Puzzles and Paradoxes in Mathematical Induction

Adam Bjorndahl

Contents

1 Flavor

Mathematical induction is a beautiful tool by which one is able to prove infinitely many things with a finite amount of paper and ink. It works by exploiting underlying structure: a complex and unwieldy problem can sometimes be broken apart along its fault lines so as to leave behind many smaller problems, each of which is more easily solved. Induction is one method of finding such fault lines and organizing the smaller pieces of a larger problem. Often, the simpler structure of the small pieces permeates the whole, and a complicated structure can be seen to operate based on the same simple rules that govern its pieces. This can lead to results that are both powerful and counter-intuitive.

Induction is only one of many techniques through which one may attempt to wrestle with infinity in finite terms (which is to say: with home field advantage), but it holds a rather distinguished position in mathematics. Conveniently, it requires very little background knowledge to learn, and for this reason it is often taught in high school and could reasonably be included in an elementary school curriculum. Its home is in the **natural numbers** : $1, 2, 3, 4, \ldots$, which are, barring geometrical objects, arguably the most intuitive of all mathematical objects. Despite its apparent simplicity, its use in contemporary mathematics is widespread. But perhaps most tellingly, a casual lunchtime conversation with my colleagues about induction revealed that everyone seemed to have their own "induction story", a tale of their first encounter with or first appreciation of mathematical induction. It is clear that induction holds a special place in the mathematician's heart, and so it is no surprise that it can be the source of so much beauty, confusion, and surprise.

2 Crunch

Let's spend a moment and get clear on what induction is and how it works in concrete terms. One often wishes to prove a certain statement true, where that statement says something about infinitely many things. For example:

Any even number squared is divisible by four.

One can check whether this holds for the first few even numbers: $2^2 = 4$, yes; $4^2 = 16$, yes; $6^2 = 36$, yes. But eventually your arm is going to get tired of writing and you'll probably be nowhere close to checking whether or not 272^2 is divisible by 4, let alone 3862456^2. Fortunately there is a easy way of rolling all this work into one:

Suppose x is an even number. Then, since being 'even' is the same as being twice some number, we know that $x = 2y$, where y is some natural number. We are interested in $x^2 = (2y)^2 = 4y^2$. The important equality here is $x^2 = 4y^2$, which says that the square of x is exactly 4 times some other number (namely

y^2, but that's not important). Since being divisible by 4 is the same as being 4 times some number, we have therefore proved that x^2 is divisible by 4.

The crucial point about this proof is that is works simultaneously to prove the result for *every* even number. We were able to do this by taking advantage of the fact that the *method* of proof doesn't depend at all on which even number you start with! So we abstracted away from the specifics by letting x stand for *some* (unnamed) even number and then proving the result for x, which must then hold true for *every* even number.

The preceding example was rather trivial, so it should at least convince you that sometimes getting infinitely many results out of a finite amount of work is not the lofty enterprise that it sounds. Mathematical induction works on the same principle of collapsing repetitive computations into a single, abstract computation which can then be applied again and again. But the implementation of induction is a bit different from the example we just saw.

Suppose that we wish to prove some result, call it R, about all the natural numbers, from 1 and up. And suppose also that we cannot find a way to do this directly; in other words, letting x stand for some natural number and then trying to prove R for x hits a dead end. We might find ourselves in the following situation:

"I can show that R is true for 1, easy enough, but that's only one step among infinitely many. But wait, I can show R is true for 2 now also, by making use of the fact that R is true for 1. And now I can show that R is true for 3, using the fact that it's true for 1 and 2."

This looks promising, but there are still an infinite number of 'steps' to take: from 1 to 2, then to 3, then to 4, and so on. The induction insight is in realizing that if the reasoning behind each of these 'steps' is the same, no matter which step it is, then instead of abstracting away from the *numbers*, we can abstract from the *steps*. More formally, an inductive proof has two stages:

1. *The Base Case.* Prove the desired result for the number 1.

2. *The Inductive Step.* Prove that *if* the result is true for the numbers 1 through n, then it is also true for the number $n + 1$.

The inductive step is proved by first assuming that the result is true for the numbers 1 through n, and then using this assumption to show that it is also true for $n + 1$. This reasoning can seem circular at first—after all, the whole point is to try to prove that the result is true, so how can we be allowed to *assume* this? The answer we've already seen: the reasoning starts from the base case, where we prove directly that the result is true for 1. Then we can apply the inductive step in the case $n = 1$ to deduce that *if* the result is true for 1 (which we just verified) *then* it is also true for 2. Now we know the result is true for 1 and 2, and so we can reapply the inductive step in the case $n = 2$ to deduce that if the result is true for 1 and 2 (which it is), then it is also true for 3. And so on. In this manner the truth of the result for every number can be established by starting at 1 and working our way up. Far from being circular, mathematical induction is a canonical example of linear reasoning.

