of people (alive at noon today). We use exponential notation for cross products whose factors are identical: P^2 is P × P; P^3 is P × P × P; and so on. Let R be the parenthood relation just mentioned. Let W be the set of all women (alive at noon today). We might want to restrict the first members of a relation S to elements of a set A; we would write $S \upharpoonleft A$ for the set of pairs <x, y> in S such that $x \in A$. Then the motherhood relation is $R \upharpoonleft W$. To restrict the second members of pairs in S to A, we write $S \upharpoonright A$. Then the daughterhood relation is $R \upharpoonright W$. To restrict both to A, we write $S \uparrow A$, so $R \uparrow W$ is the mother–daughter relation.

Aristotle's syllogistic logic is geared for unary predicates. It had long been recognized that there are arguments whose conclusions clearly follow from their premisses but where syllogistic cannot certify these arguments because the arguments' success turns on polyadic predicates. Here is an example from Augustus De Morgan in the nineteenth century:

All horses are animals.
Hence, all heads of horses are heads of animals,

where the dyadic predicate "x is a head of y" is crucial. It was not until the nineteenth century that a systematic account of relations began.

In addition to De Morgan, Charles Sanders Peirce and Ernst Schröder were central in the articulation of relations. Notation like $R \uparrow W$ is just one fruit of their work. The fact that ordered pairs were worked out set theoretically pretty much at the end of the articulation of relations shows how hard it was to command a clear view of relations.

In the seventeenth century, Newton and Leibniz focused our attention on functions. The path of a particle in, for simplicity, the plane rather than space is a continuous curve, and using Cartesian coordinates the ordinates of points along the curve can often be given as mathematical functions of the abscissae. The speed of this particle at a point along its path will be given by the derivative of such a function, and conversely, the path is given by the integral of the particle's velocity; anyone who has done some calculus knows that differentiation and integration are the meat and potatoes of Newton's and Leibniz's calculus. At school we were all programmed in algorithms for computing the sums and products of natural numbers, and addition and multiplication are also functions. Such education inclines us to think of functions in terms of ways of calculating the output, or value, of a function at its inputs, or arguments. A somewhat less intentional image of a function pictures it as a bunch of arrows, one *from* each argument *to* the value of the function for that argument; the collection of its arguments is called the function's domain, while the collection in which its values lie is called the function's range, so on this picture a function is a collection of arrows arcing from its domain into its range.

As late as Kant at the end of the eighteenth century, curves were the leading image of functions. Through the nineteenth century, people worked out an extensional conception of a function. The calculus is infinitary, and the geometrical imagination trusted since Euclid began to go awry in the infinities of the calculus. Much of nineteenth-century mathematics was given over to a process called the arithmetization of analysis, which is what calculus grows up into. The aim of this process is to replace geometry, especially in analysis, with the mathematics of number, or later, set theory. An extensional conception of a function arises by starting from the picture of a bunch of arrows arcing from its domain to its range, and then discarding everything except the ordered pairs whose first members are the arguments and whose second are the values; only input and output remain, and we don't worry about how what goes in becomes what comes out.

We write $f : A \to B$ to mean that f is a function whose domain is a set A and whose range is a set B. But we have just seen that on the extensional conception this means that f is a set of ordered pairs whose first members lie in A and whose second lie in B; that is, it is a relation between members of A and B. There are two special conditions that such a relation must meet in order to be a function. First, for every member a of A, there is at least one member b of B such that $<a, b>$ is in f, that is, f relates a to b. Second, for each a in A, there is at most one b in B such that $<a, b>$ is in f. In the crochets of logic, the first condition is that

$$(\forall a)(a \in A \to (\exists b)(b \in B \wedge <a, b> \in f)),$$

while the second is that

$$(\forall a)(a \in A \to (\forall b)(\forall c)((b \in B \wedge c \in B \wedge <a, b> \in f \wedge <a, c> \in f) \to a = c)),$$

both of which use only notions built up by logic from primitives of set theory. Motherhood, for example, is not a function from women to people, since, first, some women have no children, while, second, others have several. But age (at noon today) is a function, since (at noon today) everyone is one and only one number of years old.

Once we have the extensional conception of a function, we revert from $<a, b> \in f$ to the customary notation $f(a) = b$; the existence and uniqueness conditions underpin the singular term f(a). Like relations, functions have polyadicities. Age is unary, while addition and multiplication are binary. Remember that a unary function is a binary relation (between the domain and the range). Addition and multiplication are said to commute (or be Abelian, for the nineteenth-century Norwegian mathematician Niels Abel) because $a + b = b + a$. But exponentiation does not commute, since $2^3 = 8$ but $3^2 = 9$. So the domain of a binary function should be a set of *ordered* pairs. So in general the domain of an n-ary function will be an n-fold Cartesian product $A_1 \times \ldots \times A_n$, and if the range is A_{n+1}, the function is a set of ordered (n+1)-tuples $<a_1,\ldots, a_n, a_{n+1}>$ such that $a_1 \in A_1,\ldots, a_n \in A_n$ and $a_{n+1} \in A_{n+1}$ (that meet the existence and uniqueness conditions). Remember that an n-ary function is an (n+1)-ary relation; addition, which is a binary function, is the extension of the ternary predicate "$x + y = z$" and so a ternary relation.

A relation is a function only if each argument bears the relation to a unique value; each of us has (at noon today) a unique age in years. But a number may be the age of many different people; there are lots of five-year-olds. To rule out functions sending different arguments to the same value we require the function to be one-to-one (or injective). This means that when $f : A \rightarrow B$ and a and b are different arguments in A to f, then f(a) is different from f(b). Each natural number n has a unique successor, so the relation of a number to its successor is a function; moreover, different numbers have different successors, so the successor function is one-to-one. When a function is one-to-one, it never happens that different arguments collapse into the same value, so there are at least as many things in the range as in the domain. This fails with the age function since it is not one-to-one; there are fewer than two hundred ages but billions of people.

A relation is a function only if for each argument there is a value to which the argument bears the relation; for each of us there is a number that is our age. But a number may be the age of no one; there are no 2,000-year-old men. To rule out functions sending no argument to some member of the range we require the function to be onto (or surjective). This means that when $f : A \to B$ and b is in B, then there is an a in A such that $f(a) = b$. The successor function is not onto, since 0 is a natural number that is not the successor of any natural number; among the integers positive, zero, and negative, however, successor is onto. When a function is onto, it never happens that a member of the range is missed out by the function, so there are at most as many things in the range as in the domain (and there could be fewer if the function were onto but not one-to-one). Our mortality prevents age from being onto. Every wife has a unique husband, but there are also bachelors, so the husband-of function from wives to men is not onto, and there are more men than wives.

