The law of total probability is another probability law that gives you a way to reason about cases, which we've seen is a fundamental technique for dealing with all sorts of problems. So the point of cases, of course, is that you can prove a complicated thing by breaking it up into, if you're lucky, easy sub cases.

So here's the way to understand the law of total probability abstractly. It starts off with set theoretic reasonings. Suppose that I have a set A embedded in some larger sample space S. So A is really an event, but we're just going to think of it as a set.

Now suppose that I have three sets, B1, B2, and B3 that partition the sample space. That is B1, B2, and B3 three don't overlap. They're disjoined, and everything is in one of those three sets. So there's a picture of B1, B2, and B3, cutting up the whole sample space S, represented by the square or rectangle.

Now of course, these three sets that cut up the whole space, willy nilly cut up the set A into three pieces. The first piece is the points in A that are in B1. The second piece is the point in A that are in B2. And the third is the points in A that are in B3.

So that's why we a basic set theoretic identity that says that as long as B1, B2, and B3 have the property, that they're union is in the universe. Everything. And they are pairwise disjoined, then any set A is equal to the union of the part of A that's in B1, the part of A that's in B2, the part of A that's in B3. And this is a disjoint union, because the B's don't overlap.

That means that if I was talking about cardinality, I could add them up. But in terms of probability, I can apply the sum rule for probabilities and discover that the probability of A is simply the probability of B1 intersection A, B2 intersection A, B3 intersection A.

Now the most useful form of the law of total probabilities is when you replace this intersection, B1 intersection A, by the conditional probability using the product rule-- so let's replace it by the probability of A given B1 times the probability of B1. That's another formula for B1 intersection A.

And if I do that with the rest of them, I now have the law of total probability stated in the usual way in terms of conditional probabilities where it's most useful. Now I did it for three sets. But it obviously works for any finite number of sets. As a matter of fact, it works fine for any countable union of sets.

If I have a partition of the sample space S in to B0, B1, and so on-- a partition with a countable number of blocks-- then it's still the case that the probability of A is equal by the sum rule to the probability of these disjoint pieces, the parts of A that are in each of the different blocks of the partition.

And reformulating that as a conditional probability, I get the rule that the probability of A is the sum over all possible i of the probability of A given Bi times the probability of Bi. And that basic rule is one and we're going to get a lot of mileage out of when we turn in the next segment to analyze and understand the results of tests that are maybe unreliable.