Sometimes it happens that we are able to complete the induction step without the full assumption that the result holds for *all* the numbers 1 through to n. We will see several examples of this in the pages that follow, wherein we will only need to assume that the result is true for n in order to establish that it is also true for $n+1$.

Exercise 1 *Convince yourself that an inductive proof in this second style makes sense as a method of proving that something is true of all natural numbers.*

3 Appetizers

Before jumping in to some of the more mystifying applications of induction, let's take a look at how it works in some more straightforward (but by no means trivial) examples. Consider the following claim:

> The sum of the first n squares is equal to $(1/6)n(n+1)(2n+1)$.

Once again, this is easy enough to check by hand for the first few values of n:

$$1^2 = 1 \quad \text{and} \quad (1/6)(1)(1+1)(2\cdot 1+1) = (1/6)(2)(3) = 1,$$

yes;

$$1^2 + 2^2 = 1 + 4 = 5 \quad \text{and} \quad (1/6)(2)(2+1)(2\cdot 2+1) = (1/6)(2)(3)(5) = 5,$$

yes;

$$1^2+2^2+3^2 = 1+4+9 = 14 \quad \text{and} \quad (1/6)(3)(3+1)(2\cdot 3+1) = (1/6)(3)(4)(7) = 14,$$

yes. So the claim holds up to the sum of the first 3 squares, but already things are starting to get cumbersome, and this method of checking by hand has no hope of yielding a proof for *all* such sums. We need to find a better way to go about things.

We might try the same method here as was used in proving that all squares of even numbers are divisible by four. Let n be any number, and consider the sum of the first n squares:

$$1^2 + 2^2 + 3^2 + 4^2 + \cdots + (n-1)^2 + n^2.$$

If we could somehow add up this sum and show that the result is $(1/6)n(n+1)(2n+1)$ then we would be done. Problematically, there is no obvious way to *do* this addition (try it and see). We turn to induction.

First comes the base case $n = 1$: we must prove that the sum of the first 1 squares is equal to $(1/6)(1)(1+1)(2\cdot 1+1)$. This has already been done, above, and it was rather easy; it is often true that the base case of an inductive argument is easy or trivial.

Next comes the inductive step. We start by making our inductive hypothesis: assume that the claim is true for n. Our task is to show that it is also true for $n+1$. In other words, we need to show that the sum of the first $n+1$ squares is equal to

$$(1/6)(n+1)[(n+1)+1][2(n+1)+1], \tag{1}$$

which is just the original formula with n replaced everywhere by $(n+1)$.

Our inductive hypothesis amounts to the following equation:

$$1^2 + 2^2 + \cdots + (n-1)^2 + n^2 = (1/6)n(n+1)(2n+1).$$

Adding $(n+1)^2$ to both sides of the above equation yields

$$1^2 + 2^2 + \cdots + (n-1)^2 + n^2 + (n+1)^2 = (1/6)n(n+1)(2n+1) + (n+1)^2.$$

Notice that the left hand side of this equation is exactly the sum of the first $(n+1)$ squares! If we could show that the right hand side of this equation is equal to the formula labelled by (1), above, then we will have completed the induction step. Although it is not immediately apparent, the two formulas in question are in fact equal. The right sequence of algebraic manipulations reveals this to be so:

$$\begin{aligned}
(1/6)n(n+1)(2n+1) + (n+1)^2 &= (n+1)[(1/6)n(2n+1) + (n+1)] \\
&= (n+1)[(1/3)n^2 + (7/6)n + 1] \\
&= (n+1)\big([(1/6)n + (1/3)][2n+3]\big) \\
&= (1/6)(n+1)(n+2)(2n+3).
\end{aligned}$$

Exercise 2 *Verify the above equalities.*

The reader is invited to verify these calculations in more detail. Whatever detail is omitted here is done so for the sake of clarity: there are only so many lines of algebra one can reasonably be expected to read through before going cross-eyed. The best path to understanding is to work out the details for yourself, using the above as a guide.

This completes the induction and therefore finishes the proof: we have verified that the sum of the first n squares is always equal to $(1/6)n(n+1)(2n+1)$. It is worth noting, however, that I have left the origins of this formula a complete mystery. Induction proved quite useful in verifying that the given formula is the correct one, but how one might come to suspect that formula in the first place is another issue entirely. I refer the reader to [1], chapter 2, for a very pleasing exploration of this issue (the rest of the book is nice, too).

