Size of Sets

We often want to talk about the size of a set, to say something about how large it is. And, we want to compare the sizes of different sets, to say whether some set is larger, smaller, or just as large as some other set. How can we do this?

For finite sets this seems all rather straightforward: we can define the size of a set to simply be the number of elements in it. Accordingly, one finite set A is larger, smaller, or just as large as some other finite set B iff the number of elements in A is larger, smaller, or the same as the number of elements in B. So far so good.

Things get more complicated though once we start to consider infinite sets. For example, are there more natural numbers than even numbers? One person might say, yes, there are in fact twice as many natural numbers than even numbers. Someone else might say, no, in both cases the sets are infinitely large, and there is really no difference between the two sets in terms of how many elements are in it. A third person could say, yes, because the set of even numbers is a strict subset of the set of natural numbers. Which person is right? More generally, how should we think about the size of infinite sets, and comparing the size of two infinite sets?

The trouble is that, unlike what we did for finite sets, we cannot assign any number to the amount of elements in an infinite set. Indeed, while it is standard practice to talk about infinite sets having an 'infinite number' of elements, there is of course no such thing as an infinite number: a number is a point on a number line, and 'infinity' isn't on that line. A number is something that has a decimal representation, and 'infinity' doesn't. Of course, intuitively, 'infinity' is 'larger' than any number, and thus any infinite set is larger than any finite set, which accords with our intuitions. But, this not only makes a questionable application of the 'larger than' relationship as defined between numbers to something that isn't a number, it also doesn't help us to compare the sizes of two infinite sets. Let us also note that the suggestion of the third person above, namely to use the subset relationship to compare the sizes of infinite sets, while intuitive, doesn't help us in comparing the sizes of sets containing completely different elements. What to do?

One line of thinking regarding this matter starts with the very distinction between finite and infinite sets. How, indeed, are finite and infinite sets defined? Talking about such sets as having a finite or infinite number of elements respectively not only uses that highly questionable term of 'infinite number', but also seems to simply repeat the crucial notion of 'infinite', rather than elucidating it. Indeed, to say that a set is 'infinitely large', or to say that it has an 'infinite number of elements' is in that respect pretty much the same thing. Rather, we want something that accounts for sets being 'infinitely large', having an 'infinite size', or (God forbid) 'having an infinite number of elements' all at the same time. We want something that captures the very notion of 'infinite'.

Here is something we can do: imagine a process of going through all the elements of a set one by one, while never repeating any element. Indeed, imagine taking elements out of a set, setting them apart, one by one, while never putting them back in. A finite set is one where, at some point, one has exhausted all elements, i.e. the original set has become empty. An infinite set, however, is one where it is impossible to reach such a point: it is impossible for such a process to exhaust all elements, since no matter how many elements one has taken out of such a set, there are always more elements left in the set.

What is interesting, is that we can use this imaginary process of taking out objects from sets one by one to try and relate the size of sets as well: when comparing the size of two sets A and B, we simply imagine a process of taking out one element from A, and one element from B at the same time. Whichever set empties out the quickest, is the smallest of the two. Notice that this not only works fine for two finite sets, but it also coincides with our intuition about finite sets being smaller than infinite sets: the finite set will, trivially, always empty out before the infinite set does. So, this is already a better way of comparing sizes of sets than trying to do this in terms of numbers.

However, how does this help us with comparing two infinite sets? In that case, neither set will ever completely empty out. True, but let's look at the above process of emptying out two sets A and B a little bit differently. What we are doing here, is to uniquely associate with each element a that we take out of A the element b of B that we are taking out at the same time as a. In other words, the process of emptying out the sets A and B at the same time is effectively defining a function from A to B. Thus, the proposal is to try and compare the sizes of two sets A and B by talking about the nature of any function defined between them.

Now, for any two finite sets A and B, comparing their size in terms of functions between them is relatively straightforward. A is larger than B if and only if there is no total, one-to-one function from A to B. Or, what is logically the same thing, A is smaller or equal to B if and only if there is a total, one-to-one function from A to B. Similarly, B is larger than A if and only if there is no onto function from A to B, i.e. B is smaller or equal to A if and only if there is an onto function from A to B. And finally, A and B are of equal size if and only if there exists a one-to-one correspondence between A and B.

More importantly, we can now use these functional relationships between sets A and B to try and give proper definitions of size-comparisons between infinite sets. Thus, if there does not exist a total, one-to-one function from A to B, then it stands to reason to say that A is 'larger' in size than B, even if both are of 'infinite' size. In other words, if there is a total, one-to-one function from A to B, then A is smaller or equal in size than B. Similarly, if there exists an onto function from A to B, we can say that B is smaller or equal in size than A. And finally, if there exists a one-to-one correspondence between A and B, then A and B are equal in size.