A set A has as many members as a set B just in case there is a function that maps A one-to-one onto B. Such a function is sometimes called one–one correspondence, and $f : A \xrightarrow[\text{onto}]{1\text{–}1} B$ means that f is a one–one correspondence from A to B. Having seen one–one correspondences built up from scratch, you know that no notion of number was smuggled in surreptitiously; in particular, though a beginner might worry about the number one in one-to-one, you have seen it is about identity rather than number. Besides, it makes sense that a grasp of the as-many-as relation need not presuppose number, since it is clear that there are as many brains as spines (or as vertebrates) even when you are not clear on how many vertebrates there are.

It is an immense virtue of a number-free understanding of the as-many-as relation that it allows us to make sense of as-many-as questions about collections too big to count using 0, 1, 2, and so on. It allows us to compare infinite sets. It was not Cantor but Richard Dedekind who articulated the infinity of a set as its being in one–one correspondence with one of its proper subsets. (A proper subset of A is a subset of it to which some member of A does not belong.) Galileo observed the natural one–one correspondence between the natural numbers and their proper part, the even numbers. This looks like

$$\begin{aligned} 0 &\to 0 \\ 1 &\to 2 \\ 2 &\to 4 \end{aligned}$$

and so on; if you like arithmetical specifications, let $f(n) = 2n$. An axiom in Euclid says that the whole is always greater than any of its (proper) parts, and Galileo appealed to Euclid's authority to infer from the correspondence that there is no completed infinite totality of the natural (or even) numbers. Leibniz extended Galileo's reasoning to conclude that there are no infinite numbers either. Dedekind was inverting this reasoning. Like Cantor, he is committed to infinite sets (or systems, as he called them). So he concludes that Euclid's axiom is false, despite millennia of idolatry as *a priori*, and takes being the same size as a proper part as the hallmark of infinity. His courage inaugurates a grasp of the infinite per se.

A set with as many members as there are natural numbers is said to be countably (or sometimes denumerably) infinite. The even numbers are countable. (Some use "countable" to mean finite or countably infinite, so one needs to check how it is meant in a given text.) There are countably many integers. Interleave the integers and count them thus:

$$\begin{array}{r}0 \to 0\\ 1 \to 1\\ -1 \to 2\\ 2 \to 3\\ -2 \to 4\\ \vdots \to \vdots\end{array}$$

We don't have to, but we can arithmetize this picture by sending an even number 2n to n, and an odd number 2n + 1 to −n. Between any two rational numbers a and b there is a third, $\frac{(a+b)}{2}$; $\frac{3}{2}$ lies between 1 and 2. The rationals are thus said to be dense, and density can make one think there are more rationals than naturals. Cantor showed this is not so. Make a table of the non-negative ratios (we can put the negatives back in later by interleaving as we did with the integers). In the top row put the ratios with 0 in the numerator, in the next those with 1 in the numerator, and so on.

$$\begin{array}{cccc}\frac{0}{1} & \frac{0}{2} & \frac{0}{3} & \cdots\\ \frac{1}{1} & \frac{1}{2} & \frac{1}{3} & \cdots\\ \frac{2}{1} & \frac{2}{2} & \frac{2}{3} & \cdots\\ \vdots & \vdots & \vdots & \end{array}$$

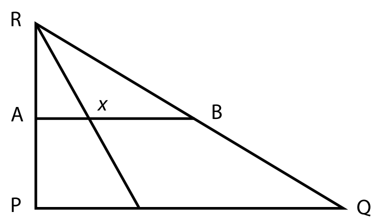
(We don't allow division by zero. This is not an arbitrary fiat among schoolteachers. The point is rather that $\frac{a}{b}$ should be the unique c such that a is b × c. But if a is positive and b is zero, there is no such c, while if a is zero, then any c will do. So division will have a unique value and be a function only if we don't divide by zero.) We next stitch the beads in this box together along the lines indicated by these arrows.

$$\begin{array}{ccccccc}
\frac{0}{1} & \rightarrow & \frac{0}{2} & & \frac{0}{3} & \rightarrow & \cdots \\
 & \swarrow & & \nearrow & & \swarrow & \\
\frac{1}{1} & & \frac{1}{2} & & \frac{1}{3} & & \cdots \\
\downarrow & \nearrow & & \swarrow & & & \\
\frac{2}{1} & & \frac{2}{2} & & \frac{2}{3} & & \cdots \\
 & \swarrow & & & & & \\
\vdots & & \vdots & & \vdots & &
\end{array}$$

Grab the thread at end $\frac{0}{1}$ and jerk it out straight and then count the beads as follows, eliminating duplicates of earlier beads.

$$\begin{array}{ccccc}
0 & 1 & 2 & \frac{1}{2} & \ldots \\
0 & 1 & 2 & 3 & \ldots
\end{array}$$

(We could arithmetize this correspondence too, but we need not.) A short line segment has as many points as a long. Put them parallel with their left ends on a perpendicular.



The line joining B and Q meets the perpendicular at a point R. If we map any point x on AB to the point where the line through R and x meet PQ, we have a one–one correspondence.

Cantor's great discovery is that some infinite sets are not the same in size. Each real number from zero up to but not including one has a decimal whose integer part is always zero, so we will ignore it. Suppose there were a one–one correspondence between these decimals and the naturals.

