Big Numbers

Tom Davis
tomrdavis@earthlink.net
http://www.geometer.org/mathcircles
September 21, 2011

1 Introduction

> The problem is that we tend to live among the set of puny integers and generally ignore the vast infinitude of larger ones. How trite and limiting our view!
>
> — P.D. Schumer

It seems that most children at some age get interested in large numbers. "What comes after a thousand? After a million? After a billion?", et cetera, are common questions.

Obviously there is no limit; whatever number you name, a larger one can be obtained simply by adding one. But what is interesting is just how large a number can be expressed in a fairly compact way.

Using the standard arithmetic operations (addition, subtraction, multiplication, division and exponentiation), what is the largest number you can make using three copies of the digit "9"?

It's pretty clear we should just stick to exponents, and given that, here are some possibilities: 999, 99^9, 9^{99}, and 9^{9^9}.

Already we need to be a little careful with the final one: 9^{9^9}. Does this mean: $(9^9)^9$ or $9^{(9^9)}$? Since $(9^9)^9 = 9^{81}$ this interpretation would gain us very little. If this were the standard definition, then why not write a^{b^c} as a^{bc}? Because of this, a tower of exponents is always interpreted as being evaluated from right to left. In other words, $a^{b^c} = a^{(b^c)}$, $a^{b^{c^d}} = a^{(b^{(c^d)})}$, et cetera.

With this definition of towers of exponents, it is clear that 9^{9^9} is the best we can do. If expanded completely, it would have more than $369,000,000$ decimal places.

It's a different story with three copies of the digit 2, however. The number 2^{22} is the best you can do.

It's fairly clear that stacking exponents makes huge numbers very rapidly, and you might at first think that if you've got 100 symbols, your best bet would be to stack 100 9s in a tower of exponents and that would pretty much beat all the other possibilities. While this would be a pretty large number, with a little cleverness you can do *far* better.

2 Factorials and Their Relatives

The factorial function, $n! = n \times (n-1) \times \cdots 3 \times 2 \times 1$ is known to generate very large numbers, but if we're counting symbols in our description of large numbers, we can probably do better with 9's than with the factorial. With no limit on factorials, for example, there is of course no limit to the size of a number we could generate with just three 9's, since we could have:

$$999!, (999!)!, ((999!)!)!, \ldots.$$

(The parenthesis are required above, since usually the doubled factorial symbol means something else: $8!! = 8 \times 6 \times 4 \times 2$ and $9!! = 9 \times 7 \times 5 \times 3 \times 1$, et cetera.)

There is also a hyper-factoral $H(n)$ function that generates even larger numbers:

$$H(n) = 1^1 \times 2^2 \times 3^3 \times \cdots \times n^n.$$

Finally, Pickover has defined a superfactorial function, $n\$$, defined as follows:

$$n\$ = \underbrace{n!^{n!^{\cdot^{\cdot^{\cdot^{n!}}}}}}_{n!\text{ copies}},$$

and even $3\$$ has an enormous number of digits.

3 Names of Large Numbers

How are large numbers named? There's some disagreement in the English language about how to name them. There are two systems, one used by Americans and the other by the English. Most of the rest of the world uses a system similar to the English one. In a sense, the English system is a bit more logical in that the prefixes "Bi", "Tri", "Quad", et cetera stand for two, three, and four groups of six zeroes. In the American system, "Bi", "Tri" and "Quad" stand for three, four, and five groups of three zeroes.

Here are the first few names in both systems:

Number	Scientific Notation	American Name	English Name
$1,000$	10^3	Thousand	Thousand
$1,000,000$	10^6	Million	Million
$1,000,000,000$	10^9	Billion	Thousand Million
$1,000,000,000,000$	10^{12}	Trillion	Billion
$1,000,000,000,000,000$	10^{15}	Quadrillion	Thousand Billion
$1,000,000,000,000,000,000$	10^{18}	Quintillion	Trillion

In the English system, sometimes the names "Milliard", "Billiard", "Trilliard", "Quadrilliard", et cetera, are used in place of "Thousand Million", "Thousand Billion", et cetera.

The names continue in the same general way, and after Quintillion, they are: Sextillion (10^{21}), Septillion (10^{24}), Octillion (10^{27}), Nonillion (10^{30}), Decillion (10^{33}), Undecillion (10^{36}), Duodecillion (10^{39}), Tredecillion (10^{42}), Quattuordecillion (10^{45}), Quindecillion (10^{48}), Sexdecillion (10^{51}), Septendecillion (10^{54}), Octodecillion (10^{57}), Novemdecillion (10^{60}), Vigintillion (10^{63}), Unvigintillion (10^{66}), Dovigintillion (10^{69}), Trevigintillion (10^{72}), Quattuorvigintillion (10^{75}), Quinvigintillion (10^{78}), Sexvigintillion (10^{81}), Septenvigintillion (10^{84}), Octovigintillion (10^{87}), Novemvigintillion (10^{90}), Trigintillion (10^{93}).

The numbers in parentheses above correspond to the American names. To obtain the English values, subtract 3 from each exponent and then double it.

The name Decillion is sometimes used to indicate 10^{303} in the American system and 10^{600} in the English system. The name Googol stands for 10^{100}—a 1 followed by a hundred zeroes—and the name Googolplex means the number 1 followed by a Googol zeroes.

4 Scientific Names

You've probably heard of prefixes like "centi" meaning "hundredth" and "mega" meaning "million". So "centimeter" is one hundredth of a meter and "megavolt" is a million volts. There is an SI-approved set of prefixes for measurements for a large number of sizes. Here is a table:

Multiplier	Prefix	Multiplier	Prefix
10^1	deca	10^{-1}	deci
10^2	hecto	10^{-2}	centi
10^3	kilo	10^{-3}	milli
10^6	mega	10^{-6}	micro
10^9	giga	10^{-9}	nano
10^{12}	tera	10^{-12}	pico
10^{15}	peta	10^{-15}	femto
10^{18}	exa	10^{-18}	atto
10^{21}	zetta	10^{-21}	zepto
10^{24}	yotta	10^{-24}	yocto

So in 20 grams of hydrogen gas, there are about 10 moles of H_2 molecules, or $10 \times 6.023^{23} = 6.023 \times 10^{24} = 6.023$ yotta molecules.

5 A Really Big Number

This is the first *really* big number I ever saw. It was shown to me by Leo Moser while I was in high school.

First, we begin by saying what we mean by $\triangle(n)$ (which we will pronounce as "triangle-n"). It will simply be defined as n^n:

$$\triangle(n) = n^n.$$

Thus, $\triangle(2) = 2^2 = 4$, and $\triangle(3) = 3^3 = 27$, et cetera.