Next I want to give an example of induction used in a very different situation:[1] to solve a puzzle which at first glance doesn't seem to have much use for induction at all. The puzzle is a tiling puzzle. It asks:

[1] Thanks to Jonathan Needleman for introducing me to this neat example.

Can a chessboard minus a rook's square be tiled with triominos?

What? The first step to solving any problem is understanding the question, so we start with that. First, a 'triomino' is a two-dimensional figure made out of three equal squares glued together in the configuration pictured in Figure 1.

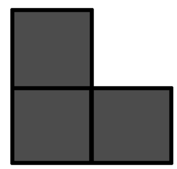

Figure 1: A triomino

A 'chessboard' is a more familiar object: it is an 8×8 square made by gluing together 64 smaller squares. Implicit in the question is the assumption that the squares that make up a triomino are the same size as the squares that make up the chessboard. A 'rook's square' is any one of the four corner squares on the chessboard. Finally, the 'tiling' that the question asks for is an arrangement of triominos on the chessboard such that every square of the chessboard (except for one corner square) is covered by a square from a triomino. No two triominos are allowed to overlap, and every triomino must be positioned so that it lies entirely over the chessboard. Can this be done?

The answer is yes, and if you can get your hands on some triominos (or make some yourself), after a little experimentation you'll be convinced of this. So we can ask a harder question:

Can every $2^n \times 2^n$ board, minus a corner square, be tiled by triominos?

The chess board example corresponds to the case $n = 3$. We can proceed in the general case by induction in a rather surprising way. That induction is useful here is perhaps not so surprising, given the nature of the claim we wish to prove (i.e. we want to prove the claim for $n = 1, 2, 3, \ldots$). But the geometrical aspect of this problem contrasts sharply with the previous example.

The base case for the induction is $n = 1$, so we must consider a $2^1 \times 2^1$ board with one of the corner pieces removed and determine whether this can be tiled by triominos. But of course it can be, since what remains of the board is precisely the same shape as a single triomino. Once again, the base case was easy.

Now the induction step. We assume the result is true for n and try to prove it true for $n+1$. This means that we can assume that the $2^n \times 2^n$ board, with one corner piece removed, can be tiled by triominos. See Figure 2.

Figure 2: A tiling of the $2^n \times 2^n$ board, minus the upper right corner square.

Note that the actual tiling pattern is not shown; at this point we need not be concerned with *how* the tiles are arranged, but merely that the board minus the corner is tiled in *some way*.

How can we use this tiling, our inductive hypothesis, to obtain the analogous result for a $2^{n+1} \times 2^{n+1}$ board? Before reading ahead to the solution, try to play around with the possibilities for a time. The answer is not complicated or difficult to understand, but it requires the right flashes of insight to come up with it.

The first important insight is that the $2^{n+1} \times 2^{n+1}$ board can be divided into exactly four $2^n \times 2^n$ boards by cutting it vertically and horizontally down the center lines. Our inductive hypothesis tells us that we know how to tile such boards with triominos, minus a corner square. So let that be done.

The second insight is best provided in picture form: we glue the four quarters back together in the configuration shown in Figure 3. This leaves one corner square (the one in the upper right) untiled, along with three central squares. But look! The three central squares that are untiled have been glued back together in exactly the shape of a triomino! Therefore, with the addition of one extra triomino in that conspicuous space, we have managed to completely tile the $2^{n+1} \times 2^{n+1}$ board, minus the square in the upper right corner. This completes the induction and finishes the proof.

Before moving on, it is worth noting a small corollary to the result we just proved. Since each triomino is composed of exactly three squares, any area that can be tiled with triominos must consist of exactly $3k$ squares, where k is the number of triominos used in the tiling.