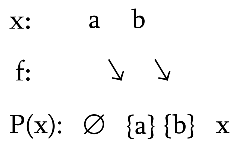
$$\begin{array}{cl}
0 \leftrightarrow & .a_1\ a_2\ a_3 \ldots \\
 & \quad\searrow \\
1 \leftrightarrow & .b_1\ b_2\ b_3 \ldots \\
 & \qquad\searrow \\
2 \leftrightarrow & .c_1\ c_2\ c_3 \ldots \\
\vdots & \qquad\vdots
\end{array}$$

where each of the as, bs, and so on is a digit, that is, among 0, 1,…, 9 (except that in order to have unique decimals we rule out those eventually constantly 9, so .250… is allowed but not .249…). We form the decimal for a real somewhere from 0 up to but not including 1 by going down the diagonal indicated by the arrows. Its first digit, α_1, is 2 if a_1 is 1, but 1 if a_1 is not 1. Its second digit, α_2, is 2 if b_2 is 1, but 1 if b_2 is not 1. And so on. The general idea is that if the natural k is assigned to the decimal $.n_1n_2n_3\ldots$, then α_{k+1} is 1 if n_{k+1} is 2, but 2 if n_{k+1} is 1. Let α be the real whose decimal is $.\alpha_1\alpha_2\alpha_3\ldots$. Then α differs from the real assigned to 0 in its first digit, from the real assigned to 1 in its second digit, and so on. This α is not a value of our supposed correspondence. We have shown that no function maps the naturals onto the reals from zero up to but not including one. This argument of Cantor's is called a diagonal argument after how we described α. Since then there have been many diagonal arguments, but Cantor's was the first.

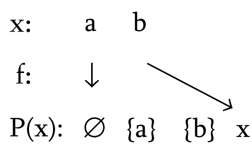
If any infinite set had as many members as any other, the arithmetic of infinity would be boring. Cantor's proof that there are more reals than naturals hints that it is not. But first we should explain "more." A set A has fewer members than or as many members as B just in case A has as many members as a subset of B, or equivalently, some function maps A one–one into B. (If you noticed that we sometimes omit "to" from "one-to-one," recall Emerson's apothegm that a foolish consistency is the hobgoblin of little minds.) Then B has more members than A just in case A has fewer members than or as many members as B, but it is not true that B has fewer members than or as many members as A. This is equivalent to requiring that A can be mapped one–one into B but A cannot be mapped onto B (or B cannot be mapped one–one into A).

Cantor proved that there are many more than two different infinite sizes. For any set x, the power set of x, written P(x), is the set of all subsets of x. Here we want to be careful about the distinction between members and subsets. So to illustrate, let x = {a, b} be a set with two members. The subsets of x are ∅ (since vacuously any member of ∅ is a member of x), the singletons of a and b, and x. So P(x) has four members. If x is a finite set with n members, then there are 2^n subsets of x, since for each of the n members there are 2 possibilities, in or out of the subset. It is an easy induction on natural numbers to show that 2^n is always greater than n ($2^0 = 1 > 0$, and if $2^n > n$, $2^{n+1} = 2 \times 2^n = 2^n + 2^n > n + 2^n \geq n + 1$; it is natural that anything to the zero power is one, for since $\frac{a^5}{a^3} = \frac{a \times a \times a \times a \times a}{a \times a \times a} = a \times a$, $\frac{a^n}{a^k} = a^{n-k}$, and then $a^0 = \frac{a^n}{a^n} = 1$).

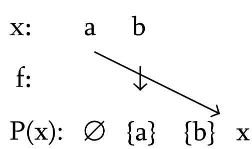
Cantor's theorem says that for any set x, its power set P(x) has more members than x. First we have to show that we can map x one–one into its power set. We do so by sending each member of x to its unit set: this is a function, and it is one-to-one because by extensionality distinct things have distinct singletons. (In our example we send a to {a} and b to {b}.) Next we show that no function maps x onto P(x). So suppose for reductio that f does map x onto P(x). Let y be the set of all members z of x such that $z \notin f(z)$. We show that y is not a value of f. But y is Cantor's big idea here, so we should pause to digest it. Suppose in our example f were



Then y is the set of things in x *not* members of what is at the end of their f arrows. In this case both a and b are, so y is ∅, and presto majesto, f does not hit ∅. Let's try again. This time



Here a is not in what's at the end of its f arrow, so a is in y, but b is in what's at the end of its f arrow, so b is not in y. Hence y is {a}, and again f does not hit {a}. Once more



Check that only b is in y, so y is {b}, and f does not hit {b}. Try other examples yourself. In general, suppose f did hit y, so for some z, $f(z) = y$. Either $z \in y$ or $z \notin y$. If $z \in y$, then since $y = f(z)$, $z \in f(z)$, so $z \notin y$ by how y is defined. If $z \notin y$, then since $y = f(z)$, $z \notin f(z)$, and then $z \in y$ again by how y is defined. So both possibilities are impossible. Hence there is no map f taking x onto P(x), so P(x) is bigger than x. (If the members of x were a, b, c,…, consider the array of conditions

$$\begin{array}{ccc}
a \in f(a) & a \in f(b) & a \in f(c) \;\ldots \\
b \in f(a) & b \in f(b) & b \in f(c) \;\ldots \\
c \in f(a) & c \in f(b) & c \in f(c) \;\ldots \\
\vdots & \vdots & \vdots
\end{array}$$

Go down the diagonal indicated by the arrows taking negations thus

$$a \notin f(a) \quad b \notin f(b) \quad c \notin f(c) \ldots$$

Then y is the set of members of x meeting these negative conditions. Cantor's proof that P(x) is bigger than x is also a diagonal argument.)

Now let A_0 be an infinite set, say the set of all natural numbers. Let A_1 be $P(A_0)$. Let A_2 be $P(A_1)$, and so on, so $A_{n+1} = P(A_n)$. Then for every n, A_n is an infinite set bigger than A_k for any k less than n. So we have as many infinite sets, no two of the same size, as there are natural numbers; we have an infinity of different infinite sizes. But there is more. Let A be the union of A_0, A_1, and so on. If A is not bigger than all of them, it has fewer members than or as many members as some A_n. (This should seem obvious, but it calls for an interesting argument.) Then it has fewer members than A_{n+1}. But A_{n+1} is one of the sets unioned to form A, so it is a subset of A and thus A_{n+1} is not bigger than A. So A is another infinite set, this one bigger than all of A_0, A_1, and so on. Don't stop now. Let B_0 be A, iterate power set through the naturals again, and again union all the B_n. Mixing power set and union in this way, a plethora of infinite sets, no two of the same size, emerges.