Using functions, we can therefore give a proper definition of size-comparisons for sets, whether they are finite or infinite. Moreover, these definitions have some nice further properties coinciding with our intuitions about size as will demonstrated below. And finally, as will also be demonstrated, this definition allows us to make some interesting and important distinctions between the different sizes of different infinite sets, for as it turns out, some pairs of infinite sets can now be shown to be of 'equal' size, while other pairs of infinite sets are not. In short, we seem to be dealing with a clean, fruitful definition that in many important coincides with our intuitions. Mathematicians therefore generally use this method to talk about size-comparisons between two sets.

Still, before doing these further demonstrations, let us make note of an important caveat. This caveat is best illustrated by going back to our earlier example of comparing the set E of even numbers, with the set N of all natural numbers. Let us first note that using the definition above, we can now provide an answer to the question as to whether these two sets are of the same size or not. For consider the function from E to N, where every number n=2k from E gets mapped to number k in N (or, if you want, f(n)= n/2). It is easy to see that this function is total, one-to-one, and onto. Hence, it is a one-to-one correspondence. And hence, using the definition, the two sets are deemed to be of equal size.

But of course, this seems rather strange to some people. While we earlier rejected the use of subsets to try and compare two sets on account of it not being applicable to sets with different elements, it is certainly applicable now, and clearly E is a strict subset of N. So, why isn't N larger than E? Why should we give preference to the functional definition that claims that the sets are in fact of equal size? More importantly, even in terms of functions we can raise some serious trouble. For instead of considering the function f(n)=n/2, consider now the function f(n)=n. This function is total, one-to-one, but not onto. Doesn't that imply that in some other, just as intuitive and meaningful way, E is smaller than N?

What is going on here? Well, go back to the case of finite sets first. Note that for any two finite sets A and B, if there exists one function from A to B that is total, one-to-one, and onto, then automatically, *any* function from A to B that is total and one-to-one will be onto. Thus, we could say that any two finite sets A and B are of equal size if and only if any total, one-to-one function from A to B is onto; if there is a function from A to B that it total, one-to-one, but not onto, then clearly B is larger than A. We could equivalently say that A and B are of equal size if and only if any total and onto function from A to B would have to be one-to-one, and if there is a total and onto function that is not one-to-one, then A is clearly larger than B. So, all this making perfect sense for finite sets, one could try and use this as a definition for infinite sets as well: A and B are of equal size if and only if all total and one-to-one functions from A to B are onto, A is larger than B if and only if there is a total and onto function that is not one-to-one, and B is larger than A if and only if there is a function from A to B that it total, one-to-one, but not onto. Once again, this is a perfectly proper definition, and it coincides with our intuitions just as much for the finite sets as the original definition. But now, as the example above demonstrates, since there apparently is a function that is total, one-to-one, but not onto, it follows that E and N are not the same size, which at first sight seems to actually square better with our intuition than the original definition.

So, the question really is: why use one particular functional relationship between sets to define some size-comparison measure instead of some other functional relationship between sets, especially if one definition leads to a completely different result than the other? Why indeed should we regard one definition as ‘the correct’ definition? Well, that’s an excellent question, and it gives us one of those rare glimpses into the nature and foundations of mathematics. In particular, what we are really dealing with here is that we are trying to get a grip, by making hard, a notion or concept that we have, in this case the concept of size, and in particular the size-comparison concepts of larger, smaller, and equal to. We are, in short, giving a theoretic account of a pre-theoretic notion. And that’s where the rub is: how do we know whether some mathematical definition really captures what we want it to capture? And let us note that this problem is really not isolated to the notion of size. Later, we will see this very same problem cropping up when it comes to the notion of computation. And, you may already be familiar with the case of sets. If you look at how mathematicians are trying to define the notion of a set, you’ll actually find quite a bit of disagreement as to what a set exactly is. And, even where there is a consensus, there is always that lingering question: did we capture exactly what we tried to capture? Maybe we missed something, or maybe we did something wrong. We may at some point find that our definition is unacceptable for any such reasons, i.e. we can convince ourselves that some definition isn’t acceptable, but not being able to find such reasons now doesn’t mean that such reasons don’t exist, so how can we ever be certain that our definition captures what it is trying to capture? Of course, maybe there simply isn’t any ‘thing’ to get ‘right’ in the first place: maybe our vague concept is just that: a vague concept. Indeed, it is not unlikely that by trying to make these intuitions hard, we are in fact changing our very intuitions and concepts. Another possibility is that it turns out that our concept was in fact internally inconsistent: something that couldn’t possibly have any kind of ‘reality’ to them other than that of being a creation of our mind. Which, by the way, raises another general issue: does my concept really coincide with your concept? Are we even working with the same concept? So, ultimately, how can we really say that someone who believes that there are more natural numbers than even natural numbers, is really wrong?