Exercise 3 *Show that the converse of this statement is false. That is, show that there are areas consisting $3k$ squares, for some natural number k, that cannot be tiled with triominos.*

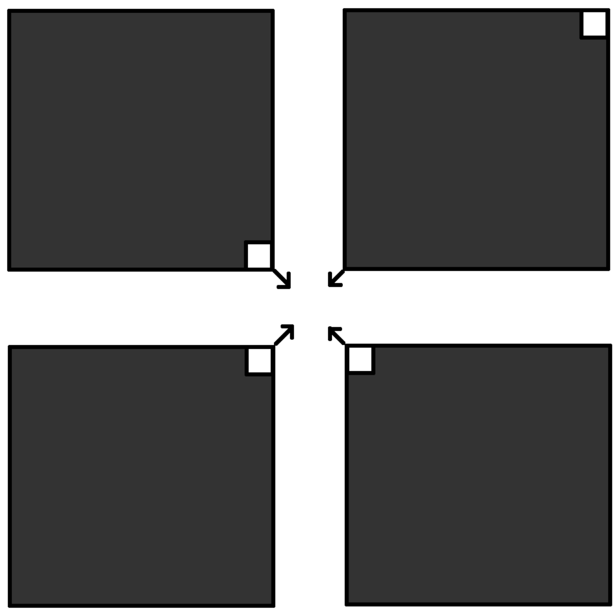

Figure 3: Gluing the four quarters back together.

The number of squares in a $2^n \times 2^n$ board is $2^n \cdot 2^n = 2^{2n}$. So a tiling of this board minus a corner square is a tiling of an area consisting of exactly $2^{2n} - 1$ squares. We can therefore deduce that $2^{2n} - 1 = 3k_n$ for some k_n (the subscript n is included to indicate that the number k_n depends on n). This isn't the most interesting result in the world, but it is not exactly obvious and we get it for free from our tiling proof.

Exercise 4 *Prove this result directly. That is, prove by induction on n without reference to tilings that for all natural numbers n the number $2^{2n} - 1$ is divisible by 3.*

Exercise 5 *Find an explicit formula for k_n in terms of n. (Hint: use the tiling proof or your answer to the previous exercise to guess the formula and then verify it.)*

4 Fly in the Soup

Now that we've seen how useful induction can be, I'd like to use it to establish a clear falsehood:

All horses are the same color.

The first hurdle to jump is figuring out how to make this an inductive argument. But that's not too hard; I will translate the original claim into the following equivalent form: for every natural number n, every group consisting of n horses is monochromatic (i.e. they are all the same color).

Now we can get our induction off the ground. The case $n = 1$ is obvious: every group consisting of 1 horse is monochromatic. Next comes the induction hypothesis: we are allowed to assume the result is true for n, and our job is to prove it true for $n + 1$. So consider a group of $n + 1$ horses. How can we show that they are all the same color? Well, our inductive hypothesis tells us that every group of n horses is monochromatic. So all we have to do is remove one horse, call him Paul, from our group of $n+1$ horses and consider the remaining group; call it G_1. There are n horses left in G_1, so by the inductive hypothesis they are all the same color. Now we have the problem that Paul is perhaps a different color than the rest. But this problem too can be overcome: simply remove a *different* horse from the original group. This leaves behind a new group of n horses which includes Paul; call it G_2. Again we apply our inductive hypothesis, this time to deduce that G_2 is monochromatic. But if Paul is the same color as all the horses in G_2, then Paul must be the same color as a horse in G_1, since every horse in G_2 *except* Paul is also in G_1. This then shows that Paul is the same color as *every* horse in G_1 (since G_1 is monochromatic), and so the original group of $n + 1$ horses is monochromatic. This completes the induction and thereby proves that all horses are the same color.

This is an excellent "proof" to show to people who have just learned about induction for the first time. Even veterans of the technique can sometimes be perturbed by this argument if they haven't seen it before in some form or another (it is quite common). What went wrong? In my experience, there are several 'stages of doubt' that one goes through in response to this seeming paradox, although the order may vary from person to person.

There is denial, something like, "Well, obviously the reformulation of the original claim isn't equivalent, so induction can't be applied." But it is equivalent. Since there are (presumably) only finitely many horses in existence, one can simply take n to be the total number of horses to regain the first claim out of the second one. Alternatively, one can show that the second claim implies the first by contrapositive: if the first claim is false, so that not all horses are the same color, then we should be able to find two horses of different colors; this gives a non-monochromatic group (with $n = 2$), contradicting the second claim. (Here we don't even rely on there being only finitely many horses).

Then there is a loss of faith in induction "Wait—you can't assume the result is true for n! That's circular!" But this assumption, the inductive hypothesis, is not the culprit. We have already discussed the logic underlying this assumption, and we have seen that induction is anything but circular.

Sometimes there is a stage of ridiculous doubt and wildly irrelevant claims: "Horses can't be grouped! And what's this about color? Horses have lots of colors, not just one!"

And perhaps, in the minds of those few whose adherence to the strictures of mathematics is unbreakable, there are even some fleeting moments of acceptance: "Maybe all horses *are* the same color. I mean, have you ever really *looked*?"