Next, we'll define "square-n": $\boxed{n}$, to be the number n surrounded by n triangles. Thus, $\boxed{2} = \triangle(\triangle(2)) = \triangle(4) = 4^4 = 256$.

In a similar way, we'll define "pentagon-n": $\pentagon(n)$, to be the number n surrounded by n squares. Thus, $\pentagon(2) = \boxed{\boxed{2}} = \boxed{256}$. So how big is "square-256"? Well, it is the number 256 surrounded by 256 triangles. To get rid of the innermost triangle, we obtain 256^{256}, a number with 616 digits, surrounded by 255 triangles. Raising this 616-digit number to its own power will leave us with "only" 254 triangles, et cetera. It is fairly obvious that this number is almost unimaginably large. The number "pentagon-2" is sometimes called just "mega", and "pentagon-10" is sometimes called "megistron".

In the same way, we can define "hexagon-n", "heptagon-n", and so on.

But at this point I become bored with my notation so I'll introduce an easier one. Define "2-sub-1", "2-sub-2", et cetera, as:

$$\begin{aligned} 2_1 &= \triangle(2) \\ 2_2 &= \boxed{2} \\ 2_3 &= \pentagon(2) \\ 2_4 &= \hexagon(2) \end{aligned}$$

and so on. The number I am *really* interested in is:

$$2_{\pentagon(2)},$$

or in other words, "2-sub-pentagon-2".

As professor Moser said, "Of course now that I've shown you this one, you can show me a larger one, but this is probably the largest number you have ever seen.", and he was certainly right for that particular group of high-school students. This huge number is sometimes called the Moser number.

6 Ackermann's Function

Note: If you are a teacher presenting this to younger students, you may find the pedagogical notes in Section 13 useful.

Let's consider an innocent-looking function which is simple to describe, but which will increase at an unbelievable rate. It is a very famous rapidly-increasing function introduced by Ackermann to settle a problem in logic. He wanted to show that there exist "general recursive" functions that are not "primitive recursive". We don't worry about the definitions of "primitive recursive" or "general recursive", but his proof consisted of the presentation of a general recursive function (Ackermann's function) that he could show to increase more rapidly than any primitive recursive function.

The function we wish to calculate is called Ackermann's function, and it is defined as follows:

$$A(m,n) = \begin{cases} n+1 & : \text{ if } \mathrm{m} = 0 \\ A(m-1,1) & : \text{ if } \mathrm{m} > 0, \mathrm{n} = 0 \\ A(m-1, A(m,n-1)) & : \text{ if } \mathrm{m} > 0, \mathrm{n} > 0 \end{cases} \tag{1}$$

The easiest way to understand it is to make a table of the values of $A(m,n)$ beginning with the easy ones. In the table below, n increases to the right, and the rows correspond to $m = 0, 1, 2, 3, \ldots$. The "*"s indicate values we have not yet determined. The first row, where $m = 0$, is easy. The general formula is obviously $A(0,n) = n+1$.

$m\backslash n$	0	1	2	3	4	5	6	7	8	9	10
0	1	2	3	4	5	6	7	8	9	10	11
1	*	*	*	*	*	*	*	*	*	*	*
2	*	*	*	*	*	*	*	*	*	*	*
3	*	*	*	*	*	*	*	*	*	*	*
4	*	*	*	*	*	*	*	*	*	*	*

The next row where $m = 1$ is a bit trickier. If $n = 0$ we can use the second line in formula 1 that defines Ackermann's function to obtain: $A(1,0) = A(0,1) = 2$.

What is $A(1,1)$? We must apply the third row in formula 1 to obtain: $A(1,1) = A(0, A(1,0)) = A(0,2) = 3$. Similarly: $A(1,2) = A(0, A(1,1)) = A(0,3) = 4$. Make sure you understand what is going on by working out a few more, and finally we can fill out the second row of the table as follows.

$m\backslash n$	0	1	2	3	4	5	6	7	8	9	10
0	1	2	3	4	5	6	7	8	9	10	11
1	2	3	4	5	6	7	8	9	10	11	12
2	*	*	*	*	*	*	*	*	*	*	*
3	*	*	*	*	*	*	*	*	*	*	*
4	*	*	*	*	*	*	*	*	*	*	*

The general formula for this row is $A(1,n) = n+2$.

The third row can be approached in the same way:

$$\begin{aligned} A(2,0) &= A(1,1) = 3 \\ A(2,1) &= A(1, A(2,0)) = A(1,3) = 5 \\ A(2,2) &= A(1, A(2,1)) = A(1,5) = 7 \\ A(2,3) &= A(1, A(2,2)) = A(1,7) = 9 \end{aligned}$$

We can continue (do so for a few more) to obtain the third row:

$m\backslash n$	0	1	2	3	4	5	6	7	8	9	10
0	1	2	3	4	5	6	7	8	9	10	11
1	2	3	4	5	6	7	8	9	10	11	12
2	3	5	7	9	11	13	15	17	19	21	23
3	*	*	*	*	*	*	*	*	*	*	*
4	*	*	*	*	*	*	*	*	*	*	*

The general formula for this row is $A(2,n) = 2n+3$.

Repeat the process for $A(3,n)$:

$$\begin{aligned} A(3,0) &= A(2,1) = 5 \\ A(3,1) &= A(2,A(3,0)) = A(2,5) = 13 \\ A(3,2) &= A(2,A(3,1)) = A(2,13) = 29 \\ A(3,3) &= A(2,A(3,2)) = A(2,29) = 61 \end{aligned}$$

If you do a few more, you will see that the table now looks like this:

$m\backslash n$	0	1	2	3	4	5	6	7	8	9	10
0	1	2	3	4	5	6	7	8	9	10	11
1	2	3	4	5	6	7	8	9	10	11	12
2	3	5	7	9	11	13	15	17	19	21	23
3	5	13	29	61	125	253	509	1021	2045	4093	8189
4	*	*	*	*	*	*	*	*	*	*	*

The general formula for this row is $A(3,n) = 2^{n+3} - 3$.

