

A beginner's guide to countable ordinals

Timothy Gowers

Contents

Prerequisites	1
Contents	1
Three problems	2
Problem 1. Showing that a continuous function on $[0, 1]$ is bounded	2
Problem 2. Generating the Borel sets	4
Problem 3. Showing that open games are determined	6
Well-ordered sets	8
Definition of well-ordered sets	8
Maps between well-ordered sets	8
Order-types of well-ordered sets	9
How to use ordinals	10
Proof that there is an uncountable well-ordered set	12
How not to use ordinals	13
Ordinals and induction	13
Solving problems 1 and 3 without ordinals.	14
A very brief sketch of how to “construct” a game that is not determined	15
How can infinity plus one not be infinity?	16

Prerequisites

It would help to know basic real analysis, up to the definition of open and closed sets. But even without this, it might be possible to understand the general discussion and at least one of the examples.

Contents

When one meets ordinals for the first time in a university course on set theory, they can seem quite forbidding. In particular, the notion of a transitive set (a set X such that every element of X is also a subset of X) takes a little getting used to, and seems rather artificial. In fact, it *is* rather artificial – one would like to define an ordinal to be an equivalence class of well-ordered sets (if you don't know what I mean by this, then you can come back to these preliminary remarks after reading the explanations below) but this creates difficulties of a the-set-of-all-sets kind, and is therefore not allowed. Instead one is forced to

come up with a way of constructing a well-ordered set of each order-type, and this can be done in various ways.

The aims of this page are to show

- that it is possible to *use* ordinals without knowing how they are constructed (rather as one can do real analysis without knowing how real numbers are constructed);
- that the countable ordinals, which are sufficient for many applications, can be constructed rigorously in a way that requires more or less no knowledge of set theory;
- that ordinals crop up naturally in many contexts, so that it is worth understanding them, at least in the naive way outlined below.

Three problems

The idea of this section is to show how ordinals arise naturally in three contexts. Quite a bit of the discussion, particularly with Problems 2 and 3, is about the contexts themselves rather than about ordinals. I have therefore given various opportunities for the reader to jump to the end of the section and start reading the more general theory. The first one is now.

Problem 1. Showing that a continuous function on $[0, 1]$ is bounded

I have discussed this problem in detail on a different page. Here I shall be a little briefer about the real analysis but much more detailed about the aspect of the problem that relates to ordinals.

If f is continuous on $[0, 1]$ and bounded on $[0, t]$, then, by continuity at t , one can find $s > t$ such that f is bounded on the interval $[0, s]$. Let us call this the **basic lemma**. Using the basic lemma repeatedly, one can build a sequence of numbers $0 < t_1 < t_2 < t_3 < \dots$ in such a way that f is bounded on the interval $[0, t_n]$ for every n (though so far we do not know whether f is bounded on the union of all those intervals).

Since the t_n all live in $[0, 1]$, they form a monotonic sequence which is bounded above. Hence, they converge to some limit, which, with the considerable benefit of hindsight, I shall call t_ω . By continuity, there is some interval around t_ω on which f is bounded. Since this interval contains t_n for all sufficiently large n , we can deduce that f is bounded all the way to t_ω . Thus, f *was* bounded on the union of the intervals $[0, t_n]$, though what we have now proved is very slightly stronger as it includes the end point t_ω .

Now we can continue the sequence “beyond infinity” by using the basic lemma repeatedly, starting at t_ω . This gives us a second sequence $t_\omega < t_{\omega+1} < t_{\omega+2} < \dots$ such that f is bounded on the interval $[0, t_{\omega+n}]$ for every n .

It may look as though there is something not quite rigorous going on here. If you think of ω as being infinity (as it is quite reasonable to do) then is it legitimate

to start adding $1, 2, 3, \dots$ to it? Doesn't infinity plus one equal infinity?

I shall return to this question, but for the time being let me say only this: there is nothing to stop me using expressions like $\omega + 17$ if all I do is regard them as *notation*. Instead of writing $t_{\omega+n}$ I could have written u_n . Or I could have written t_{0n} for t_n and then t_{1n} for $t_{\omega+n}$.

Thus reassured, let us continue. By the same argument as before, the sequence $(t_{\omega+n})$ converges, and it feels natural to write $t_{2\omega}$ for its limit. Again, as before, f is bounded on the interval $[0, t_{2\omega}]$.

Now we can continue the sequence – or perhaps one should say generalized sequence – with $t_{2\omega+1}$, $t_{2\omega+2}$ and so on. These converge to a number we can call $t_{3\omega}$, and if we keep on going we will end up producing, for every $m > 0$ and $n > 0$, a number $t_{m\omega+n}$ (I shall interpret t_0 as 0) such that f is bounded on the interval $[0, t_{m\omega+n}]$. Furthermore, $t_{m\omega+n}$ will be greater than $t_{k\omega+l}$ if and only if $m > k$ or $m = k$ and $n > l$.

What can we say now? What if none of these numbers is bigger than $1/100$, as may well be the case? Well, we are not forced to stop, because we could consider the increasing sequence $t_\omega, t_{2\omega}, t_{3\omega}, \dots$. This has a limit, which is easily seen to be bigger than any of the numbers $t_{m\omega+n}$, and what's more there is a natural name for this limit – t_{ω^2} . Once again, f will be bounded on the interval $[0, t_{\omega^2}]$.