Fortunately, in this particular case, we can convince ourselves that the alternative proposed definition really is problematic. Indeed, it may be one of those cases where our intuitions are inconsistent. For consider what would happen if we were to adopt the alternative definition, and consider the function $f(n)=n+1$ as defined from the natural numbers to the natural numbers. This function is total, one-to-one, but not onto. So, according to the alternative definition, this would mean that the set of natural numbers is larger than itself! That, of course, is not something we can live with, and makes us reject the alternative definition as a useful definition. Consequently, we are left with the original definition. And, as the following results demonstrate, it is a definition that actually has some nice further features, which seems to indicate that it does its job. But, the moral is that it is very hard to say that some particular definition is ‘correct’. Indeed, maybe this definition will lead to some absurd consequence as well. Or, someone finds an alternative definition that works just as well, but gives different results. Strange as it sounds, these things are always possible in mathematics. And, we should note that the results from mathematics are only so useful as its definitions and axioms are. Thus, while the mathematical and logical reasoning may be impeccable, the question as to what the results really mean depends for a large part on what the definitions and axioms capture or not capture.

Even better news (even if a bit of a copout) is that for the purposes of this course, we really don't need any answer to the question of whether certain sets are larger, smaller, or equal in size than other sets. Instead, we will be using the definition from above to define a notion that we are going to call cardinality, and leave it up to interested people to debate whether this mathematical notion of cardinality captures our size-comparison concepts.

Cardinality of Sets

We will say that every set has a certain *cardinality*. For a set S, its cardinality is written as $|S|$. For finite sets, we will even define this cardinality: for a set S with n elements, $|S| = n$. For infinite sets, we can't use numbers, and in fact we are not going to define what their cardinality *is* in terms of something we are familiar with. What we are going to do, however, is to define binary relationships $\leq$, $\geq$, and $=$ on the cardinalities of sets. In particular, for any sets A and B:

$|A| \leq |B|$ if and only if there exists a total and one-to-one function from A to B

$|A| \geq |B|$ if and only if there exists an onto function from A to B

$|A| = |B|$ if and only if there exists a one-to-one correspondence between A and B

$|A| < |B|$ if and only if $|A| \leq |B|$ but not $|A| = |B|$ (similar for $>$)

If $|A| = |B|$, then A and B are said to be *equinumerous*

Using these definitions, we can show that $=$ is an equivalence relation, and that $\leq$ and $\geq$ are total (linear) order relations, which is just what we would like them to be (we would most likely reject any definition that would not satisfy these properties, such as the earlier proposed alternative definition of 'smaller than'). That is, we can prove that for any sets A and B:

0. $|A| \leq |B|$ if and only if $|B| \geq |A|$
1. $|A| = |A|$, $|A| \leq |A|$, and $|A| \geq |A|$ (reflexivity)
2. If $|A| = |B|$ then $|B| = |A|$ (symmetry)
3. $|A| \leq |B|$ or $|B| \leq |A|$ (totality)
4. $|A| \leq |B|$ and $|B| \leq |A|$ if and only if $|A| = |B|$ (same for $\geq$) (anti-symmetry; from $|A| \leq |B|$ and $|B| \leq |A|$ to $|A| = |B|$ is difficult: this is the Schröder–Bernstein theorem)
5. If $|A| \leq |B|$ and $|B| \leq |C|$, then $|A| \leq |C|$ (same for $\geq$ and $=$) (transitivity)

Exercise:

Prove 5

Enumerability

A set S is *enumerable* if and only if there exists a mapping f: N → S such that f is onto. Informally, a set S is enumerable iff there exists a (linear, but possibly infinite) list such that each element of S occurs somewhere in that list. If so, we say that the list enumerates all of the members of the set. An enumerable set is said to have a *countable* number of elements. Finally, any infinite set that is enumerable is said to be *denumerable*.

Examples:

I. Trivially, every set S with a finite number of elements is enumerable.

II. The set E of even positive integers is enumerable: 2 4 6 …

III. The set N is enumerable. Each of the following lists exhausts all members of N:

1 2 3 … (trivial)

2 1 4 3 6 5 … (order doesn't matter: as long as each element is in the list, it's ok)

1 1 2 2 3 3 4 4 … (having duplicates doesn't matter: as long as …)

1 _ 2 _ 3 _ … (having gaps (because f may be a partial function, i.e. f(n) may be undefined for some n) doesn't matter: as long as …!)

IV. The set Z of all integers is enumerable: 0 1 –1 2 –2 …

V. The set Q of all rational numbers is enumerable. To see this, notice that any rational number q can be written (by definition) as a number m/n where m and n are integers. Now consider the following array:

	1	2	3	…
0	0/1	0/2	0/3	
1	1/1	1/2	1/3	
2	2/1	2/2	2/3	
:				

Since every rational number will appear somewhere in this array, every rational number will appear somewhere in the list that we construct by writing down every element that we encounter by weaving through the array as indicated below (start top left):

	1	2	3	…
0	→	↓	→	
1	↓	←	↑	
2	→	→	↑	

Theorem: A set S is denumerable iff S is equinumerous with N.