There is a natural correspondence between the reals from zero up to but not including one and the sets of positive integers. (The successor function from n to n + 1 is a one–one correspondence between the naturals and the positive integers, so there is a one–one correspondence between sets of naturals and sets of positive integers.) A decimal is fixed by a function whose arguments are positive integers and whose values are digits, that is, among 0, 1,..., 9, except that we exclude those f eventually constantly 9. Then for any real r from zero up to but not including one, there is a unique f of this sort such that

$$r = \sum_{n=1}^{\infty} \frac{f(n)}{10^n}.$$

But there is nothing magic here about 10 and the digits 0, 1,..., 9. Suppose we took 2 and the digits 0, 1. Then excluding those f eventually constantly one, for any real r from zero up to but not including one, there is a unique f of the new binary sort such that

$$r = \sum_{n=1}^{\infty} \frac{f(n)}{2^n}.$$

For any set S of positive integers, let c_s be the function (from positive integers to {0, 1}) such that for any positive integer k, $c_s(k) = 0$ if $k \in S$ but $c_s(k) = 1$ if $k \notin S$. In effect, taking 0 as truth and 1 as falsehood, $c_s(k) = 0$ says it is true that k is in S, while $c_s(k) = 1$ says it is false that k is in S; c_s is called the characteristic function of S. For any set S of positive integers, let

$$r = \sum_{n=1}^{\infty} \frac{c_s(n)}{2^n}.$$

This fixes a one–one correspondence between sets of positive integers and reals from zero up to but not including one; S goes to the binary real whose n^{th} digit says whether n is in S. This correspondence shows a way Cantor's theorem generalizes his proof that there are more reals than natural numbers.

Lines, planes, and space are continua. The idea is that a continuum has no gaps. An old tradition distinguishes so firmly between the continuous and the discrete that, while of course accepting that a point may lie on a line, it insists that a continuum is not granular and so does not consist of points. But Descartes founded analytic geometry on the assumption that there is a one–one order-preserving correspondence between the points on a line and the real numbers. "Order-preserving" means that if f is the correspondence and point p on the line lies to the left of point q, then the number f(p) is less than the number f(q); the left-to-right relation is reflected in the less-to-greater relation. Actually, it is perhaps an anachronism to state Descartes's assumption in terms of the real numbers. If pressed, Descartes might have explained his coordinates geometrically as lengths (and so numbers) of segments. The real numbers were not isolated from geometry until the arithmetization of analysis during the nineteenth century. But once the real numbers were liberated from geometry, good conscience recognized analytic geometry as centered in the assumed correspondence just mentioned. The range of the correspondence is the set of points on the line, and since the nineteenth century we have ceased to see much difference between a line and the set of points on it. It has become customary, especially in discussions focused on size, to call the reals and the power set of the natural numbers continua too. (The function that sends x to $\frac{x}{(x-1)}$ shows that there are as many non-negative reals as reals from zero up to but not including one.)

Cantor proved that there are more reals than natural numbers. This leaves open whether there is a set with more members than there are natural numbers but fewer members than there are reals; is there a size between those of the naturals and the reals? Cantor conjectured that there is not; this has come to be called the continuum hypothesis. There is also a generalized continuum hypothesis conjecturing that for no infinite set A is there a set larger than A but smaller than P(A). In 1900 the German mathematician David Hilbert, addressing the powers of mathematics that then were, put the continuum hypothesis at the head of the agenda for twentieth-century mathematics. In the 1930s, Kurt Gödel proved that if prevailing set theories are consistent, the continuum hypothesis is not refutable in them, and in the 1960s, Paul Cohen proved that if prevailing set theories are consistent, neither is the continuum hypothesis provable from them. Gerald Sacks called the generalized continuum hypothesis a rock of undecidability. Hugh Woodin has an elaborate program for winkling out plausible axioms to settle (probably by refuting) the continuum hypothesis.

So far we have talked about fewer, as many as, and more, relations that suggest number, but we have not considered how to get from such relations to numbers. Cantor himself thought that we abstract a number as something common to all and only those sets any one of which has as many members as any other. The notion of such abstracting has a very long history, but that does not make it clear. Sometimes it sounds perceptual, as if we saw a single thing common to all these sets, but it feels silly to look for the number five both in the digits on one's left hand and in the digits on one's right. Sometimes it sounds more intellectual, as if we infer to the existence of a single thing common to all these sets. Then one wants to see the logic of the inference.

The irony is that set theory lays out this logic (and Frege, unlike Cantor, uses it). The as-many-as relation is called an equivalence relation, which means that it is reflexive, symmetric, and transitive. So next we explain these terms. Recall that a (binary) relation is a set of ordered pairs. Let R be a binary relation. The field of R is the set of all those things that either bear R to something or to which something bears R, so the field of the parent-of relation is the set of living things, while the field of the gravitational attraction relation is all massive bodies. To get the idea of an equivalence relation, keep in mind examples like as-tall-as and as-hot-as. (In Welsh, in addition to the three standard degrees absolute [hot], comparative [hotter], and superlative [hottest], there is a fourth [as hot as].) A relation R is reflexive just in case every member of its field bears R to itself; each of us is as old as himself. R is symmetric just in case for any a and b in its field, when a bears R to b, b bears R to a; if a is as old as b, b is as old as a. R is transitive just in case for any a, b, and c in its field, when a bears R to b and b bears R to c, then a bears R to c. It should be clear that our as-many-as relation has all three properties and so is an equivalence relation.

Now suppose A is a set. By a partition of A we mean a collection P of subsets of A with two features: first, for any member a of A there is a set X in P such that a is in X; and, second, different members of P are disjoint. You might liken a partition to a tiling of A that covers it completely (first feature) without any tiles overlapping (second feature). Let A be a set and let R be an equivalence relation whose field is A. For example, A might be the set of people, and R, the as-old-as relation. We will see ages as members of a partition of A reflecting R. For any member a of A, let [a] be the set of all b in A such that a bears R to b. Then [a] is called the equivalence class of a. If several equivalence relations with field A are in play, we say [a] is the equivalence class of a under R and sometimes write $[a]_R$ to make the connection with R explicit. Let P be the set of all these equivalence classes, one for each a in A. We will show that P is a partition of A. To establish the first feature, note that since R is reflexive, each a in A bears R to itself, so a is always in [a], which is a member of P. The reflexivity of an equivalence relation reappears as the first feature (full coverage) of its corresponding partition. Next, suppose that [a] and [b] are different members of P. Since they are sets, this means that some c in A is in one but not the other. That can happen in two ways; we will look at one and you can do the other to fix ideas. So suppose c is in [a] but not in [b]. By how we defined equivalence classes, this means that

(1) a bears R to c

and

(2) b does not bear R to c.

We have to show that [a] and [b] are disjoint, that is, no d in A is in both. So suppose for reductio that some d is. Then

(3) a bears R to d

and

(4) b bears R to d.