For convenience, let me state and prove a **second basic lemma**, the one that we use for limits of sequences. It says that if we have an increasing sequence of numbers u_1, u_2, \dots in $[0, 1]$ and if f is a continuous function that is bounded on the interval $[0, u_n]$ for every n , then the u_n converge to a limit u and f is bounded on the interval $[0, u]$.

The fact that the u_n converge is just the monotone-sequences axiom for the real numbers. As for the boundedness of f on $[0, u]$, I proved it earlier, but let me repeat the argument. By continuity, f is bounded in some interval about u . This interval contains u_n for all sufficiently large n , and in particular for *some* n . Hence, putting together the interval $[0, u_n]$ and the interval surrounding u , we find that f is bounded on $[0, u]$.

Now that I have gone this far, it should be clear how to continue, obtaining numbers $t_{l\omega^2+m\omega+n}$ for any $l,m,n > 0$. And, of course, this is still not the end. One can construct $t_{p(\omega)}$ for any polynomial p with positive integer coefficients, in such a way that they come in the obvious order (a neat way to describe it is that $t_{p(\omega)} > t_{q(\omega)}$ if and only if $p(n) > q(n)$ for all sufficiently large integers n) and such that f is bounded on all the intervals $[0,t_{p(\omega)}]$. Then we can apply the second basic lemma to the sequence $t_\omega, t_{\omega^2}, t_{\omega^3}, \dots$ obtaining a limit, which is called – as you have probably guessed – t_{ω^ω} . I now leave it to you, if you are interested, to work out how to continue further, to $t_{\omega^{\omega^\omega}}$ and beyond, and then to take a limit t_{ϵ_0} , where ϵ_0 is the standard notation for what one might otherwise call ω to the ω to the ω to the ω to the ω to the ... and so on.

What is the point of all this? There is nothing to force even t_{ϵ_0} to be close to

1, and it begins to look as though, however ingenious we are in extending our notation in order to lengthen the “sequence”, we will never manage to reach 1 and hence never prove the theorem.

Is it conceivable that we could go *so far* with the process that we eventually succeeded in forcing the “sequence” to go beyond 1? In one way the answer is yes, and in another it is no.

The following observation gives us cause for optimism. If we could explain how to extend the sequence *uncountably* many times, then we would have proved the theorem. That is because every number in the “sequence” is followed by an open interval that contains no others, and these open intervals are disjoint. Since any open interval contains a rational number, there can be at most countably many of them. This would give us a contradiction, deduced from the assumption that we never passed 1 together with our method, whatever it might be, for continuing the process of extending uncountably many times.

On the other hand, a second observation should give us pause for thought. If you look at the discussion above, you will see that at various points I had to make a “notational leap”. Each of the following symbols required of me a small explanation, after which it became obvious for a while how to proceed: ω , $\omega + 1$, 2ω , ω^2 , ω^ω , ϵ_0 . Given that there are only countably many finite strings of symbols, it will not be possible to continue to find an adequate notation for uncountably many terms in the sequence (hypothetical terms that is – we eventually hope for a contradiction). That does not rule out a sort of infinitely long pseudo-notation (such as we use for real numbers – it is somehow comforting to write $\pi = 3.14159265358979323846264338327950288\dots$ and imagine the sequence going on for ever) but it does suggest that we should not be too *reliant* on notation. There is in fact a theorem of logic (whose precise statement, alas, I do not know) which says something like that one does indeed have to continue making notational leaps, and that eventually it ceases to be possible.

[Click here](#) if you would now like to skip to a general discussion of well-ordered sets and ordinals.

Problem 2. Generating the Borel sets

A *Borel set* of real numbers is a set that can be built out of open or closed sets by repeatedly taking countable unions or countable intersections. To be precise, open sets and closed sets are Borel, and if B_1 , B_2 , B_3 , \dots are Borel, then so is their union and their intersection.

Let me give a non-trivial example of a Borel set (i.e., one that isn’t something like the union of an open set with a closed set). First, here is a subset A of $(0, 1)$ which has three properties that might at first seem hard to reconcile: it is open and dense, but is nothing like the whole of $(0, 1)$. If you think that a dense open subset of $(0, 1)$ *must* be the whole of $(0, 1)$, then you should consider the union of $(0, 1/2)$ and $(1/2, 1)$ and see where your argument breaks down. The point is

that, although every point in some dense set is surrounded by a little interval, you can make the width of that interval depend on the point in order to leave other points out of your set.

Once you have spotted that, then you can go much further. Let q_1, q_2, q_3, \dots be an enumeration of the rationals in $[0, 1]$ and for each k let I_k be an interval of width r^{-k} centered on q_k . Let $A'(r)$ be the union of all the sets I_k and let $A(r)$ be the intersection of $A'(r)$ with $(0, 1)$. For $r > 2$ the sum of the widths of the I_k is less than 1, so $A'(r)$ cannot possibly be the whole of $(0, 1)$. (For those who know what this means, it has measure less than 1.) The larger r is, the less of $(0, 1)$ is contained in $A(r)$.