Beginning with the next line, things begin to get ugly:

$$\begin{aligned} A(4,0) &= A(3,1) = 13 = 2^{2^2} - 3 \\ A(4,1) &= A(3,A(4,0)) = A(3,13) = 65533 = 2^{2^{2^2}} - 3 \\ A(4,2) &= A(3,A(4,1)) = A(3,65533) = 2^{65536} - 3 = 2^{2^{2^{2^2}}} - 3 \\ A(4,3) &= A(3,A(4,2)) = A(3,2^{65536} - 3) = 2^{2^{65536}} - 3 = 2^{2^{2^{2^{2^2}}}} - 3 \end{aligned}$$

The general form for $A(4,n)$ is this: $A(4,0) = 2^{2^2} - 3$, $A(4,1) = 2^{2^{2^2}} - 3$, and in general, each time we increase the value of n, the height of the tower of exponents of 2 increases by 1. If we denote by $T(n)$ the value of a tower of exponents of height n, where all the exponents are 2, then $A(4,n) = T(n+3) - 3$:

$m\backslash n$	0	1	2	3	4	5	6	7	8
0	1	2	3	4	5	6	7	8	9
1	2	3	4	5	6	7	8	9	10
2	3	5	7	9	11	13	15	17	19
3	5	13	29	61	125	253	509	1021	2045
4	$2^{2^2} - 3$	$2^{2^{2^2}} - 3$	$2^{2^{2^{2^2}}} - 3$	$2^{2^{2^{2^{2^2}}}} - 3$	$2^{2^{2^{2^{2^{2^2}}}}} - 3$	$\cdots$			

If you've worked a bunch of these examples by hand, you can see the general pattern. The first number in each row is the second number in the row above it. Each successive number in a row is found by looking at the previous number and going that many steps ahead in the row above.

The fifth row thus begins with $A(5,0) = 65533$. The second number, $A(5,1)$ requires that you evaluate a tower of exponents of 2 having height 65536, et cetera. We can convert this fantastically rapidly growing function of two variables into a single-variable function as follows: $\mathcal{A}(n) = A(n,n)$.

7 An Alternative Notation

If we examine the rows in the tables of values of Ackermann's function, we can see a pattern of growth. For Ackermann's function, the rows always seem to involve 2 (multiplication by 2, powers of 2, towers of powers of 2, et cetera). The other annoying thing about this function are the extra "-3" values that appear in the equations.

Ackermann's function is interesting for historical reasons, but if we're simply interested in large numbers, there's a cleaner and more general way to represent this type of number. We'll do it as follows, following Knuth and Conway:

$$\begin{aligned} nm &= n+n+\cdots+n \quad (m \text{ copies of n}) \\ n\uparrow m &= nnn\cdots n = n^m \quad (m \text{ copies of n}) \\ n\uparrow\uparrow m &= n\uparrow n\uparrow n\cdots n\uparrow n \quad (m \text{ copies of n}) \\ n\uparrow\uparrow\uparrow m &= n\uparrow\uparrow n\uparrow\uparrow n\cdots n\uparrow\uparrow n \quad (m \text{ copies of n}) \\ n\uparrow\uparrow\uparrow\uparrow m &= n\uparrow\uparrow\uparrow n\uparrow\uparrow\uparrow n\cdots n\uparrow\uparrow\uparrow n \quad (m \text{ copies of n}) \end{aligned}$$

As was the case with exponents, we will evaluate the expressions above from right to left. The first two lines are fairly straight-forward, but let's look at some examples from the third row:

$$\begin{aligned} 3\uparrow\uparrow 3 &= 3\uparrow 3\uparrow 3 = 3^{3^3} = 3^{27} = 7625597484987 \\ 2\uparrow\uparrow 4 &= 2\uparrow 2\uparrow 2\uparrow 2 = 2^{2^{2^2}} = 2^{16} = 65536 \end{aligned}$$

It's fairly clear that this third row corresponds to towers of exponents, but with the n and m, we can easily specify towers of any number to any height.

The fourth row also behaves much like the Ackermann function:

$$3\uparrow\uparrow\uparrow 3 = 3\uparrow\uparrow 3\uparrow\uparrow 3 = 3\uparrow\uparrow 7625597484987$$

which is an exponent tower of 3s of height 7625597484987. Let's just call this giant number X.

It is difficult even to think about the next row:

$$3\uparrow\uparrow\uparrow\uparrow 3 = 3\uparrow\uparrow\uparrow 3\uparrow\uparrow\uparrow 3 = 3\uparrow\uparrow\uparrow X$$

This will consist of a list of X 3s with two up-arrows between each pair. This huge number, $3\uparrow\uparrow\uparrow\uparrow 3$, comes up again in Section 9.2.

Using the up-arrow notation above, the Ackermann numbers $\mathcal{A}(n)$ are something like $1\uparrow 1$, $2\uparrow\uparrow 2$, $3\uparrow\uparrow\uparrow 3$, $4\uparrow\uparrow\uparrow\uparrow 4$, and so on.

8 Still Larger Numbers

The notation above represents a number using two other numbers plus certain number of up-arrows between them. We might as well represent the number of arrows as a number as well.

The following notation, following Conway and Guy, does just that. Let:

$$a \to b \to c$$

represent the number a followed by c up-arrows followed by b. Thus $3\to 4\to 5$ is the same as $3\uparrow\uparrow\uparrow\uparrow\uparrow 4$. Using this notation, the Ackermann numbers (as described in the previous section) look like $n\to n\to n$.

What we would like to do is describe what is meant by a chain of these right-pointing arrows. If there is just one (as in $n\to m$) this will mean n^m. The situation with two is described above, and if there are more than two, here is the meaning:

To evaluate

$$a\to b\to c\to\cdots\to x\to y\to z+1 \tag{2}$$

check the value of z, If $z=0$ then Equation 2 is the same as

$$a\to b\to c\to\cdots\to x\to y.$$

Otherwise the value depends on y. The following lines indicate the value for $y=1,2,3$, et cetera:

$$\begin{aligned} y=1: &\quad a\to\cdots\to x\to z \\ y=2: &\quad a\to\cdots\to x\to(a\to\cdots\to x)\to z \\ y=3: &\quad a\to\cdots\to x\to(a\to\cdots\to x\to(a\to\cdots\to x)\to z)\to z \end{aligned}$$

These numbers are *huge*. Consider $3 \to 3 \to 3 \to 3$:

$$\begin{aligned} 3 \to 3 \to 3 \to 3 & \\ &= 3 \to 3 \to (3 \to 3 \to (3 \to 3) \to 2) \to 2 \\ &= 3 \to 3 \to (3 \to 3 \to 27 \to 2) \to 2 \\ &= 3 \to 3 \to (3 \to 3 \to (\cdots)) \to 2 \end{aligned}$$

where the "$(\cdots)$" represents a 27-deep nesting of $(3 \to 3)$. You can imagine that by the time this whole thing is expanded, there will be an absolutely mind-boggling number of up-arrows.

9 Famous Large Numbers

In this section we'll consider a few historically famous large numbers.

9.1 Skewes' Number

When the author was a child, various math books claimed that Skewes' number was the largest that had ever come up in a "practical" sense, meaning that it came up in the proof of some important result.

