On Choosing

go figure

Five-year-old Harry and his family are going on a road trip. "Bring something to read on the drive, Harry," his mother tells him. Harry is ecstatic – that means he gets to bring his rabbit books! He fetches his 12 favorite books about rabbits. His mother is not excited about the prospect of lugging around all those books for the whole trip. "Not all of them, Harry," she tells him. "You can bring *four*." Four!? Harry has a conundrum on his hands. How many ways can he choose four to bring for the drive?

The first thing we might notice about this problem is that it's not asking for a sequence of books, where the order of the books matters. In this problem, it doesn't matter what order the books are in. But this notion of selecting looks rather like our thought process from the factorials page. So we might start by saying, "Okay, we can still use the ordering strategy from the factorials page but let's stop when we get past the first four. There are 12 ways to pick the first book, and then there are 11 books left and so 11 ways to pick the second book, and then there are 10 books left and so 10 ways to pick the third book, and then there are 9 books left and so 9 ways to pick the fourth book, and the order of the rest of the books doesn't matter because they're staying home, so we get a total of

$$12 \cdot 11 \cdot 10 \cdot 9 = 11,880 \text{ ways.}"$$

Well, that's a good start. But let's look more closely at the problem. When we say order doesn't matter in here, we're not just talking about the books that stay home: we're also talking about the books that Harry chooses for the drive. A choice of Books A, B, C, and D is the same as a choice of Books D, C, B, and A, which is the same as a choice of Books B, D, A, and C. If you put those four books in a big sack and jumbled them up, you wouldn't be able to tell one ordered sequence from another: they would be *indistinguishable*.

Let's think a little more carefully about what we're doing. We'll start with a simpler case. Suppose Harry could only choose three books out of six. Now if we followed our initial approach, we'd say there are 6 ways to pick the first book, times 5 ways to pick the second book, times 4 ways to pick the third book, for a total of $6 \cdot 5 \cdot 4 = 120$ ways. We're still overcounting, but now the numbers are more manageable. If Harry chooses Books A and then B and then C, it's the same as choosing Books B and then C and then A: either way, the same three books are coming on the drive.

Basically what we're doing is lumping together all of the possible sequences that would be indistinguishable if you jumbled them up in a sack. So what possible sequences could produce this jumbled sack? Let's list them:

<div align="center">
A, B, C
A, C, B
</div>

B, A, C
B, C, A
C, A, B
C, B, A

And how many options are there? 6, which is $3!$. It's the number of ways there are to arrange the letters. So if we're trying to pick 3 objects using our initial approach, we're really counting six things when we should only be counting one. That means our answer 120 is six times as big as it should be. For our real answer, then, we can start with $6 \cdot 5 \cdot 4$, and then divide by $3! = 6$ to correct for overcounting:

$$\frac{6 \cdot 5 \cdot 4}{3!} = 20.$$

If you're not convinced, write all the options out for yourself!

Now let's return to Harry's conundrum. Remember, he needs to choose 4 books out of 12. We now know how to tackle this: we start with 12 ways to pick the first book time 12 ways to pick the second times 10 ways to pick the third times 9 ways to pick the fourth, and then divide by $4!$ to correct for overcounting. The result is:

$$\frac{12 \cdot 11 \cdot 10 \cdot 9}{4!} = \frac{11,880}{24} = 495.$$

And that's our answer.

When we choose r objects from a group of n objects, and the order of the r objects doesn't matter, we call it "n choose r," and we write it like this:

$$\binom{n}{r}$$

Let's write a general formula for the logic we just went through. The first thing we did to calculate $\binom{n}{r}$ was to multiply $(n)(n-1)(n-2) \cdots$ until we had r numbers being multiplied. That gave us a numerator of

$$(n)(n-1)(n-2) \ldots (n-r+1)$$

We can do better than that, though. To get rid of the "..." in the middle, we can think of this expression as $n!$ without the last $n - r$ terms. Then we can rewrite the expression as

$$\frac{n!}{(n-r)!}$$

Now that looks like a fraction, not a numerator. But bear with me for a minute. Our next step was to divide the numerator by $r!$ to correct for overcounting. The result is:

$$\frac{n!/(n-r)!}{r!} = \frac{n!}{r!(n-r)!}$$

And there we have it. That thing on the right is our formula.

The powerful thing about having a formula like that at our disposal is that now we can straightaway plug in the numbers whenever we want to choose things from a group, and we can use it to prove identities algebraically. However, there are often ways to prove these identities that are much less ugly.

For instance, we could prove that $\binom{n}{r} = \binom{n}{n-r}$ algebraically, like so:

$$\binom{n}{n-r} = \frac{n!}{(n-r)!(n-n+r)!} = \frac{n!}{r!(n-r)!} = \binom{n}{r}$$

Or we could ask ourselves what the identity really tells us. And we find that $\binom{n}{r} = \binom{n}{n-r}$ really means that Harry choosing some number of books to bring on the road trip is the same as Harry choosing the remaining number of books to stay home. Choosing 4 out of 12 is the same as choosing 8 out of 12. And that's a lot more intuitive, isn't it?

Here's another example. Suppose Harry reasons with his mother for a bit, and she changes her ruling: "Okay, Harry, you can bring six books; but let's pack two of them in the back of the car." Harry can pick six books in total and then pick two of those to get locked in the trunk. Or, he could pick the two books first, and from his remaining collection choose the four he can read in transit. Let's create expressions for both methods.

For the first approach, Harry has $\binom{12}{6}$ ways to pick 6 books out of 12, and then for whichever such choice he makes, he has $\binom{6}{2}$ ways to pick the 2 that will be stored in the trunk. Multiplying these will give us our total number of options: $\binom{12}{6}\binom{6}{2}$.

On the other hand, he could first pick 2 from 12 to go in the back of the car – $\binom{12}{2}$ ways.

Then from the remaining 10 books in his collection, he picks 4 to read on the drive, in one of $\binom{10}{4}$ ways. The total number of options is then $\binom{12}{2}\binom{10}{4}$. But this counts the same number of options as our first expression, so

$$\binom{12}{6}\binom{6}{2} = \binom{12}{2}\binom{10}{4}$$

Put more generally:

$$\binom{n}{r}\binom{r}{k} = \binom{n}{k}\binom{n-k}{r-k}$$

An algebraic approach would have involved expanding all of these out and multiplying. It's not so elegant, but it certainly works. I'll leave ye intrepid readers to work that one out yourself.

Perhaps the most famous of these counting identities is Pascal's Identity:

$$\binom{n}{r} + \binom{n}{r+1} = \binom{n+1}{r+1}$$

We could prove this algebraically, or we could devise a clever counting argument. But there's another way, one that's much more streamlined – and it has to do with the reason it's called Pascal's Identity. Why is it called Pascal's Identity? Read on….