Now let A be the intersection of all the $A(r)$, for $r = 1, 2, 3, \dots$. Since all the $A(r)$ were open, A is a Borel set. What is A ? You might guess that A was just the set of all rationals in $(0, 1)$, but in fact this is quite wrong. (It is an interesting exercise to see why.) Since A contains all rationals but is not the whole of $(0, 1)$ (in fact, it has measure zero) it cannot possibly be closed. So it is a good example of a countable intersection of open sets which cannot be described any more simply than that.

Returning to more general Borel sets, let us think what a typical Borel set might look like. Let U_0 be the class of all open sets, and let U_1 consist of all intersections of countably many sets in U_0 . We cannot produce more sets by taking countable intersections of sets in U_1 (since such sets will themselves belong to U_1) but there is no guarantee that a countable *union* of sets in U_1 will belong to U_1 . In fact, examples can be constructed that don't.

We can therefore define U_2 to be the class of all countable unions of sets in U_1 , and in general can define U_n to be the set of all countable unions/intersections of sets in U_{n-1} according to whether n is even/odd.

Does the process stop there? Not at all. We can then define U_ω to be the class of all unions of sets A_1, A_2, \dots such that A_n belongs to U_n . This is a countable union of sets we have so far built, and there is no reason for it to belong to any of the classes U_n . (Incidentally, I am using the word "class" just to distinguish subsets of \mathbb{R} from sets of subsets of \mathbb{R} – not because there is any danger of running into Russell's paradox.)

Now we can let $U_{\omega+1}$ consist of all countable intersections of sets in U_ω , and so on. In fact, just as with Problem 1, one ends up with a hierarchy that goes well beyond infinity, and about which it gradually becomes harder and harder to speak. I leave it to the reader to work out which sets belong to U_{ω^ω} .

[Click here](#) if you would now like to skip to a general discussion of well-ordered sets and ordinals.

Problem 3. Showing that open games are determined

Many two-player games of skill are special cases of the following general one. Write $\mathbb{N}^{\mathbb{N}}$ for the set of all sequences of positive integers, and let X be a subset of $\mathbb{N}^{\mathbb{N}}$. Two players of the X -game take turns choosing positive integers, building a sequence (n_1, n_2, n_3, \dots) . The first player wins if this sequence belongs to X and the second wins if it doesn't.

A *strategy* for the first player is a function f that takes sequences (n_1, n_2, \dots, n_k) (with k even) to positive integers. To use the strategy f , all the first player does is look at the sequence (n_1, n_2, \dots, n_k) so far and choose $f((n_1, n_2, \dots, n_k))$ as the next move. A *winning strategy* for the first player is a strategy f such that, using it, the first player is guaranteed to win. Similarly one can define strategies and winning strategies for the second player. A game is called *determined* if one of the two players has a winning strategy.

How, you might wonder, could there possibly be a game that was not determined? If the first player has no winning strategy, does that not mean that the second player is guaranteed to win, and is that not the same as saying that the second player has a winning strategy?

The sketchy argument of the last paragraph doesn't work for the following reason. What we can deduce from the first player having no winning strategy is that, given any strategy, the second player can come up with another strategy that defeats it. (This strategy can be very simple – something like choose 2 then 8 then 673 then ... etc.) But there is no guarantee that the first player will stick to any one strategy.

But surely, you might still think, whatever the first player does is the result of *some* strategy, even if it isn't declared in advance. That is true, but it doesn't help the second player find a strategy. What we are trying to do is start from the premise "For every first-player strategy f there is a second-player strategy g that defeats f " to the conclusion "There is a second-player strategy g that defeats every first-player strategy f ". In other words, we are trying to interchange two quantifiers and obtain a g that is *independent* of f – which certainly shouldn't be trivial.

With the axiom of choice one can in fact build a game that is not determined – using a "just do it" proof of a kind explained here. The proof depends on well-orderings so I will give it later on.

An *open* game is defined to be one where, given any sequence (n_1, n_2, n_3, \dots) in X there exists a k such that all sequences that begin (n_1, n_2, \dots, n_k) are also in X . In other words, if the first player wins, then there was some point after which they would have won whatever moves they had played. Here are a few examples of open games. Although these games go on for ever, I shall think of them as finishing when they reach a finite sequence all of whose continuations lie in X – after which point they are not very interesting.

1. Let X be the set of all sequences with at least 300 occurrences of the number 10. It is of course very easy for player 1 to win this game – one strategy is to choose the number 10 for the first 300 moves. More to the point, if player 1 wins, there must be some point at which the number 10 has appeared 300 times, after which nothing can affect the outcome of the game.
2. Let X be the set of all sequences (n_1, n_2, n_3, \dots) with at least n_{10} occurrences of the number 10. Again, player 1 has an easy strategy: wait for player 2 to choose n_{10} and then keep choosing 10s until there are enough of them. Notice a difference between this game and the previous one though: whereas in the previous one player 1 could guarantee to win after 300 moves, in this one player 2 can make the game last arbitrarily long simply by choosing n_{10} arbitrarily large.
3. Let X be the set of all sequences (n_1, n_2, n_3, \dots) with at least $n_{n_{10}}$ occurrences of the number 10. Once again, player 1 has an easy strategy: wait for player 2 to choose n_{10} , then wait for player 2 to choose $n_{n_{10}}$ and then choose that number of 10s. Whereas with the previous time player 1 could say, “After 10 moves of the game I will be able to tell you how long it will take me to win,” for this game no such assurance can be given. The best that can be said is, “After 10 moves I will be able to tell you when I will know how long the game will take.”