If you don't understand the following details, don't worry too much, but basically the idea is this. Let $\pi(x)$ denote the prime number counting function. It is defined to be the number of prime numbers less than or equal to x. So $\pi(3) = 2$, $\pi(9) = 4$, et cetera.

Let $li(x)$ be the logarithmic integral function[1]:

$$li(x) = \int_0^x \frac{\mathrm{d}t}{\ln t}.$$

Anyway, for "small" numbers, $\pi(x) - li(x) < 0$, but Littlewood proved that this is not always the case, and that in fact, as $x \to \infty$, the sign of the expression on the left changes infinitely often. Skewes' number is is an upper bound on the smallest x such that $\pi(x) - li(x) \geq 0$, but Littlewood's proof contained no estimates for what that value of x might be.

Skewes, Littlewood's student, proved in 1933 that this number must be less than:

$$e^{e^{e^{79}}} \approx 10^{10^{10^{34}}},$$

Assuming that the Riemann hypothesis is true. This is sometimes called the "first Skewes' number, and was the largest value to appear in a mathematical proof for a long time. The second Skewes' number is even larger, and it was his best result assuming that the Riemann hypothesis is false. The second number is approximately:

$$10^{10^{10^{1000}}}.$$

Much better bounds are known today, namely $1.397162914 \times 10^{316}$, but Skewes was the first to find a bound.

9.2 Graham's Number

Graham's number is more modern, and dwarfs Skewes' number. It is also an upper bound, but as we shall see, it is a terrible upper bound. Before we describe it, we need to take a short digression to graph theory.

[1]The function as described here has a singularity at $x = 1$ if $x >= 1$ so we assume that the value is the Cauchy principal value, for all the nit-pickers out there:

$$li(x) = \lim_{\epsilon \to 0} \Big(\int_0^{1-\epsilon} \frac{\mathrm{d}t}{\ln t} + \int_{1+\epsilon}^{x} \frac{\mathrm{d}t}{\ln t} \Big)$$

A standard problem that appears in almost every introduction to graph theory is this:

Suppose there is a set of six people, and every pair either knows each other or does not know each other. Show that there is either a set of three people, all of whom know each other, or a set of three people, none of whom know each other.

The problem is usually solved by drawing a mathematical graph with six vertices that represent the six people, and each pair of vertices is connected with a line. The line is red if the two people know each other, and blue if they do not. The problem reduces to showing that there must either be a triangle with all red lines or a triangle with all blue lines. Suppose that's not the case. Consider any point P, and there will be five lines connecting it to the other points. At least three of these lines are red, or three of them are blue. Suppose there are at least three red lines connecting P to points Q, R and S. (The situation is similar for three blue lines.) Then to avoid any pure red triangles, the lines QR, RS and SQ must all be blue. But then we have a pure blue triangle QRS, so we have a contradiction.

With fewer than six people, it is possible to have no sets of three acquaintances or non-acquaintances, so six are required. This number six is called a "Ramsay number" for this problem. It turns out that the question can be turned around a little as follows: "What is the smallest number of people required so that there is at least a set of n acquaintances or a set of n non-acquaintances? Whatever that number is would be called the Ramsay number for n for this sort of problem. Even for this problem, for relatively small n, the exact Ramsay number is unknown, but it is known that some such number exists.

There are hundreds of situations like this related to bounds on sizes of graphs with colored vertices such that some condition holds, and that's exactly what Graham's number is. Here is the problem that defines Graham's number.

Given an n-dimensional hypercube, connect every pair of vertices to obtain a complete graph on 2^n vertices. Color these edges with two different colors. What is the smallest value of n for which every possible coloring contains a single-colored complete sub-graph with 4 vertices that lies in a plane?

Again, if you don't understand the exact problem, that's unimportant; the key thing is to see that it is similar to our first example, and this is just a Ramsay number for sub-graphs of size 4.

Graham proved that this number must be smaller than the number that we call Graham's number, and that number is huge. Using the up-arrow notation introduced in Section 7 (where we, in fact, looked at the value of g_1 as defined below) define:

$$\begin{aligned} g_1 &= 3 \uparrow\uparrow\uparrow\uparrow 3 \\ g_2 &= 3 \uparrow^{g_1} 3 \\ \ldots &= \ldots \\ g_{64} &= 3 \uparrow^{g_{63}} 3 \end{aligned}$$

The number $G = g_{64}$ is Graham's number. By the way, an exponent on the "$\uparrow$" symbol means that the symbol is repeated that many times. Using Conway's notation, we have:

$$3 \to 3 \to 64 \to 2 < G < 3 \to 3 \to 65 \to 2.$$

What is sort of amazing about Graham's number is that as of now, it is the best known bound for this particular Ramsay number. But the best lower bound at the time this article was written is 11. That's right: the number 11, so the unknown true bound may be as small as 11, and as large as Graham's number G.

9.3 The Busy Beaver Function

Here's one more interesting function that seems to generate absolutely huge numbers, but one of the things that makes it so interesting is that there is no effective way to compute it. An effective computation is one that is known to halt with the correct answer in a finite number of steps. The Busy Beaver function is not one of these.

Again, to describe the Busy Beaver function, we'll need to take a short digression to talk about Turing machines, first described by Alan Turing. We will describe here a very simple Turing machine.

Imagine a machine with a certain number n of internal states, one of which is the initial state, and an additional "halted" state, so there are actually $n + 1$ states. If the machine ever gets to the "halted" state, computation ceases.

At any point, the machine is "looking at" one position on an arbitrarily long tape which initially contains zeros in every position. Each position can contain either a zero or a one. The machine has a set of instructions, and each instruction tells it, for every internal state what to do if it is looking at a zero or looking at a one. The "what to do" consists of three things:

1. Whether to write a zero or one into the current position. (To leave the value at the position the same, just write the value that is there.)
2. Whether to move left or right one position along the tape. (There has to be movement.)
3. What the new state of the machine will be after the move.

If there are n states, there are only a finite number of Turing machines. The instructions have to exist for $2n$ conditions, and there are $4n+4$ actions that the machine can take for each of these conditions, so there are at most $(4n + 4)^{2n}$ machines.

Some of them halt instantly, and some run forever. As a machine that runs forever, imagine a machine that always stays in the initial state, writes a one, and then moves to the left.

Let's look at an example of a very simple Turing machine with just two states (three, including the "halted" state). We will label the states with letters: A, B, C, et cetera, and we will use H for the "halted" state. Let's also assume that the machine begins in state A. We will label the movements R and L for "move right" and "move left," respectively, and we will indicate with a 0 or a 1 whether to write a zero or a one at the current position before moving. Here is a simple two-state machine:

	0	1
A	$1RB$	$1LB$
B	$1LA$	$1RH$

Here's the interpretation: If the machine is in state A and sees a zero, write a one, move right, and change to state B. A macine in state A that sees a one will write a one (in other words, leaving the tape unchanged), will move left, and will change to state B. The second row similarly describes what the machine will do if it is in state B and sees a zero or a one.