Since R is symmetric, it follows from (3) that

(5) d bears R to a.

Then since R is transitive, it follows from (4) and (5) that

(6) b bears R to a

and from (6) and (1) that

(7) b bears R to c.

Since (7) contradicts (2), we have reached an absurdum. Symmetry and transitivity in the equivalence relation reappear as the second feature (nonoverlap) of its corresponding partition. An equivalence relation is said to induce the corresponding partition of its field. Conversely, if we start from a partition P of a set A and define a relation R with field A such that for a and b in A, a bears R to b just in case a and b are both in some member of the partition, then we can show R is an equivalence relation. Equivalence relations and partitions are different ways to package the same information.

In *Our Knowledge of the External World*, Russell urges a principle he says could as well be called the principle that dispenses with abstraction. The idea is that we give up metaphysical worries about what, say, the age five-years-old might be and instead go with the set of all people as old as any given five-year-old. He sometimes calls this replacing inferring to entities by logical construction, but that is an excess of enthusiasm; sets are entities to whose existence we infer. But equivalence classes give us a uniform and articulate setting for ages, weights, heights, and so on. Frege thought they would also do for numbers (though our setting is not his). Frege and Russell founded a constructional tradition in philosophy that includes Carnap, Goodman, David Lewis, and others. But the core, and perhaps sum total, of that tradition is that one object the bearers of an equivalence relation have in common is their equivalence class.

When A has as many members as B, what they have in common is their number of members. So (roughly) Frege and Russell took the number of members of A as the set of all B with as many members as A.[1] (This is rough because Frege did not think of himself as working with sets and because Russell layered his sets, so his numbers are not unique.) An object is a number just in case for some A it is the number of members of A. Then 0 is the number of members of $\varnothing$, 1 is the number of members of {0}, and 2 is the number of members of {0, 1}. A number n is less than or equal to a number k, $n \le k$, just in case there are A and B such that n is the number of As, k is the number of Bs, and there are fewer As than Bs or as many As as Bs. Define a function s so that when n is a number, s(n) is the number of numbers $k \le n$. If we follow Frege and Russell and focus on familiar numbers beginning with the natural numbers, we may isolate them as the members of all sets A such that $0 \in A$ and $s(n) \in A$ when $n \in A$. Among the naturals, s(n) is the number of members of {0, 1,…, n} and is, as one expects, n + 1. But if n is the number of members of an infinite set, then s(n) = n; Cantor's generalized continuum hypothesis conjectures that the power set of an infinite set is of the next, or successor, infinite size.

These numbers are cardinals. As Frege puts it, a cardinal number is an answer to the question how many. There are (or would be if Greek mythology were true) twelve Muses. This does not depend on the Muses being arranged in some special order; there are twelve Muses even if Erato (lyric poetry) has no priority over Clio (history) nor Clio over Erato. But if we are going to count the Muses, we pick a first Muse to match with "one," a second to match with "two," and so on up to "twelve." When we count Muses, either we pick Erato before Clio or we pick Clio before Erato, and while we can do so without favoritism, we must pick one first. The terms "first," "second," "third," and so on reflect positions in an ordering; this is the central idea behind ordinal (rather than cardinal) numbers.

It is set theory, and more specifically the theory of relations, that articulates order. Let R be a binary relation with field A. We say that R

[1] In *Labyrinth of Thought*, 2nd ed. (Basel: Birkhäuser, 2007), 87, José Ferreirós says that by the 1850s, Dedekind was using equivalence classes, and that Gauss and Riemann had done so earlier. But when Dedekind came to define the natural numbers (*Essays on the Theory of Numbers* [New York: Dover, 1963], 68, article 73), he did so in terms of abstraction, like Cantor, not in terms of equivalence classes, like Frege.

partially orders A just in case R is irreflexive and transitive. Irreflexivity means that nothing in A bears R to itself. The ancestor-of relation partially orders a family tree, and the order is partial because siblings (and spouses) are incomparable in the ordering. The proper-subset-of relation partially orders a set's power set, and the order is partial because when the set is {a, b}, its subsets {a} and {b} are incomparable in the ordering. A partial ordering is total just in case any two members of its field are comparable, that is, either a is R to b, or b is R to a, or a is b. (These alternatives are exclusive.) So the less-than relation totally orders the natural numbers, and with larger fields "it" totally orders the integers, rationals, and reals. (The "it" is in scare quotes here because each of these relations is from the extensional point of view an extension of, and so different from, the orders mentioned earlier.)

There are no infinite descending sequences of natural numbers, for if the sequence starts with n, then since there are only n natural numbers less than n, there aren't enough naturals to fill out an infinite sequence descending from n. Another way to put this is that any nonempty set of natural numbers has a least member. A total order well-orders its field just in case any nonempty subset B of its field has an R-least member, that is, there is a b in B such that for any a in B different from b, b bears R to a. So less-than well-orders the natural numbers. But it does not well-order the integers, since there is no least negative number, nor does it well-order the rationals or the reals, since there is neither a least positive rational nor a least positive real.

Let R be a binary relation with field A and let S be a binary relation with field B. Let f be a one–one correspondence mapping A onto B. Then we say f preserves R (in S) just in case for any x and y in A, x bears R to y just in case f(x) bears S to f(y). When we count the Muses, we lay out an ordering of the Muses and a correspondence of them with the numerals from "one" to "twelve" that preserves the usual ordering of the numerals in our ordering of the Muses, and the ancestor-of relation between family members is preserved between their names in the family tree. Relations are said to be similar when there is a one–one correspondence between their fields that preserves one relation in the other. Similarity is an equivalence relation, so it partitions (binary) relations into equivalence classes, and the equivalence class of a relation under similarity is called its type.

Cantor named the types of some total orders. He used ω for the type of less-than on the natural numbers. He used $^*\omega$ for the type of less-than on the negative integers. We can add types of the total orders by concatenating an order of the first type with a copy of the second chosen so they have disjoint fields. So he used $^*\omega + \omega$ for the type of less-than on the integers. An order of type $^*\omega + \omega$ has no end points, while an order of type $\omega + {}^*\omega$ has a terminus at either end, so such an addition does not commute. Cantor used η for the order type of less-than on the rationals, a dense total order without end points of a countable set. The names $^*\omega$ and η are no longer much used in mathematics, but a philosopher might do well to remember them.