As is becoming clear, there is a hierarchy of complexity appearing. Let us say that a game has complexity n if player 1 can guarantee to win after n moves. We could describe this inductively by saying that the complexity is 0 if player 1 wins no matter what happens and complexity n if after both players have moved once the complexity of the resulting position is $n - 1$. We can then say that a game has complexity ω if after the first moves the resulting position has complexity n for *some* n . Next, it has complexity $\omega + n$ if after the first moves the resulting position has complexity $\omega + n - 1$.

Just to illustrate this: the first game above has complexity 300 and the second has complexity $\omega + 4$. To see why the second statement is true, notice that on the fifth pair of moves the second player will be forced to decide on n_{10} after which the length of the game will be known and hence the complexity will be some positive integer.

A game has complexity 2ω if after the first move the resulting position has complexity $\omega + n$ for some n . It has complexity ω^2 if after the first move the game has complexity $a\omega + b$ for some a, b . (When I say “has complexity \dots ” I mean “has complexity no worse than \dots ”.) It has complexity ω^ω if after the first move the game has complexity some polynomial in ω .

And so it goes on, just as with the previous examples. What would player 1 be able to guarantee if the game had complexity ω^2 ? They would find it rather hard to explain, but might make an attempt like this: “I will be able to tell you how long it will be until I can tell you how long it will be until I can tell you

how long it will be until ... and so on ... I can tell you how long it will be until I win the game, and after the second player's first move I will be able to tell you how many times I needed to repeat the words 'I can tell you how long it will be until.'" I leave it to the reader to give an intuitive description of games of complexity ω^ω .

Before moving on, let us see why every game of complexity ω^2 is determined. I shall assume that every game of complexity $m\omega + n$ is determined (as an inductive hypothesis). If player 2 has no winning strategy, there must be some first move for player 1 such that player 2 still has no winning strategy. After they have both had their first turn, the position is a game of complexity $m\omega + n$ for some m, n , and is therefore determined. Since player 2 does not have a winning strategy, player 1 must. But this gives player 1 a strategy for the whole game – do that first move and then apply whatever winning strategy is appropriate after player 2's first move.

Well-ordered sets

Definition of well-ordered sets

A *total ordering*, or just ordering, on a set X is a transitive relation $<$ such that for every x, y in X exactly one of $x < y$, $x = y$ or $y < x$ is true. A total ordering is called a *well-ordering* if, in addition, every non-empty subset of X has a minimal element. That is, if Y is a non-empty subset of X , there exists some y in Y such that every z in Y is greater than or equal to y .

The most familiar well-ordering is the usual ordering on the natural numbers: every non-empty subset of \mathbb{N} has a least element. (This fact is discussed a bit more below.) A more complicated one was defined earlier, namely the set of all polynomials p with non-negative integer coefficients, with $p < q$ if and only if $p(x) < q(x)$ for all sufficiently large x . To see that this is a well-ordering, let P be a set of such polynomials. From P we can pick all polynomials of least degree. (Note that if $\deg(p) < \deg(q)$ then $p < q$.) Amongst those we can select with smallest leading coefficient. (Note that if p has a smaller leading coefficient than q and they are of the same degree, then $p < q$.) Amongst those we can take the ones with smallest next coefficient, and so on. We end up with a polynomial p which is smaller in the ordering than every other q in P .

Maps between well-ordered sets

A set with a total ordering is called a totally ordered set. A map f between two totally ordered sets X and Y is called an *order-isomorphism* if it is a bijection and $f(x) < f(y)$ if and only if $x < y$. In other words, it is an invertible map that preserves order. If there is an order-isomorphism between X and Y then X and Y are said to be *order-isomorphic*.

Let X be a well-ordered set. An *initial segment* of X is a subset of the form $I(z) = \{x : x < z\}$. A very important fact about well-ordered sets is the

following: if X and Y are well-ordered sets then either X is order-isomorphic to Y or one of X and Y is order-isomorphic to an initial segment of the other. Here is (a sketch of) how to prove it.

Roughly what we want to do is to take the smallest element of X and pair it with the smallest element of Y , then the next smallest then the next smallest and so on. Of course, this “and so on” is one of our problematic ones, because we may need to pair off the ω th smallest elements and so on well beyond infinity. However, we can avoid getting into those details by using the well-ordering property.

Let Z be the set of all z in X such that $I(z)$ is order-isomorphic to an initial segment of Y . If z is in Z and $w < z$ then it is easy to see that w is in Z . Hence, either Z is all of X or it is the initial segment $I(u)$, where u is the least element of X that does not belong to Z .

Next, I claim that if z is in Z , then there is exactly one isomorphism from $I(z)$ to an initial segment of Y . For if not, let w be minimal such that there are two distinct order-isomorphisms f and g from $I(w)$ to initial segments of Y . Let v be minimal such that $f(v)$ does not equal $g(v)$. Then f and g must agree on the initial segment $I(v)$, and $f(v)$ and $g(v)$ are then forced to be the least element of Y that does not belong to $f(I(v))$.