Let's follow the action of this machine one step at a time.

start	$\cdots$	0	0	0	0	$0A$	0	0	0	0	$\cdots$
1	$\cdots$	0	0	0	0	1	$0B$	0	0	0	$\cdots$
2	$\cdots$	0	0	0	0	$1A$	1	0	0	0	$\cdots$
3	$\cdots$	0	0	0	$0B$	1	1	0	0	0	$\cdots$
4	$\cdots$	0	0	$0A$	1	1	1	0	0	0	$\cdots$
5	$\cdots$	0	0	1	$1B$	1	1	0	0	0	$\cdots$
6	$\cdots$	0	0	1	1	$1H$	1	0	0	0	$\cdots$

At the start, we have a tape filled with zeros, the machine in state A, and it is looking at one of the zeroes. In the chart above, the state is written in the tape position the machine is looking at. Since the state is A and there's a zero in that slot, the table that describes the machine says to write a 1 and move to the right. The second line in the table above shows the situation at that point. Follow along and see that after six steps the machine finally halts and at that point has written four 1's.

Here is an exercise to see if you understand how these machines work. Below is a three-state machine. Use the same representation as above to see how many 1's are written by this machine before it halts:

	0	1
A	$1RB$	$1RH$
B	$0RC$	$1RB$
C	$1LC$	$1LA$

Try to do this yourself, but there is a solution in Section 14.

Among all the machines with n states that do eventually halt, one of them must halt with the longest string of ones? The length of that string of ones is the output for the Busy Beaver function of n, usually written $\Sigma(n)$.

Here are some known facts:

- $\Sigma(1) = 1$.
- $\Sigma(2) = 4$. This can be proved by enumeration of all possible machines.
- $\Sigma(3) = 6$. This is *not* easy to prove.
- $\Sigma(4) = 13$.
- For $n > 4$, nobody knows $\Sigma(n)$. $\Sigma(5) \geq 4098$, and $\Sigma(6) \geq 3.514 \times 10^{18276}$. As n increases, these numbers grow astronomically.

Here is a web page that lists some results of the "Busy Beaver Competition," including descriptions of the record-holding machines:

`http://www.logique.jussieu.fr/~michel/bbc.html`

In the next section we will give an example of a fairly simple machine (that could easily be converted into a Turing machine) that generates surprisingly large numbers.

10 Goodstein's Theorem

10.1 Hereditary base-k notation

To state Goodstein's theorem, we first need to describe what is meant when we express a number in hereditary base-k notation. Let us begin using base-2 as an example.

The usual way to write the number 143 in base-2 is:

$$143 = 2^7 + 2^3 + 2^2 + 2 + 1,$$

with the understanding that we could write 2 as 2^1 and 1 as 2^0.

But if we are trying to avoid using numbers larger than 2, we'd like to get rid of the 7 and the 3 that appear in the exponents above, so we can always write $7 = 2^2 + 2 + 1$ and $3 = 2^1 + 1$:

$$143 = 2^{2^2+2+1} + 2^{2+1} + 2^2 + 2 + 1.$$

The expansion above represents 143 in hereditary base-2 notation.

There is nothing special about base-2; if we're interested in base-k, we'd like to use the variable k only when it requires an exponent, and otherwise use only values ranging from 0 to $k - 1$. Let's try the same number, 143, in hereditary base-3:

$$143 = 3^4 + 2 \cdot 3^3 + 2 \cdot 3 + 2 = 3^{3+1} + 2 \cdot 3^3 + 2 \cdot 3 + 2.$$

If this idea isn't perfectly clear, there will be many other examples in this section that display expansions of numbers using various hereditary bases.

10.2 Goodstein sequences

A Goodstein sequence begins with any integer g and is constructed as follows:

1. Set n, the current value of the sequence, to g. Set k, the current hereditary base, to 2.
2. if $n = 0$ the sequence terminates; otherwise:
3. Write the current element n of the sequence in hereditary base-k. The new value of n, and the next element of the sequence, is obtained by changing all occurrences of k to $k + 1$ in this expression, and then subtracting one from that value.
4. Increase the value of k by 1, and go to step 2 above.

Let's step through an example: we'll construct the first few terms of the Goodstein sequence beginning with the number 4:

Term	Hereditary notation	Value	Next term
1	2^2	4	$3^3 - 1$
2	$2 \cdot 3^2 + 2 \cdot 3 + 2$	26	$2 \cdot 4^2 + 2 \cdot 4 + 2 - 1$
3	$2 \cdot 4^2 + 2 \cdot 4 + 1$	41	$2 \cdot 5^2 + 2 \cdot 5 + 1 - 1$
4	$2 \cdot 5^2 + 2 \cdot 5$	60	$2 \cdot 6^2 + 2 \cdot 6 - 1$
5	$2 \cdot 6^2 + 6 + 5$	83	$2 \cdot 7^2 + 7 + 5 - 1$
6	$2 \cdot 7^2 + 7 + 4$	109	$2 \cdot 8^2 + 8 + 4 - 1$
7	$2 \cdot 8^2 + 8 + 3$	139	$2 \cdot 9^2 + 9 + 3 - 1$
8	$2 \cdot 9^2 + 9 + 2$	173	$2 \cdot 10^2 + 10 + 2 - 1$
9	$2 \cdot 10^2 + 10 + 1$	211	$2 \cdot 11^2 + 11 + 1 - 1$
10	$2 \cdot 11^2 + 11$	253	$2 \cdot 12^2 + 12 - 1$
11	$2 \cdot 12^2 + 11$	299	$2 \cdot 13^2 + 11 - 1$

To make sure you understand what is going on, you should verify by hand that the Goodstein sequences beginning with 1, 2 and 3 are: $\{1, 0\}$, $\{2, 2, 1, 0\}$ and $\{3, 3, 3, 2, 1, 0\}$, respectively. If you'd like, try to show that the first few terms for the Goodstein sequence beginning with 5 are: $\{5, 27, 255, 467, 776, ...\}$.