An ordinal number is the order type of a well-ordering. So ω, the type of less-than on the naturals, is an ordinal number because less-than well-orders the naturals. It is the smallest infinite – or, as Cantor put it, transfinite – ordinal number. But "smallest" makes sense here only if we have an ordering of the ordinal numbers. We say that a well (or even just total) order R with field A is an initial segment of a well (or total) order S with field B just in case A is a subset of B, R is S restricted to A (which we called $S \upharpoonright A$ earlier) and when $a \in A$ and bSa for some $b \in B$, then $b \in A$ (so R omits nothing S-below anything in its field; R has no S-gaps). Your pre-Revolutionary ancestors are an initial segment of your ancestors, and the naturals less than 100 are an initial segment of the naturals, though the even numbers are not. R with field A is a proper initial segment of S with field B just in case it is an initial segment and A is a proper subset of B. An ordinal number α is less than an ordinal number β just in case there is a well-order R of type α with field A, a well-order S of type β with field B, and R with A is similar to a proper initial segment of S with B.

The less-than relation, <, well-orders the ordinals. Cantor's arguments here are a bit sketchy, so we will fill in some blanks. We do so using principles that come with well-orderings. Let R be a well-order with field A. Suppose that whenever all b in A such that b is R to a are in a set C, then a is in C too; then all members of A are in C. For otherwise let a be the R-least member of A not in C; then everything in A R-below a is in C, so by the supposition a is in C after all. This is proof by R-induction. We can also explain expressions by R-induction. Suppose that we state conditions on whether a member a of A, for example, satisfies a predicate in terms of how satisfaction of the predicate distributes across a's R-predecessors (if a has no R-predecessors, we state outright whether a satisfies the predicate); then we have satisfaction conditions for the predicate over all of A. For otherwise let a be the R-least member of A on which the predicate is unsettled; then it is settled on everything R-below a, so by the supposition it is settled on a after all.

Less-than partially orders the ordinals. Suppose $\alpha < \beta$ and $\beta < \gamma$. Then there is a well-order R with field A of type α, a well-order S with field B of type β, and a well-order T with field C of type γ, and there are functions f and g such that f shows R similar to a proper initial segment of S and g shows S similar to a proper initial segment of T. For any a in A, let $h(a) = g(f(a))$; h is called the composition of g on f. Then h shows R similar to a proper initial segment of T, so $\alpha < \gamma$. Next we show that < is irreflexive. For any set A, let i_A be the function whose value on any a in A is a. We show that for any well-order R with field A, i_A is the only function that shows R similar to an initial segment of R. For any member a of A, let A_1 be the set of members of A R-below a, let R_1 be $R \upharpoonright A_1$, and suppose that i_{A_1} is the only function that shows R_1 similar to an initial segment of R_1. To extend i_{A_1} to a similarity with domain $A_1 \cup \{a\}$, we have to assign to a a member of $A - A_1$, and to avoid gaps, the only choice is a itself. So if A_2 is $A_1 \cup \{a\}$ and R_2 is $R \upharpoonright A_2$, then i_{A_2} is the only function that shows R_2 similar to an initial segment of R_2. Our claim follows by R-induction. If $\alpha < \alpha$, a well-order of type α is similar to a proper initial segment of itself, and this is ruled out by the claim just established. We can show that only one function can show a well-order similar to an initial segment of another. It is perhaps worth noting that while an infinite set is always the same size as many of its proper subsets, the only initial segment of a well-order to which it is similar is itself, which is not a proper initial segment.

To show that less-than totally orders the ordinals, we show that $\alpha < \beta$, or $\alpha = \beta$, or $\beta < \alpha$. Let R be a well-order with field A of type α and let S be a well-order with field B of type β. To get started, suppose first that A is empty. Then the empty set shows A similar to the empty initial segment of S. So if B is empty, $\alpha = \beta$, while if B is not empty, $\alpha < \beta$. Likewise, if B is empty, $\alpha = \beta$. So if either A or B is empty, α and β are comparable. If neither A nor B is empty, suppose we have a binary relation $F \subseteq A \times B$ that is a similarity between an initial segment R_1 with field A_1 of R and an initial segment S_1 with field B_1 of S. If $A_1 = A$ and B_1 is a proper subset of B, then $\alpha < \beta$. If A_1 is a proper subset of A and $B_1 = B$, then $\beta < \alpha$. Otherwise A_1 is a proper subset of A, and B_1, of B. Let a be the R-least member of $A - A_1$, and b, of $B - B_1$. Then $F \cup \{\langle a, b \rangle\}$ is a similarity between the initial segment of R with field $A \cup \{a_1\}$ and the initial segment of S with field $B \cup \{b_1\}$. We thus define by R and S inductions a relation F that is either a similarity between R and S, so $\alpha = \beta$, or between R and a proper initial segment of S, so $\alpha < \beta$, or between S and a proper initial segment of R, so $\beta < \alpha$. (Showing cardinal numbers comparable is more demanding.)

To show that less-than well-orders the ordinals, let A be a nonempty set of ordinals. We must show there is a least ordinal in A. If not, there is a sequence $\alpha_1, \alpha_2, \ldots$ of ordinals in A such that for all n, $\alpha_n > \alpha_{n+1}$. Then for each n there is a well-order R_n with field A_n of type α_n such that R_{n+1} is similar to a proper initial segment of R_n. Taking compositions of these similarities, it follows that for each n there is a well-order R_n with field A_n such that for all n, R_{n+1} is a proper initial segment of R_n and of R_1 where A_{n+1} are nonempty. For each n, let a_n be the R_1-least member of $A_n - A_{n+1}$. Then the set $\{a_1, a_2, \ldots\}$ of all these a_n is a nonempty subset of A_1 with no R_1-least member, which is impossible since R_1 well-orders A_1.