This observation allows us to define an order-isomorphism from Z into Y – each z in Z maps to the least element of Y not included in $f(I(z))$. Then either $f(Z) = Y$, in which case we are done, or $f(Z)$ is a proper subset of Y , in which case Z must be the whole of X or we’d be able to extend f .

Order-types of well-ordered sets

As you may well have noticed, there is an obvious notation for at least some of the elements of a well-ordered set X . We could call the first element 1, the next 2, the next 3 and so on. (I can always talk about “the next”: it just means the minimal element not yet mentioned.) If I’ve gone through all the natural numbers and still haven’t included all of X , then I’ll call the next element ω , the next $\omega + 1$ and so on. Then after those I’ll have 2ω , $2\omega + 1$ and so on. Then 3ω , 4ω , \dots , ω^2 , \dots , ω^ω , \dots , ω^{ω^ω} , \dots , ϵ_0 , \dots . I didn’t mention earlier that there are also ϵ_1 , ϵ_2 , \dots , ϵ_ω , ϵ_{ϵ_0} , \dots and even $\epsilon_{\epsilon_{\dots}}$.

The notion of order-isomorphism gives us what we’d like to call an equivalence relation on the set of all well-ordered sets. Unfortunately, we can’t call it that because, for example, every set of the form $\{A\}$, where A is a set, can be well-ordered, and the object $\{\{A\} : A \text{ is a set}\}$ is too big to be a set. Anyhow, in the back of our minds is the idea that the *order type* of a well-ordered set is the isomorphism class of that set, but the difficulty just mentioned forces us to define it differently – by picking a representative from each class. That is, we’d like to choose a whole lot of well-ordered sets in such a way that every well-ordered set is order-isomorphic to exactly one of our special ones. We will call them 1, 2, 3,

$\dots, \omega, \omega + 1, \dots$ etc. and the general name for them will be ordinals.

How to use ordinals

Let us think once again about the three problems we were grappling with earlier. What we were looking for was an *indexing set* – that is, a set X which we could use to label the terms of an infinite list, or rather something like a list. In each case, although the list would sometimes go on for ever but still not have finished, we could at any stage talk about the *next* item on the list. This suggests that what we really needed was a well-ordered set. The natural numbers are a very useful well-ordered set and are often used as an indexing set, but for these applications we wanted something larger.

In this section I shall show how the following statement does the job. For the moment we can think of it as an axiom and later I shall prove it.

Axiom: *There exists an uncountable well-ordered set.*

Let me show how this axiom, together with the two basic lemmas, solves Problem 1 for us, and allows us to conclude that continuous functions on $[0, 1]$ are bounded. Recall that we were building a sequence t_1, t_2, \dots which went on well beyond infinity. It is not hard to see that, at least to the point that we took it, this sequence forms a well-ordered subset of \mathbb{R} (under the usual ordering). Indeed, it is order-isomorphic to the set we used to label the sequence – which, though we thought of it as mere notation, does come with a natural ordering.

With that procedure in the back of our minds, let X be an uncountable well-ordered set and let us imagine that we have tried to build a well-ordered subset T of $[0, 1]$ indexed by X , so that a typical element of T will be written t_x for some x in X , and the map x goes to t_x will be an order-isomorphism.

Now I can make more formal an argument I gave earlier. For every x in X there is a *successor*, that is, a least element y of X which is greater than x . Since the map x goes to t_x is an order-isomorphism, there is no element of T between t_x and t_y . Therefore, that interval contains a rational q_x and all those rationals are distinct. That contradicts the fact that X , and hence T , were supposed to be uncountable.

This shows that there is no uncountable well-ordered subset of $[0, 1]$, which means that the “sequence” I constructed must eventually come to an end – by reaching 1. To make this formal, let us define the sequence explicitly as follows. First enumerate all the rationals in $[0, 1]$. Now, if I have defined t_x then let y be the successor of x . The first basic lemma tells us that there is an $s > x$ such that f is bounded on the interval $[0, s]$. The interval $(x, s]$ contains many rational numbers. For the sake of definiteness let t_y be the first rational q in the enumeration such that f is bounded on the interval $[0, q]$. If y is not the successor of any x , then define t_y to be the supremum of all t_x such that $x < y$. Since f is bounded up to every t_x with $x < y$ and since these get arbitrarily close to t_y , f is bounded up to t_y .

As long as t_y never equals 1 this defines a unique order isomorphism from X into $[0, 1]$. (If not, let y be the first element of X where anything goes wrong. But I've just shown unambiguously how to define t_y once I've defined t_x for all previous x .) It follows that for some y $t_y = 1$ and this shows that f is bounded on $[0, 1]$.

Now let us return to Borel sets. We can measure the complexity of such sets as follows. Say that a set has complexity at most 1 if it is open or closed, complexity at most 2 if it is a countable union of closed sets or a countable intersection of open sets, and in general complexity at most \mathbb{N} if it is a countable union or intersection of sets of complexity at most $n - 1$. If in addition the set is not of complexity at most $n - 1$ then say that the set is of complexity n . More generally still, given an element x of X , say that a set has complexity at most x if it is a countable union or intersection of sets each of which has complexity strictly less than x . [As before, this definition makes sense – or there would be a smallest X for which it does not make sense, which there clearly isn't.]