The Goodstein sequence beginning with 4 increases, but apparently not too rapidly. To show what usually occurs, let's look at the sequence beginning with 19 and keep in mind that 19 is a fairly small number:

Term	Hereditary form	Value
1	$2^{2^2} + 2 + 1$	19
2	$3^{3^3} + 3$	7625597484990
3	$4^{4^4} + 3$	$\approx 1.3 \times 10^{154}$
4	$5^{5^5} + 2$	$\approx 1.8 \times 10^{2184}$
5	$6^{6^6} + 1$	$\approx 2.6 \times 10^{36305}$
6	7^{7^7}	$\approx 3.8 \times 10^{695974}$
7	$7 \times 8^{7\times8^7+7\times8^6+7\times8^5+\cdots+7\times8+7}$ $+7 \times 8^{7\times8^7+7\times8^6+7\times8^5+\cdots+7\times8+6}$ $+7 \times 8^{7\times8^7+7\times8^6+7\times8^5+\cdots+7\times8+5} + \cdots$ $+7 \times 8^{8+2} + 7 \times 8^8 + 1$ $+7 \times 8^8 + 7 \times 8^7 + 7 \times 8^6 + 7 \times 8^5 + \cdots + 7 \times 8 + 7$	$\approx 6 \times 10^{15151335}$
8	$7 \times 9^{7\times9^7+7\times9^6+7\times9^5+\cdots+7\times9+7}$ $+7 \times 9^{7\times9^7+7\times9^6+7\times9^5+\cdots+7\times9+6}$ $+7 \times 9^{7\times9^7+7\times9^6+7\times9^5+\cdots+7\times9+5} + \cdots$ $+7 \times 9^{9+2} + 7 \times 9^{9+1}$ $+7 \times 9^9 + 7 \times 9^7 + 7 \times 9^6 + 7 \times 9^5 + \cdots + 7 \times 9 + 6$	$\approx 4.3 \times 10^{369693099}$

With even such a tiny number as 19 as the first term, this Goodstein sequence takes off like a rocket! What is surprising, perhaps, is that even our first example where we began at 4 does get quite large. In fact it eventually increases to $3 \cdot 2^{402653210} - 1$ after which it eventually decreases to zero, and thus terminates.

What is even more surprising is that if we examine the Goodstein sequence beginning with any integer, every one of them eventually converges to zero. For large starting values, this will take a long time and the numbers in the sequence will get incredibly large before they start down. For example, in the sequence beginning with 19, the eighth term is already larger than the largest value obtained by the sequence beginning with 4.

We will not present a completely formal proof of Goodstein's theorem which states that Goodstein sequences all terminate at zero, but we will provide an outline of the proof and some comments on it later, in Section 10.3.

Assuming that Goodstein's theorem is true: that all Goodstein sequences eventually terminate at 0, we can define the following function: Let $G(n)$ be the largest value obtained by the Goodstein sequence that begins with n. Thus $G(0) = 0$, $G(1) = 1$, $G(2) = 2$, $G(3) = 3$, $G(4) = 3 \cdot 2^{402653210} - 1$, and $G(5)$ is ever so much more vastly larger than that. As with functions like Ackermann's, what we have constructed is a very rapidly increasing function. This one, however, grows much faster than Ackermann's. $G(1000000)$ will be pretty big, and by now every reader of this article will be able to produce much larger numbers, given that they can use this function G.

10.3 Proof of Goodstein's theorem

This is not a formal proof, but if you know about infinite ordinal numbers and the fact that they are well-ordered, you will be able to expand the outline below to a formal proof. If you don't know about ordinal numbers, the outline attempts to show you the basic ideas.

Unfortunately, the construction of the infinite ordinal numbers requires a much more powerful theory (Zermelo-Fraenkel set theory) than what is usually required to prove theorems about the arithmetic of the natural numbers. Usually all that is required to prove almost every theorem you know about the natural numbers are the so-called "Peano postulates".

What is doubly unfortunate is that it is impossible to prove Goodstein's theorem from the Peano postulates, although the proof of the fact that such a proof is impossible is far beyond the scope of this article.

10.4 The infinite ordinals

The ordinals are basically an extension of the natural numbers through infinite values. You can get to all the natural number by starting with 0 and obtaining the next by adding 1 to the previous. If those are the only things you're allowed to do, then only the finite natural numbers are accessible:

$$\{0, 1, 2, 3, 4, 5, \ldots\}.$$

To obtain the infinite ordinals, we also allow you to construct a new ordinal that is just larger than any infinte sequence of ordinals previously obtained. The first infinite ordinal is called "omega" and is indicated by the Greek letter of the same name: "ω". Such ordinals are called "limit ordinals", of which ω is the first.

But once ω is allowed, since we are able to add one, we can obtain: $\omega + 1$, $\omega + 2$, $\omega + 3$ and so on. Since $\omega + n$ is a valid ordinal for any natural number n, we can include the limit of ω, $\omega + 1$, $\omega + 2$, $\ldots$, and it will be $\omega + \omega$, which is usually denoted by $\omega \times 2$, the second limit ordinal. Then of course we can get $\omega \times 2 + 1$, $\omega \times 2 + 2$, $\ldots$, and the limit of that sequence of ordinals will be the third limit ordinal: $\omega \times 3$.

In a similar way, we can obtain the limit ordinals $\omega \times 4$, $\omega \times 5$, $\ldots$, and we can then construct another limiting sequence:

$$\omega, \omega \times 2, \omega \times 3, \omega \times 4, \ldots.$$

The limit of that will be $\omega \times \omega = \omega^2$. Then we're back to $\omega^2 + 1$, $\omega^2 + 2$, $\ldots$, $\omega^2 + \omega$, $\omega^2 + \omega + 1$, $\ldots$, $\omega^2 + \omega \times 2$, $\ldots$, $\omega^2 \times 2$, $\omega^2 \times 2 + 1$, $\ldots$, ω^3, $\ldots$, ω^4, $\ldots$, ω^ω, $\ldots$, $\ldots$.

Obviously, we've skimmed over a lot, but all this can be done in a very formal and logical way.

10.5 Well-ordering of the ordinals

The amazing thing is that this set of infinite ordinal numbers is "well-ordered", and that means that any decreasing series of ordinals is of finite length: it is impossible to construct an infinite set of decreasing ordinals, no matter how large a one you start with.

The idea of well-ordering is obvious if we just look at the natural numbers: If you begin with some number and each successive number in the sequence is smaller, you must stop at zero after a finite number of steps. If you start with a million and start stepping down, you can't possibly take more than a million steps. Obviously you can make a decreasing sequence of natural numbers as long as you want, but every one is finite.

The infinite ordinals behave the same way. Let's look at some examples. First, suppose you start "infinitely far along", at ω itself. What is the next smaller ordinal in your sequence?

Well, ω is the first infinite ordinal, so every smaller ordinal is finite. Once you take that first step down, to a million, or a billion, or to a googleplex, there are only a finite number of additional steps left to go. When you take that one step down, it is enormous.