Every ordinal is the type of an initial segment of the ordinals. In particular, α is the type of $< \restriction \{\beta \mid \beta < \alpha\}$. (If R is a relation, the field of $R \restriction A$ is always A, so we need not mention the field of $R \restriction A$ separately. Indeed, since the field of a relation is always uniquely determined by the relation, we never need mention a relation's field separately. But it is often easier to follow what is going on if we do so.) For any well-order R with field A we may sort members of A into three kinds. A member of A may have no R-predecessors; there are never two members of A of this kind. Otherwise a member a of A has R-predecessors. One sort of this kind has an immediately predecessor, so there is a b that is R to a but there is no c such that b is R to c and c is R to a. Another sort has predecessors, but no immediate predecessor, so there is a b that is R to a and for any b R to a there is a c such that b is R to c and c is R to a. Members of A of the first kind are R-successors, and members of A of the second kind are R-limits. It follows by R-induction that if a member of A with no R-predecessors is in a set C, if the R-successor of a member of C is in C, and if R-limits are in C when all their R-predecessors are in C, then all members of A are in C. Among ordinals, 0 is the type of $< \restriction \varnothing$. Since $< \restriction \varnothing$ has no proper initial segments, there is no ordinal $\alpha < 0$. So $\{\alpha \mid \alpha < 0\} = \varnothing$, and thus 0 is the type of $< \restriction \{\beta \mid \beta < 0\}$, as expected. For any ordinal α let $s(\alpha)$ be the type of $< \restriction \{\beta \mid \beta \leq \alpha\}$. Suppose α is the type of $< \restriction \{\beta \mid \beta < \alpha\}$. The relation $< \restriction \{\beta \mid \beta < \alpha\}$ is a proper initial segment of $< \restriction \{\beta \mid \beta \leq \alpha\}$. Thus $\alpha < s(\alpha)$, which separates s on cardinals from *s* on ordinals. No proper initial segment of $< \restriction \{\beta \mid \beta \leq \alpha\}$ properly extends $< \restriction \{\beta \mid \beta < \alpha\}$. Thus α is the immediate predecessor of $s(\alpha)$, and $s(\alpha)$ is the successor of α. Moreover, $s(\alpha)$ is the type of $< \restriction \{\beta \mid \beta < s(\alpha)\}$, as expected. Now suppose λ is a limit ordinal and that for all $b < \lambda$, β

is the type of $< \restriction \{\gamma \mid \gamma < \beta\}$. $< \restriction \{\beta \mid \beta < \gamma\}$ properly extends $< \restriction \{\gamma \mid \gamma < \beta\}$ for all $\beta < \lambda$, but no proper initial segment of $< \restriction \{\beta \mid \beta < \lambda\}$ properly extends $< \restriction \{\gamma \mid \gamma < \beta\}$ for all $\beta < \lambda$. So the type of $< \restriction \{\beta \mid \beta < \lambda\}$ is the least type greater than the types of $< \restriction \{\gamma \mid \gamma < \beta\}$ for all $\beta < \lambda$, and since each of these types is β, the type of $< \restriction \{\beta \mid \beta < \lambda\}$ is λ, as expected. So, as we said, for any α, α is the type of $< \restriction \{\beta \mid \beta < \alpha\}$. In particular, each ordinal is greater than every member of the initial segment of the ordinals of which it is the type.

The finite ordinals are the members of all sets A such that the ordinal $0 \in A$ and $s(\alpha) \in A$ when $\alpha \in A$. Let F be the set of finite ordinals. Then ω, the least transfinite ordinal, is the type of $< \restriction F$, and the transfinite ordinals are the ordinals $\alpha \geq \omega$. Let N be the set of natural numbers. Set $f(\omega) = N$. When $s(\alpha)$ is transfinite, let $f(s(\alpha))$ be $P(f(\alpha))$ and when λ is a limit, let $f(\lambda)$ be the union of the $f(\beta)$ for $\beta < \lambda$. Then when α is transfinite, $f(\alpha)$ is an infinite set, and when $\alpha < \beta$, $f(\beta)$ is larger than $f(\alpha)$. For each transfinite α, let c_α be the cardinal number of $f(\alpha)$. Since we can map the transfinite ordinals one–one into the infinite cardinals, there are at least as many infinite cardinals and transfinite ordinals. To show the converse, we use a new principle, the axiom of choice. This says that if A is a set of nonempty sets, there is a function f with domain A such that for any x in A, $f(x) \in x$. The idea is that f picks from each x in A a member of x; f is called a choice function for A. When A is finite, we can prove it has a choice function, so we need a new axiom only when A is infinite, and then only when the members of A are opaque to us. Russell makes this point by saying that when A is an infinite set of pairs of shoes, we can pick the left one uniformly, but when A is an infinite set of pairs of socks, we need the axiom of choice.[2] It follows from the axiom of choice that every set can be well-ordered, that is, for any set A there is a well-order whose field is A. Let P be the set of nonempty subsets of A, and let f be a choice function for P. Suppose we have a one–one function g whose domain is an initial segment S of the ordinals and whose values lie in A, and let B be the range of g on S. If B is a proper subset of A, let $g' = g \cup \{<\alpha, b>\}$ where α is the type of $< \restriction S$ and $b = f(A - B)$, while if $B = A$, let $g' = g$. Then g' is a one–one function

[2] Russell's less felicitous version occurs in "On Some Difficulties in the Theory of Transfinite Numbers and Order Types," reprinted in his *Essays in Analysis*, ed. Douglas Lackey (New York: George Braziller, 1973), 157–58. This essay was first published in 1905. His more felicitous version (using socks) occurs in his *Introduction to Mathematical Philosophy* (London: George Allen & Unwin, 1919), 126.

whose domain is the initial segment $A \cup \{\alpha\}$ of the ordinals and whose values lie in A. By induction there is a one–one g from an initial segment s of the ordinals and whose range is A. Since g maps S one–one onto A, $h = \{\langle a, \alpha\rangle \mid g(\alpha) = a\}$ maps A one–one onto S. Let R be $\{\langle a, b\rangle \mid a, b \in A, \text{ and } h(a) < h(b)\}$. Then R well-orders A. (To go the other way, let A be a set of nonempty sets, and let U be the union of the sets in A. If R well-orders U, for any $x \in A$ let f(x) be the R-least member of x. So if any set can be well-ordered, the axiom of choice follows.) For any infinite cardinal c, let S be a set in c (and so of size c), let R be a well-order of S, and let α_c be the type of R. If c is less than k, a well-order of a set of size c is similar to a proper initial segment of a well-order of a set of size k, so $\alpha_c < \alpha_k$. So we can map the infinite cardinals one–one into the transfinite ordinals. Hence there are as many infinite cardinals as transfinite ordinals.