But in the end the fact remains that induction is not broken, yet horses do differ in color. Before reading on to the solution printed below, I suggest you let the problem torment you for a while.

A hint to what went wrong in the proof can be gleaned from solving the following simpler puzzle, which relies on the same misleading language to confuse the reader:

"How many pets do I have if all of them are dogs except two, all of them are cats except two, and all are parrots except two?"

This teaser comes from [2], where many others can be found and many hours can be spent on a joyful see-saw ride between befuddlement and insight.

The answer to this puzzle is that I have exactly three pets: one dog, one cat, and one parrot. The misleading use of the word 'all' to refer to only one thing is the source of the trouble here, as it is in the horse paradox. Let us return to the group G_2 and examine my assertion that "every horse in G_2 *except* Paul is also in G_1". True enough, but quite vacuous in the case where $n = 1$ and the group G_2 consists of Paul alone. But this was the crucial claim that allowed me to conclude that Paul is the same color as the other horses in the original group of $n + 1$. When $n = 1$, the original group is a group of 2, and the logic fails entirely, since G_1 and G_2 have no horses in common.

This is the hole in the argument. It is not a problem with induction, but the inductive form of the argument does allow it to be easily concealed. The inductive step goes through for all values of n *except* 1. Every step in the reasoning—from 2 to 3, from 3 to 4, from 4 to 5—they all make sense, all except the very first step, from 1 to 2. And indeed, if we lived in a world where every group of two horses was monochromatic, then we *would* be living in a world where all horses are the same color. It would be a strange world, to be sure, but one thing is certain: mathematical induction would work just as well there as it does in our world.

5 Entrées

We have a feel for induction now. We've seen how it works, we've seen it in action, and we've seen how easy it can be to make a mistake in applying it. We will now make use of it—carefully—to solve three puzzles. Each solution is, at best, surprising, and at worst, counter-intuitive. Induction will be the means by which we extend our intuitions from simple situations, where they are strong, to the apparently complex, where they often falter. The chains of reasoning involved are made up of very simple, repetitive pieces, but they can extend so long as to be completely unintelligible without some systematic method of analysis. Induction plays precisely this role.

The first puzzle I want to consider comes from [2]. It is about a game.

Two people sit facing each other, call them Alexander and Kathleen; these are the players. A third person secretly writes two consecutive natural numbers on two slips of paper, and tapes each piece on the two players' foreheads (one on each). The third person then leaves the room (or sits quietly); his role in the game is finished.

Alexander can see the number taped to Kathleen's forehead, and likewise she can see the number taped to his forehead. So they both know the number that's not their own. They also both know that the two numbers are consecutive.

One player, say Alexander, begins the game by asking Kathleen if she knows what her number is. If she does, she says so and the game ends. If not, Alexander's turn ends and Kathleen gets her chance to ask him if he knows his number. As before, if he does then he says so and the game ends. Otherwise, it becomes his turn again, and he repeats his original question to Kathleen. This back and forth questioning continues until someone finally says "Yes", if ever. The question is:

Does this game ever end?

Sometimes? Never? Always?

One caveat: we must assume that the two players are 'perfect reasoners', so that *if* there was some way for either of them at any point to deduce their own number then they would do it. Without this assumption, whether or not the game ever ends might depend on whether or not one of the players is clever enough, and this kind of question doesn't lend itself to a mathematical answer.

Intuitively, we might imagine the game running as follows. Say Alexander's number is 12. Kathleen can see this, so she knows that her number is either 11 or 13. But there's no way to figure out which, so when Alexander asks her if she knows what her number is, she is forced to respond in the negative. Alexander is in the same position: he can see Kathleen's number and so he can narrow the possibilities for his own number down to two, but he can't decide between them. And since asking the same person the same question over and over again doesn't tend to generate any new answers (barring annoyed quips), it seems that this game is doomed to *never* end.

But this isn't quite true, as you may have already realized. What if one of the players, say Kathleen, has the number 1 taped to her forehead? Then when Alexander sees it, he can reason that his number is either 0 or 2—but wait! The number 1 is the lowest of the natural numbers,[2] so that rules 0 out. Thus Alexander knows that his number must be 2, so the game will end as soon as he is asked, which will be either on the first or the second turn, depending on who goes first.

In fact, the game *always* ends. This is especially surprising because it seems reasonable to expect that if the game doesn't end on the first two turns, then it will never end, since thereafter the players are just repeating the same question over and over. But in fact each response of "No" by one of the players adds a little piece of genuinely new information into the mix. Imagine that Alexander sees the number 2 instead of 1 on Kathleen's forehead. Then he is unable figure out whether his own number is 1 or 3. However, when he asks Kathleen whether or not she knows her own number, if she says no then this tells him that she can't be seeing the number 1 on his forehead! If she were, then she would know that her number is 2, as we reasoned before. This then allows Alexander to deduce that his number is in fact 3.