So ω won't work as an ordinal from which you can make an infinite number of steps down. It's clearly pointless to start from $\omega + k$, where k is finite, since after at most k steps you'll be back at ω, and from there it's only a finite number of steps to the bottom. How about $\omega \times 2$?

Well, the first step down will take you to $\omega + k$, where k finite, so that's no good. There are similar problems with $\omega \times 3$, $\omega \times 4$, or $\omega \times k$, where k is finite: it's only a finite number of steps down to get rid of each multiple of ω.

How about ω^2? Well, the first step down has to have a largest term of the form $\omega \times k$, so that's no good, either. We have to be a little careful here, since here's a number smaller than ω^2:

$$\omega \times 1000 + \omega \times 999 + ... + \omega \times 2 + \omega + 1000000,$$

but the expression will only have a finite number of terms, and they must be knocked off, one by one.

In fact, you can sort of imagine an inductive proof[2] that there are always a finite number of steps to the bottom. Suppose that the ordinal κ is the first one from which there is an infnte descending sequence. If κ is of the form $\lambda + 1$, then the first step down is to λ, from which there are only a finite number of steps to the bottom or to a number even smaller than λ, from which there are also at most a finite number of steps down. Similarly, if κ is one of the "limit ordinals", the first step down will be to one from from which there are only a finite number of steps to zero, so κ will share that property.

Try to figure out what might happen with ω^ω – a pretty large ordinal[3]. You can play around with this a bit if you like, and perhaps that will make things clearer.

10.6 Proof of Goodstein's theorem

Anyway, if you believe that the infinite ordinals are well-ordered, the proof that Goodstein sequences all terminate at zero is not too hard. What we will do is replace every term of such a sequence in hereditary notation by an ordinal number that is clearly larger than it. We will show that the sequence of ordinals thus obtained is a decreasing sequence, so the Goodstein sequence will be dominated by a sequence of ordinals that we know tends to zero, and hence the dominated sequence will also tend to zero.

The dominating ordinal is simple: just replace all occurrences of the base by ω. So for example, in our example of the Goodstein sequence beginning with 19, the first few terms, and their dominating ordinals are:

[2] This is actually not a proof by finite induction that you're probably familiar with, but rather a proof by "transfinite induction". It is, however, very similar to the usual proofs by finite induction.

[3] Well, "pretty large" is optimistic: almost all ordinals, of course, are larger.

Term	Hereditary form	Ordinal
1	$2^{2^2}+2+1$	$\omega^{\omega^\omega}+\omega+1$
2	$3^{3^3}+3$	$\omega^{\omega^\omega}+\omega$
3	$4^{4^4}+3$	$\omega^{\omega^\omega}+3$
4	$5^{5^5}+2$	$\omega^{\omega^\omega}+2$
5	$6^{6^6}+1$	$\omega^{\omega^\omega}+1$
6	7^{7^7}	ω^{ω^ω}

On the next step, of course, we take a *huge* step down in the ordinals. The largest term in the exponent of the bottom ω contains only finite powers of ω; not ω^ω. If you want to see what it looks like, just substitute ω for every 8 in the expansion of term 7 that we did in the previous section for the Goodstein sequence beginning with 19.

10.7 Behavior of Goodstein sequences

Every Goodstein sequence behaves similarly in the following sense: as long as the base itself appears in the expansion with even a multiple of 2, the sequence will increase. That's because when the base increases for the next term, the increase will be multiplied by at least 2, and only 1 is subtracted.

Goodstein sequences finally reach a point where they have the following form:

$$B \times 1 + k,$$

where B is the base and k is a (usually huge) constant. At this point, the sequence stays constant for k steps, since at each step, although k is reduced by 1, B is increased by 1. Finally, the term looks like $B \times 1 + 0$ and after 1 is subtracted from that, the base will not appear in the expression; only a constant will. After this, the sequence steps down to zero, one unit at a time.

So every Goodstein sequence increases (often incredibly rapidly at first), and keeps increasing until it obtains the form above. Then it is constant for a long time, after which it reduces to zero by one at each step.

11 Are We Cheating?

In the previous sections we have presented some very large numbers in a very compact way, but since each of these methods required a few paragraphs of explanation, aren't we really cheating? Shouldn't we count the description of the method?

If you have a contest to "describe the largest number you can using n symbols", you've got to say at the beginning what the allowable symbols are. Otherwise you can imagine a many-page description of a huge number where the last sentence is, "Call that number X." Then the description is a single character, "X", and that's not really fair.

One good way to structure such a contest might be to describe a virtual calculator that has certain buttons on it, and the contest is something like, "Make the largest number in the display you can with at most n button-presses." If the calculator has a factorial button on it, then $9!$ might be the winning candidate for the two-button-press version of the contest.

If the calculator is programmable, then you would need to count the strokes necessary to create the program that is run with an additional button press, et cetera.

In any case, to end this article with an interesting example, imagine a calculator that has a bunch of variables in it, (say, A, B, C, ..., all of which are initially zero), and only allows the following sorts of operations:

- $A++$ Increment the value of variable A by 1.

- $A--$ Decrement the value of variable A by 1 unless $A = 0$.
- $A :: \mathrm{x}$ If $A \neq 0$ go to line number x; otherwise, continue to the next line.

Here is a 23 line program written in the language above. Try to figure out how large the numbers A, B, C, D, and E get by the time the program terminates. It is fairly impressive, given that essentially the only allowable operations are "add one", "subtract one", and "branch if non-zero".

1:	$E++$	9:	$A++$	17:	$D--$
2:	$E++$	10:	$B--$	18:	$D :: 5$
3:	$E++$	11:	$B :: 8$	19:	$D++$
4:	$E++$	12:	$C--$	20:	$C--$
5:	$B++$	13:	$C :: 5$	21:	$C :: 19$
6:	$A--$	14:	$C++$	22:	$E--$
7:	$A :: 5$	15:	$A--$	23:	$E :: 5$
8:	$A++$	16:	$A :: 14$		

12 Even Larger Numbers

In a sense, all the numbers described in previous sections are tiny compared to what we will encounter here. Up to this point, the text in this article contains fewer than 6000 words, and think of all the different large numbers we have defined.

As a closing idea, consider the function $F(n)$ which is defined to be "the largest integer than can be defined in n or fewer English words." This is obviously very sloppy, but the idea is not so bad. We could define a perfectly formal language and make a definition like this one precise, but that requires a lot of work. Let us just stick with English for now.

What is $F(1)$? Perhaps "googolplex"? It's at least that big. We know that $F(6000)$ is far more than sufficient to describe all the numbers here so far, and it's clear that *far* larger numbers could have been defined. But with this F, we can do all the tricks again. What if $G(1) = F(1)$, and $G(n) = F(G(n-1))$, for example? We already know that $G(2)$ is at least as big as the largest number that could be described with a googolplex words.