We can also that show there are at least as many ordinals as cardinals without choice. For any cardinal c, let L_c be the set of ordinals with fewer than c predecessors. Then L_c is a segment, and if we let α_c be the type of $< \upharpoonright L_c$, then sending c to α_c maps the cardinals one–one into the ordinals. But it was as well to have been introduced to choice, and, as we shall see, we have already used it elsewhere.

Let us also sketch maps between cardinals and ordinals that will have later reflections. For any cardinal c, let L_c be the set of ordinals with fewer than c predecessors. L_c is a segment. If we let α_c be the type of $< \upharpoonright L_c$, then α_c is the least ordinal with c predecessors, and sending c to α_c maps the cardinals one–one into the ordinals. To go the other way, first let S_c be the set of ordinals with c or fewer predecessors. Then S_c is a segment, and if we let ν_c be the type of $< \upharpoonright S_c$, then ν_c is the least ordinal with more than c predecessors. So if we let $\sigma(c)$ be the cardinal of the set of predecessors of ν_c, then $\sigma(c)$ is the least cardinal greater than c. (So the generalized continuum hypothesis says that if c is infinite, the power set of L_c has $\sigma(c)$ members.) Let z be the cardinal of $\varnothing$ and let ζ be the type of $< \upharpoonright \varnothing$. Let $f(\zeta) = z$, and $f(s(\alpha)) = \sigma(f(\alpha))$; when λ is a limit, for each $\beta < \lambda$, let A_β be the segment of ordinals less than $\alpha_{f(\beta)}$, and let $f(\lambda)$ be the cardinal of $\cup_{\beta<\lambda} A_\beta$. Then f maps the ordinals one–one into the cardinals. John von Neumann used α_c to rethink the cardinals in the aftermath of the paradoxes.

When a function maps a set A one–one into a set B and another function maps B one–one into A, there is a function that maps A one–one onto B. This claim is sometimes called the Cantor–Bernstein–Schroeder theorem. All its proofs are interesting, so it is not obvious or trivial. We use the axiom of choice to show that cardinal numbers are comparable, that is, that for any cardinals c and k, c is less than k or c is k or k is less than c. To do so, let A be of size c and B be of size k. Use choice to get an R that well-orders A and an S that well-orders B. Our proof that ordinals are comparable yields a relation $F \subseteq A \times B$ that is either a similarity of R to an initial segment of S, so c is less than or equal to k, or a similarity of S to an initial segment of R, so k is less than or equal to c. This shows c and k comparable. We have often assumed cardinals comparable without saying so. The first time was when we said that if the union of the set of naturals, its power set, and so on is not larger than all these sets, it is smaller than or the same size as one of them. It follows from the comparability of cardinals that any set can be well-ordered, which yields the axiom of choice. For any set x let h(x) be the set of ordinals α such that $\{\beta \mid \beta < \alpha\}$ can be mapped one–one into x. Then h(x) is a segment of ordinals. Let α be the type of $< \restriction h(x)$. Since α is greater than all members of h(x), $\alpha \notin h(x)$. So by comparability there is an f that maps x one–one into $\{\beta \mid \beta < \alpha\}$. Let R be $\{\langle a, b\rangle \mid a, b \in x$, and $f(a) < f(b)\}$. Then R well-orders x.

David Hilbert called Cantor's realm of infinite cardinals and transfinite ordinals a paradise. Such praise might have made Cantor nervous (though he was seven years dead when Hilbert gave it). Cantor distinguished three kinds of infinity: potential, actual, and absolute. By the potential infinite Cantor meant variables. He wrote long before Frege's scrupulous syntax sank in, and by variables he meant things that vary, like lengths. But, he said, variables always have ranges of variation (like non-negative reals for lengths), such ranges are sets, and an infinite set is an actual infinity; the potential infinite presupposes the actual. Another way to imagine a potential infinity might be a sequence $u_1, u_2, \ldots$ of possible worlds each accessible to all its successors, all containing only finitely many inhabitants, but each containing more inhabitants than its predecessors; then no possible world is infinite, but however many things there might be, there could be yet more. Cantor would grant all this, and observe that it requires an actual infinity of possible worlds. What we might call Cantor's thesis, that there won't be a potential infinity of any sort unless there is also an actual infinity of some sort, deserves more attention than it has received. Cantor said that only God is absolutely infinite. He was anxious that his set theory not poach on absolute preserves, and many of the theologians he cited on the infinity of divinity were Catholic (like Origen, Augustine, Aquinas, and Frs. Magnon, Saguens, and Libertore). For Hilbert to call it a paradise might have seemed to threaten his theological discretion. But most philosophers find the infinite so irresistible that they are likely to agree with Hilbert.

But there was a serpent in Eden, and there were three in Cantor's paradise. Three paradoxes arise in Cantor's naïve set theory. The first to be published appeared in 1897 and is due to Cesare Burali-Forti. Let On be the set of ordinals guaranteed by comprehension as the extension of "is an ordinal." Then $<$ well-orders On, so let Ω be the type of $< \upharpoonright$ On. On the one hand, since Ω is the type of the segment On, Ω is greater than all ordinals in On, but on the other hand, Ω is an ordinal, so it is a member of On. Then $\Omega < \Omega$, which is impossible since $<$ is irreflexive. The second was first described in a letter Cantor wrote to Dedekind in 1895. Let U be the universe, the set of everything, guaranteed by comprehension as the extension of "is self-identical." On the one hand, by Cantor's theorem, U is strictly smaller than its power set P(U), but on the other hand, everything in P(U) is self-identical and so a member of U, and then P(U) is smaller than or the same size as U. The third struck Russell while thinking about diagonal arguments, and he sent Frege a letter describing it in 1902. Some sets, like the set of all lions, are not members of themselves (since it is not a lion), while others, like the set of all sets, are members of themselves. Let R be the set of all sets of the first sort, that is, not members of themselves. R is the extension of "$x \notin x$" guaranteed by comprehension. Then any set is a member of R just in case it is not a member of itself. So R is a member of R just in case R is not a member of R, which is a contradiction.

When Hilbert spoke of Cantor's paradise, what he said was that no one shall be able to drive us from the paradise that Cantor created for us. That trope might well have made Cantor very nervous of Rome's judgment. But Hilbert's point was that Cantor's theory of sets had for the first time made systematic sense of the infinite per se, so even if in its naïve state that theory was beset by paradox, still, what it reveals is such a joy that we will insist on sophisticating it to avoid paradox while retaining its insight into the infinite. But if naïveté is often unreflective, sophistication deliberates.