The reasoning above can be generalized to produce an inductive argument: we will prove that the game always ends in a number of turns no more than twice the lower of the two numbers written on the slips of paper. More formally, let n denote the lower of these two numbers; we will prove by induction on n that the game will end in no more than $2n$ turns.

The 'base case' $n = 1$ corresponds to the setup where one of the numbers is 1 and the other is 2, which we have already seen leads to a game which ends in no more than 2 turns.

Now the inductive step: assume that if the lower of the two numbers is less than or equal to n, then the game ends in at most $2n$ turns. We need to show that if the lower of the two numbers is $n + 1$, then the game ends in no more than $2(n + 1)$ turns.

[2] Be careful: according to some conventions, 0 is counted as the lowest natural number. Both conventions are popular so you're likely to run into each of them depending on what you read and whom you talk to. Notational disagreements like this one are a *hilarious* source of confusion in mathematics, so get used to them!

So suppose that the two numbers are $n+1$ and $n+2$. Suppose also that turn $2n$ has passed and the game has not ended. It is now turn $2n+1$. By the inductive hypothesis, this means that the lower of the two numbers must be at least $n+1$, since otherwise the game would have already ended by turn $2n$. Therefore the player whose turn it is to answer the question, say Kathleen, can deduce that the lowest possible number written on either of the slips of paper is $n+1$. There are two cases to consider.

First, if Kathleen sees the number $n+1$ on Alexander's head, then she can figure out that her own number must be $n+2$, since n was already ruled out as a possibility.

On the other hand, if she sees the number $n+2$ on Alexander's head, then perhaps she can't tell whether her own number is $n+1$ or $n+2$. The turn might pass to Alexander, making it turn number $2n+2$. Alexander looks out and sees the number $n+1$ on Kathleen's head, which allows him to deduce that his own number must be $n+2$, for the same reasons given above. Thus the game ends in at most $2n+2 = 2(n+1)$ turns. This completes the induction.

Exercise 6 *The inductive proof above not only answered the question of whether or not the game would always end, but it also gave an upper bound on how long any particular game would take. Why was this included in the proof? (Hint: It was not just to show off.)*

Exercise 7 *We now have an upper bound of $2n$ on the number of turns a game starting with the numbers n and $n+1$ might last. Is this also a lower bound? If not, can you find a more precise result regarding how long such games will last?*

These exercises are challenging.

The second puzzle I want to discuss is a popular one that has appeared in many forms. One particularly nice formulation of it can be found as an exercise in [3], pages 34-35. I reproduce it here:

University B. once boasted 17 tenured professors of mathematics. Tradition prescribed that at their weekly luncheon meeting, faithfully attended by all 17, any members who had discovered an error in their published work should make an announcement of this fact, and promptly resign. Such an announcement had never actually been made, because no professor was aware of any errors in her or his work. This is not to say that no errors existed, however. In fact, over the years, in the work of every member of the department at least one error had been found, by some other member of the department. This error had been mentioned to all other members of the department, but the actual author of the error had been kept ignorant of the fact, to forestall any resignations.

One fateful year, the department was augmented by a visitor from another university, one Prof. X, who had come with hopes of being offered a permanent position at the end of the academic year. Naturally, he was apprised, by various members of the department, of the published errors which had been discovered. When the hoped-for appointment failed to materialize, Prof. X obtained his revenge at the last luncheon of the year. “I have enjoyed my visit here very much,” he said, “but I feel that there is one thing that I have to tell you. At least one of you has published an incorrect result, which has been discovered by others in the department.” What happened the next year?

Like the previous puzzle, we can use induction to break this problem up into more manageable chunks. We can start to get a feel for how induction might apply by examining some simpler versions of the same question; specifically, we can change the total number of professors from 17 to any other number n and see what happens.

The case $n = 1$ is a silly one, since in this situation there would be no one to apprise Prof. X of the single tenured professor’s mistake to begin with. So let’s move on.