These numbers, of course, will not be computable in the same sense that the Busy Beaver numbers in Section 9.3 are not. But they exist, and are guaranteed to be unimaginably huge.

It's always good to keep in mind, however, what is sometimes called "The Frivolous Theorem of Arithmetic": Almost all numbers are very, very, very large. In fact, even considering all the particular huge numbers listed in this article, almost all numbers are much, much larger than any of them....

13 Pedagogical Notes

For younger students (in middle school, for example), it's probably a bad idea just to write down the definition of Ackermann's function without an introduction.

This is a good time to do a review of functional notation. For example, just review some standard, simple function definitions, emphasizing the idea that the definition provides a rule (or rules) for determining the output given the input.

If, for example, the function is defined as:

$$f(x) = x^2 - 3,$$

then to evaluate f for any particular input, you simply substitute the input value for the x on the right-hand side of the equation above and evaluate it. In this example, you simply square the input value and subtract 3, so $f(5) = 5^2 - 3 = 25 - 3 = 22$, or $f(20) = 20^2 - 3 = 400 - 3 = 397$.

Ackermann's function, unfortunatley, is more complex in two ways. First, it's a function of *two* variables, and second, it is defined recursively. Rather than jump right in, introduce the two ideas in two steps. First, look at functions of more than one variable.

An example might be this: The ticket price at a movie theatre is \$10 for an adult, and \$6 for a child. What we seek is a formula to tell us the admission cost for a mixed group of adults and children. Here is the answer:

$$f(a, c) = 10a + 6c,$$

where $f(a, c)$ represents the admission cost for a group consisting of a adults and c children. Make it clear why this is the correct function, and why, to determine the numerical output, you need to know both the number of adults and children.

But even without complete information, you can simplify the formula with partial information. For example, suppose you know that the group is going to consist of 4 adults, but you don't yet know the number of children. If we let c stand for the (as yet unknown) number of children, the total cost will be:

$$f(4, c) = 10 \cdot 4 + 6c = 40 + 6c.$$

Using this simpler function, all you need to do is plug in the value of c to obtain the final cost.

The other type of function, a so-called "recursive function" is a little more interesting. Here is perhaps the best example to use, since the function is already known to the students, but probably never in the form shown below. The function F, defined below, is defined for all positive integers.

$$F(n) = \begin{cases} 1 & : \text{ if n} = 1 \\ n \cdot F(n-1) & : \text{ if n} > 1 \end{cases} \tag{3}$$

The only time you know the answer immediately is if the input value happens to be zero: $F(1) = 1$. But what is the value of $F(4)$? Well, you just follow the rule stated in the second line of Equation 3. The input value, 4, is not equal to zero, so you must use the second rule:

$$F(4) = 4 \cdot F(3). \tag{4}$$

This doesn't seem to help much since we don't know the value of $F(3)$, but if we have faith, we can just reuse the definition. (A mathematician would say that we are using Equation 3 recursively.) To evaluate $F(3)$ we first note that the input value, 3, is not zero, so $F(3) = 3 \cdot F(2)$. Substituting this value for $F(3)$ into Equation 4 yields:

$$F(4) = 4 \cdot (3 \cdot F(2)), \tag{5}$$

and we're now left with the problem that we don't know the value of $f(2)$. But we can reuse Equation 3 yet another time to obtain $F(2) = 2 \cdot F(1)$, so the original Equation 4 now looks like:

$$F(4) = 4 \cdot (3 \cdot (2 \cdot F(1))),$$

and now we're in good shape because we know that $F(1) = 1$, so we have:

$$F(4) = 4 \cdot (3 \cdot (2 \cdot 1)) = 4! = 24.$$

Try evaluating $F(6)$ using the same method, but as soon as you get to:

$$F(6) = 6 \cdot (5 \cdot F(4)),$$

you can tell the kids that we don't need to go farther since we already worked out the value of $F(4)$, which was $4! = 24$. Thus:

$$F(6) = 6 \cdot 5 \cdot 4! = 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 6!$$

Finally, point oiut that for this recursive function at least (and this will usually be the case), the easiest way to figure out what's going on is not to start with large values, although that will work, but to start

with small ones. For this example, we know the value of $F(1)$ immediately, so work on $F(2)$ which we find to be $2 \cdot 1$. Once we know the value of $F(2)$, we can see that $F(3) = 3 \cdot F(2) = 3 \cdot 2 \cdot 1 = 3!$. Then we will see that $F(4)$ can easily be evaluted in terms of $F(3)$, and so on. After just a few steps like this, it's easy to see (and to prove, if you wish) that $F(n) = n!$.

Notice that the recursive definition is in some ways nicer than the usual formula for $n!$:

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1,$$

since we never need to write the somewhat vague "$\cdots$".

In any case, after this introduction, you can look at the more complicated Ackermann's function which is both a function of two variables *and* recursive.

14 Turing Solution

Here is the simulation for the following Turing machine. This machine, in fact, is an example of a three-state machine that achieves the longest row of 1's before halting. Here's the machine:

	0	1
A	$1RB$	$1RH$
B	$0RC$	$1RB$
C	$1LC$	$1LA$

And here's the simulation:

start	$\cdots$	0	0	0	0	$0A$	0	0	0	0	$\cdots$
1	$\cdots$	0	0	0	0	1	$0B$	0	0	0	$\cdots$
2	$\cdots$	0	0	0	0	1	0	$0C$	0	0	$\cdots$
3	$\cdots$	0	0	0	0	1	$0C$	1	0	0	$\cdots$
4	$\cdots$	0	0	0	0	$1C$	1	1	0	0	$\cdots$
5	$\cdots$	0	0	0	$0A$	1	1	1	0	0	$\cdots$
6	$\cdots$	0	0	0	1	$1B$	1	1	0	0	$\cdots$
7	$\cdots$	0	0	0	1	1	$1B$	1	0	0	$\cdots$
8	$\cdots$	0	0	0	1	1	1	$1B$	0	0	$\cdots$
9	$\cdots$	0	0	0	1	1	1	1	$0B$	0	$\cdots$
10	$\cdots$	0	0	0	1	1	1	1	0	$0C$	$\cdots$
11	$\cdots$	0	0	0	1	1	1	1	$0C$	1	$\cdots$
12	$\cdots$	0	0	0	1	1	1	$1C$	1	1	$\cdots$
13	$\cdots$	0	0	0	1	1	$1A$	1	1	1	$\cdots$
14	$\cdots$	0	0	0	1	1	1	$1H$	1	1	$\cdots$