Let us call an element x in X *countable* if the initial segment $I(x)$ is countable. I now claim that every Borel set has complexity x for some countable x . Let us say that such a set has *countable complexity*. To prove that, all I need to do is show that a countable union or intersection of sets of countable complexity also has countable complexity. Then it follows that, no matter how much one takes countable unions or intersections, starting with open and closed sets, the result will always have countable complexity.

Let A_1, A_2, \dots be sets with countable complexities x_1, x_2, \dots . I claim first that there is an element x of X that exceeds all of the x_n . To see this, first note that the union of the sets $I(x_n)$ is not the whole of X , since it is a countable union of countable sets and X is uncountable. Now let x be the smallest element of X that belongs to none of the intervals $I(x_n)$. It is possible that $x = x_n$ for some n , but we can at least say that the successor y of x exceeds every x_n , from which it follows that the union or intersection of the A_n has complexity at most y .

It is not obvious, but can be shown, that for every countable x in X there is a Borel set of complexity x . (Earlier we showed this for $x = 2$, with an unnecessarily complicated example – the rational numbers is another.)

Using X , we can also measure the complexity of winning strategies for player I in open games. By X I now mean the uncountable well-ordered set and not the set of all winning sequences for the first player, which I shall call W . All we have to say is that a strategy is of complexity 0 if every sequence is a win for player 1, that is, if $W = \mathbb{N}^{\mathbb{N}}$, so that player 1 has won before the game even starts. It is of complexity at most x if, whatever player 2 plays for the first move, player 1's strategy is then of complexity at most y for some $y < x$. (For example, it is of complexity at most w if after the first pair of moves player 1 can declare, for

some n , that the game will definitely be won in the next n moves.) A strategy has complexity X if it has complexity at most x and for no $y < x$ does it have complexity at most y .

Here is a proof that, for every open game, either player 2 has a winning strategy or player 1 has a winning strategy of some countable complexity. Suppose that player 1 does not have such a strategy. Then, whatever player 1 plays, player 2 must have a move such that player 1 still does not have a winning strategy of countable complexity. Why is this? Well, otherwise player 1 could play some n such that, for every possible move nmn of player 2 there was then a winning strategy f_m of complexity some countable x_m . Then the following is a winning strategy for player 1: play n and follow with strategy f_m if player 2 plays m . Since the x_m are countable, there is some x in X that exceeds all of them, so this strategy has complexity at most x .

This proves my claim that player 2 has a move that ensures that player 1 still has no winning strategy of countable complexity. But this argument can now be repeated. In other words, player 2 can play in such a way that player 1 never has a winning strategy of countable complexity, and in particular never wins the game (or else the complexity would be zero). If player 2 plays like this, the result will be some infinite sequence (n_1, n_2, n_3, \dots) . If it belonged to W , then there would be some k such that all sequences beginning (n_1, n_2, \dots, n_k) were also in W – by the definition of an open game. But this contradicts the fact that player 1 never wins. Hence, player 2 has a way of ensuring that the game results in a sequence not in W – that is, has a winning strategy.

Proof that there is an uncountable well-ordered set

Recall from earlier that we wanted to define ordinals to be order-isomorphism classes of well-ordered sets, but that set-theoretical considerations forbade this. This problem can be avoided if all we need is countable ordinals. Here is a definition that will do: a *countable ordinal* is an isomorphism class of well-ordered subsets of \mathbb{R} . (We proved above that every well-ordered subset of \mathbb{R} must be countable.) What, for example, is $\omega + 2$? It is the set of all sets of the form $\{t_1, t_2, \dots, u_1, u_2\}$, where the t_n are a bounded increasing sequence, and every t_n is less than u_1 which is less than u_2 .

How can we now prove that there is an uncountable well-ordered set? All we have to do is use ordinals, as just constructed. Let us see why they come with a natural well-ordering and why there are uncountably many of them.

The ordering is as follows. Write a, b for typical well-ordered subsets of \mathbb{R} and $[a], [b]$ for their isomorphism classes. Then $[a] < [b]$ if a is isomorphic to an initial segment of b . Clearly this gives a well-defined total ordering. To see that it is a well-ordering, let A be a set of ordinals and let $[a]$ be an element of A . Then either $[a]$ is minimal or there is a non-empty subset B consisting of all $[b]$ such that b is isomorphic to an initial segment $I(t)$ of a . Pick $[b]$ such that t is

minimized (which can be done as a is a well-ordered set of reals). Then $[b]$ must be a minimal element of A .

Now let us prove that there are uncountably many countable ordinals. (There is no paradox here – just as it is not paradoxical that there are infinitely many finite numbers.) If there were countably many, then we could write them as $[a_1], [a_2], \dots$. Now it is easy to construct an order-preserving bijection from \mathbb{R} to the open interval $(-1, 1)$ (something built out of \tan^{-1} will do the job). Hence, given a well-ordered subset a of \mathbb{R} and an open interval (c, d) , we can find a well-ordered subset of (c, d) that is isomorphic to a .