Consider the case $n = 2$. Here we have two tenured professors, each with a published error that they don’t themselves know about, but each knows about the error that the other has published. When Prof. X leaves in a huff, he informs the two that at least one of them knows of an error published by the other one. Then what? At the next luncheon meeting, the two professors stare at one another. Both can reason along the following lines:

“If my colleague over there *didn’t* know of any errors I have published, then she would know that Prof. X was referring to me when he said that one of us *did* know of an error. That would lead her to deduce that she has, in fact, published an error. But she hasn’t deduced that! She’s just staring at me. That means that she *must* know of an error that *I’ve* published. Which means that I’ve published an error, so I have to resign.”

Since both professors can reason in this way, both can deduce that they have published an error, and so both will end up resigning.

Now consider the case $n = 3$. Prof. X obtains his revenge by informing them that at least one of them has published an error that others are aware of. Each of the three can now reason as follows:

"If my colleagues *didn't* know of any errors I had published, then they would know that Prof. X was referring to an error published by one of them. But in that case they could reason along exactly the same lines as outlined above, for the case $n = 2$, between the two of them. This would cause both of them to resign. But they're *not* resigning, we're all just staring at each other! That means that they must know of an error that *I've* published, so I have to resign."

Once again, since each of the three can reason like this, all three will end up resigning.

At this point, you may be a bit confused. It can take some time to wrap one's head around even simpler cases like $n = 2$ or $n = 3$, let alone $n = 17$. The key is to forget about trying to understand the case $n = 17$ directly, and instead focus on just two things. First, how does it work in the base case ($n = 2$)? And second, how can knowledge of a simpler case lead to knowledge of a more complicated case (such as reasoning from a solution for $n = 2$ to a solution for $n = 3$)?

Exercise 8 *Convince yourself that if the base case $n = 1$ of an inductive argument is replaced with the base case $n = k$, for some natural number k, then the proof still shows that the result in question holds true of the natural numbers greater than or equal to k.*

Exercise 9 *Provide an inductive proof (starting at $n = 2$) that answers the original question: all 17 professors will end up resigning.*

As noted, this question appears as an exercise in [3]. The very next exercise, marked as one of the hardest in the book, reads as follows:

> Each member of the department already knew what Prof. X asserted, so how could his saying it change anything?

This question points to an apparent paradox; namely, although we have an inductive *proof* that Prof. X caused all the tenured professors to resign, we also have common sense and basic reasoning telling us that Prof. X couldn't have changed anything. The spectre of a disconnect between mathematics and logic looms ominously. I invite you to put it to rest.

The last of the three puzzles I would like to examine picks up where the previous one left off. As it turns out, Prof. X visited many universities, and as a result scores of mathematicians found themselves unemployed. A group of the less scrupulous ones aligned themselves and formed an international network of math-thieves (people who use math somehow to steal stuff—it's a booming industry). After a particularly successful heist, the group found themselves in possession of 1 million dollars. A meeting was called to distribute the money to the members. Exactly how many members there are is not known outside of the group itself, so we'll just have to say that there are n members and leave it at that. What *is* known is that each member has a unique rank in the organization, from 1st ranked (the leader) all the way down to nth ranked (the last-in-command).

As it turns out, a very precise Code is in place that governs how surplus income is to be distributed. To begin with, the 1st ranked member decides on a potential distribution of the wealth. Each member must be assigned a whole dollar amount (no cents), with 0 dollars of course being allowed. This potential distribution is then put to a secret vote, wherein each member, including the leader, gets to cast exactly one ballot: Yes or No. The members cannot communicate or strategize amongst themselves; it is every ex-mathematician for themselves.

If the vote passes or is a tie, then the money is distributed according to the proposed distribution. The catch is this: if the vote fails, then the 1st ranked member is ousted from the organization forever. Every other member is promoted by exactly one rank to fill the power vacuum, and the *new* 1st ranked member (who used to be 2nd ranked) repeats the process by indicating a new potential distribution and putting it to a vote. This continues until one of the distributions is passed, at which point the members take whatever money was allotted to them by that distribution.

Each member is very invested in this international network, and would rather get no share of the money at all than be ousted from the organization. Each member would also prefer not to oust too many people, if possible, so if all else is equal (i.e. if they would get the same payoff either way), then a member will vote Yes rather than No on a given distribution. Of course, if they figure that they can get even a single extra dollar by voting No on the current plan, they will do it. That's the way the world works, at least among secret math thieves.

Now for the question:

How much cash can the leader pocket?

The answer, surprising as it may be, is *all of it.* The proof of this, much less surprising at this point, is by induction on n, the number of members.

As usual, we can get a feel for how the induction will work by examining simpler cases. If $n = 1$ then there is only one member and she is the leader! She votes to give all the money to himself (being a stickler for the rules), and the vote passes.

If $n = 2$, the leader can still propose a distribution that apportions all the money to herself. The second in command won't like it, since he could get a lot more if the current leader were ousted and he took command, but there's not much he can do about it. His vote of No versus the leader's vote of Yes results in a tie, which means a pass.

If $n = 3$, the leader can *still* get away with giving herself all the money. As before, the 2nd ranked member won't like it one bit, and will vote No. But the 3rd ranked member will realize that *if* she votes No and the current leader is ousted, then the situation will revert to the $n = 2$ case, with her playing the role now of 2nd ranked. In this case, she still gets 0 dollars, and so she will vote Yes to the original proposed distribution, since she gets the same amount (0 dollars) either way.

Now it should be clear how to proceed with the induction. The base case is done and then some. Next is the inductive step: suppose that if there are exactly n members, then the 1st ranked can take all the cash. We need to show that this holds true if there are $n+1$ members, too. So suppose there are $n+1$ members. If the leader apportions all the cash to himself, then the 2nd ranked will vote No, but everyone else will vote Yes because they get the same payoff (i.e. nothing) either way. It's as simple as that; this completes the induction.

There are several variants on this puzzle that can make it more challenging.

Exercise 10 *How much cash can the leader pocket if tie votes result in oustings?*

Exercise 11 *How much cash can the leader pocket if members vote No rather than Yes if they get the same payoff either way?*

6 Icing on the Cake

I will close this article with three fun examples of paradoxes based loosely on mathematical induction. I qualify my words because in each case induction is not really a pivotal part of the paradox; it appears sometimes merely to grease the wheels of the description, and other times more maliciously, to confuse the reader even more. Moreover, the solutions to these paradoxes do not rely on induction in any particular way, and so they will be omitted entirely! (Or, as mathematicians often prefer to put it, they will be "left as an exercise".)

First we have the proof that everyone is pretty much bald. It goes by induction: we will prove that for all n, if you have n hairs on your head then you are pretty much bald. The base case is easy: if you have 1 hair on your head, then certainly you're pretty much bald.

Now suppose inductively that if you have n hairs on your head, then you're pretty much bald. We need to show that the same is true for someone with $n+1$ hairs on their head. But certainly if someone has only 1 hair more on their head than someone else who is pretty much bald, then that first person is also pretty much bald. This completes the induction!

Next we have the proof that every natural number is interesting. Perhaps you already believe this to be so, but a proof certainly wouldn't hurt! We proceed by induction on n. The base case $n = 1$ is obvious because of course 1 is a very interesting number. Now suppose that all the numbers from 1 to n are interesting. We need to show that $n+1$ is interesting. But consider this: if $n+1$ were *not* interesting then by the inductive hypothesis it would be the very *smallest* uninteresting number—and that makes it very interesting! This completes the induction.

Finally, we'll take a quick look at the paradox of the unexpected examination.[3] This one is more serious than the previous two, though how much more is hard to say. The setup is as follows:

Friday afternoon, just before school lets out, a teacher promises his class that they will have a quiz on one of the five days of the coming week. Moreover, he guarantees the students that the quiz will be a surprise in that they won't be able to predict the night before that it will happen the next day. The class is dismayed until one of the students realizes that something fishy is going on. She reasons:

"The quiz can't be given on Friday, for sure, because that's the last possible day, so we would be able to predict it Thursday night. So Friday is out. That makes Thursday the last possible day the quiz can be given. But then the quiz also can't be given on Thursday, because Wednesday night we would know it was coming the next day! And in the same way we can eliminate Wednesday, Tuesday, and even Monday from the list of possible days for the quiz."

This argument is good enough to convince the rest of the class, who gleefully go about their business, content in the certainty that there can be no surprise quiz. Tuesday morning comes, however, and the teacher hands out a quiz sheet to each student. There are, of course, objections: "You can't give this quiz! We already figured out that you couldn't make it a surprise no matter what day you gave it on!"

But the teacher is unperturbed. "You figured that out, did you? Well, here's the quiz. Aren't you surprised?" The students reluctantly agreed that they were. But where did their logic go awry?

[3]Thanks to Alexander Nabutovsky for introducing me to this paradox in his Introduction to Mathematical Logic class at the University of Toronto.

References

[1] Richard Friedberg. *An Adventurer's Guide to Number Theory.* McGraw-Hill Book Company, United States of America, 1968.

[2] Martin Gardner. *Aha! A Two Volume Collection.* The Mathematical Association of America, United States of America, 2006.

[3] Michael Spivak. *Calculus.* Publish or Perish, Inc., United States of America, third edition, 1994.