We can therefore build a set as follows. For each \mathbb{N} let b_n be an isomorphic copy of a_n that lives in the interval $(n-1, n)$ and let b be the union of the b_n . It is easy to see that b is well-ordered – given any non-empty subset first find the minimal \mathbb{N} such that it intersects $(n-1, n)$ and then find the minimal element of that intersection.

Next, I claim that $[b]$ is at least as big as every $[a_n]$. If this were not the case, then we would be able to find \mathbb{N} such that b was isomorphic to an initial segment of $[a_n]$. But since a_n embeds into b via the isomorphism with b_n this would show that $[a_n]$ was order-isomorphic to a subset of an initial segment $I(t)$ of itself. But it is not hard to show that this is impossible. Let f be the supposed isomorphism and let s be minimal such that $f(s) < s$. (Such an s exists – t is an example.) But then $f(s) = u$ for some $u < s$ and $f(u) \geq u$, so f is not order-preserving.

Finally, if $[b]$ is at least as big as every $[a_n]$ then the successor of $[b]$ cannot be one of the $[a_n]$. (The successor is obtained by taking a bounded isomorphic copy of b , adding in an upper bound and taking the isomorphism class.)

How not to use ordinals

Now comes the moment to admit that my ‘applications’ of countable ordinals were, in a sense, a con. The application to Borel sets wasn’t really solving a problem – it was just classifying the Borel sets in quite an interesting way. As for the other two results – that continuous functions on $[0, 1]$ are bounded and that open games are determined – it is downright silly to use ordinals for their proofs and very easy to remove them. This is almost always true of proofs that use countable ordinals. Though there are probably several counterexamples to this assertion, I myself know of only one theorem proved with countable ordinals for which a neater ordinal-free proof has not been discovered, and even there I am convinced that it exists.

Ordinals and induction

It is a familiar fact about induction that it can be done in two ways. Given a sequence of statements $P(n)$, either you prove $P(0)$ and that $(P(l) \text{ for } l < k) \implies P(k)$ or you say, ‘Let k be minimal such that $P(k)$ is false’ and derive a contradiction. In both cases the actual work is the same – you get the

contradiction by showing that $P(k)$ follows from the truth of all previous $P(l)$, unless $k=0$ in which case you must argue separately.

What we were doing above is very similar to induction except that there are two sorts of induction step. The more straightforward one is showing that a statement $P(y)$ (for some ordinal y in X) implies $P(x)$, where X is the successor of y . The other one is a ‘limiting’ stage, where we show that if X is not the successor of any y , then we can at least deduce the truth of $P(x)$ from the truth of all $P(y)$ for which $y < x$.

If $P(0)$ is true and both sorts of induction step are proved, then $P(x)$ is true for every X in X . Why? Because if not then there must be some minimal X for which $P(x)$ is false. But then either X is the successor of y for which $P(y)$ is true, or $P(y)$ is true for every $y < X$ (in fact, the second statement is enough on its own) and we have a contradiction.

Although induction and picking a minimal counterexample are formally the same thing, they feel different. With induction one imagines some infinitely long process taking place as one gradually proves more and more of the statement, whereas with minimal counterexamples one has the illusion that the proof takes only one step. For this reason, choosing a minimal counterexample is often cleaner. The same is true for generalized ‘ordinal induction’.

Solving problems 1 and 3 without ordinals.

Returning to the proof that continuous functions are bounded on $[0, 1]$, here are two ways of removing ordinals. The first is to change slightly how I defined the sequence t_x . If y is the successor of x , a more natural way to choose t_y is to make it as large as possible, subject to the constraint that $|f(s) - f(t_x)|$ should be at most 1 for every s in the interval $[t_x, t_y]$. Luckily, this makes sense – the set of all numbers satisfying the constraint attains its upper bound. Now suppose we generate the sequence t_1, t_2, \dots in this way. Then either it reaches 1 or it converges to some limit t . But there is an interval about t in which f oscillates by at most $1/2$, and that interval must contain some t_n , which proves that t_{n+1} is bigger than t – a contradiction.

In other words, it was only because of a perverse definition of t_y (using an enumeration of the rationals) that there was ever any chance of going beyond w in the ordinal hierarchy.

But once we have noticed this simpler proof, a yet simpler one suggests itself. After all, the number 1 was rather arbitrary above. Why not simply define t_{n+1} to be as large as possible subject to the constraint that f should be bounded on the interval $[t_n, t_{n+1}]$? Again, this makes sense. If f is bounded on the interval $[0, s]$ for every $s < t$, then f is bounded on the interval $[0, t]$ (since it is bounded on an interval surrounding t). Setting $t_0=0$, what will t_1 be? It can’t be less than 1 because then t_1 is surrounded by an interval on which f is bounded,

contradicting maximality. So it's 1, and there isn't even a sequence.

This is basically the proof I gave in the page on the boundedness of continuous functions on $[0, 1]$, and it can be thought of as looking for a minimal counterexample – focusing on the infimum of the set of all t such that f is not bounded on $[0, t]$, and deriving a contradiction. This is not meant as a formal statement, as I wasn't taking the minimal element of a subset of a well-ordered set, and in fact I do not know of a formal procedure for removing ordinals from proofs, though again I feel that something might be possible (and might be known to set theorists).

I devoted some time to explaining why it wasn't obvious that all games are determined. However, I do not wish to suggest that one's intuition that they are is entirely stupid. Why, for example, is it obvious in chess that either white has a winning strategy or black has a strategy for forcing at least a draw? Well, if white has no winning strategy, then after white's first move there must be some way for black's first move to result in a position for which white still has no winning strategy (or else white could win by seeing what black did and applying the appropriate strategy from that point on). Similarly, after white's second move there must be a move for black that does not give white a winning strategy, and so on. In this way black can avoid defeat, but since the number of moves in a chess game is bounded (if the same position is reached three times then it is a draw), that is all black needs to do to force a draw.

In other words, what one *can* conclude from player 1 not having a winning strategy is that player 2 has a strategy that makes it possible to avoid losing at any finite stage of the game. If the game is open, this will in fact be a winning strategy. To see this, suppose that the game comes to an end with a sequence (n_1, n_2, n_3, \dots) and that at no stage has player 1 won. Then this sequence cannot be a win for player 1, or there would have been some k such that player 1 had won after the k^{th} move. Therefore, it is a win for player 2. Note that I reasoned identically at the end of the earlier, ordinals proof. What I have now shown is that the ordinals were a red herring. It is interesting to classify strategies by their ordinal complexity, but quite unnecessary if all one wants is the easy result that open games are determined.

A very brief sketch of how to “construct” a game that is not determined

As I mentioned earlier, this needs the axiom of choice, so this section requires slightly more knowledge than the rest of the page. (I should add that when I say "needs the axiom of choice" I am speaking informally. Some set theorists are fans of the so-called axiom of determinacy, which states that all games are determined. This is definitely not equivalent to the negation of the axiom of choice, even though, as we shall see, it implies it.)

We use the axiom of choice in the following form: every set S can be well-ordered in such a way that all initial segments of the well-ordering have smaller cardinality

than that of S . In our case, we consider the set S of all possible strategies for both players of the game, which has the cardinality of the continuum. Imagine that it has been well-ordered in the way specified. Now take the first strategy and suppose that it is a strategy for the first player. Run the game with the first player using this strategy and the second making some arbitrary set of moves. This generates a sequence which we shall place outside X (thus ensuring that the second player's moves result in a sequence not in X and thereby defeat the first strategy.) Now take the second strategy in the well-ordering. Suppose it is a strategy for the second player. Choose arbitrary moves for the first player and run the game with this strategy, *except* that the first player's moves should be chosen in such a way that the resulting sequence is not the same as the sequence we first produced. This sequence is easy to avoid, though, as there are continuum-many choices for the moves of the first player. Having generated the second sequence, put it in X (ensuring now that the second strategy is not a winning strategy for the second player).

Continue this way, at each stage taking the minimal strategy (in the well-ordering) not yet considered. Since one has at each stage constructed strictly fewer than continuum-many sequences, it is easy to choose moves for the player playing against the strategy under consideration so as to produce an entirely new sequence, which can then be placed in or out of X as is appropriate. When all strategies have been considered, the remaining sequences can be assigned to X or X^c arbitrarily. And now, in a trivial just-do-it way we have made sure that no strategy is a winning strategy.

How can infinity plus one not be infinity?

There was a time when this question bothered me and made me feel uneasy about ordinals. However, it has a simple and precise answer, one that is quite revealing about the abstract method in mathematics.

Why do people say that infinity plus one equals infinity? Presumably because if X is an infinite set, y is not in X and $Y = X \cup \{y\}$, then Y can be put into one-to-one correspondence with X . Or, even more concretely, the natural numbers can be put into one-to-one correspondence with the non-negative integers via the map $f(n) = n - 1$. So it looks, for example, as though $\omega + 1$ ought to be the same size as ω .

Well, what *is* correct about that last assertion is that a well-ordered set of order type ω can be put into one-to-one correspondence with a well-ordered set of order-type $\omega + 1$. However, the one-to-one correspondence will not be order-preserving. In other words, there are two notions of isomorphism that one can choose to be interested in. One is appropriate for sets: two sets are isomorphic if there is a bijection between them. The other is appropriate for well-ordered sets: two well-ordered sets are isomorphic if there is an *order-preserving* bijection between them. To say that ω and $\omega + 1$ are distinct is to say that they are not order-isomorphic. (A small complication: as I defined

countable ordinals, they were not well-ordered sets, but equivalence classes of well-ordered sets. To make the above discussion strictly correct, one can change the definition as follows: take the uncountable well-ordered set defined earlier and define a countable ordinal to be any proper initial segment of it. Alternatively, define two ordinals to be order-isomorphic if any two well-ordered sets with the corresponding order types are order-isomorphic.) As a category theorist might put it, the category of sets with arbitrary functions between them is different from the category of well-ordered sets with order-preserving functions between them. Since the second has much more structure, it is not too surprising that it has a wider variety of infinite objects.

One thing this demonstrates is that there is not just one single notion of "infinity" waiting there for its properties to be explored. Rather, we can think of finite numbers in many different ways, and different ways of thinking about them lead to different generalizations to infinite "numbers". Thinking about the sizes of integers leads to cardinals and thinking about their order properties leads to ordinals. In both cases the word "numbers" is slightly misleading.