Set theory for intro discrete math

Randall Holmes

August 26, 2022

This document discusses simple set theory foundations for discrete math.

I do believe that some axioms and definitions are appropriate, and I give some. But I try to restrict them to what is actually useful for discrete math at a beginning undergraduate level.

Something interesting happens (which is part of the reason I am writing this). The sum of what we teach in a discrete math course does add up to an axiomatic set theory, close to Zermelo set theory though not identical to it, and adequate for the foundations of classical mathematics.

The idea here is to exhibit situations in discrete math where the set and function abstractions come up naturally, and develop the exact axioms and definitions that are appropriate in this context.

Natural finite examples of the set and function abstractions come up in combinatorics.

When I ask, how many subcommittees of three people can be formed from a committee of ten people, I am actually asking, how many subsets of size three does a subset of size ten have?

And there is something disingenuous here...of course, two subcommittees can have the same roster. We are asking a question about sets in the background so we want to presume the identity condition for sets...two sets are the same iff they have the same elements. A semi-real-world question along the same lines which avoids this issue is...a subcommittee of three is to be formed from a committee of ten people...how many possible rosters of members are there? Of course we know that the order of the members doesn't matter – but this is another way one could become confused as to what the correct answer is here if the set concept isn't familiar.

Another thing to notice in any discrete math application is that the size (at least of a finite set) is an important part of the interface of the data

type of sets (at least, finite sets). A lot can be done with just the notions of membership and equality, but in discrete math we are interested in counting.

1 First lecture: some definitions, axioms, basic operations

I shall begin with the simplest assumptions about sets.

primitive properties and relations: Some objects in our world are *sets*.

Primitive notion 1 (sets and individuals): Some objects are *sets*. Objects which are not sets we call *individuals*[1].

Primitive notion 2 (membership): There is a relation of *membership*, written $\in$, which holds between general objects and sets they "belong to".

Primitive notion 3 (equality): We presume familiarity with the general notion of equality, and with its basic logical properties ($x = x$, and "if $x = y$ and $P[x]$ then $P[y]$").)

We say "x is an element of a" or "x is a member of a", "x belongs to A" or "x is contained in a" to mean $x \in a$. "x is included in a" has another meaning for us. We write $x \notin A$ for "x is not an element of A".

identity criterion: With any data type, we want to be able to tell when two objects of that type are the same.

If A and B are sets, $A = B$ holds if and only if A and B have the same elements, that is, for any x, $x \in A$ holds exactly when $x \in B$ holds. Another way of putting this is, There is no element of A which does not belong to B, and there is no element of B which does not belong to A. This avoids vacuous quantification.

This is summarized in

Axiom 1 (extensionality): If A and B are sets, and no element of B is not an element of A, and no element of A is not an element of B, then $A = B$.

[1] what I call "individuals" have also been called "atoms" or "urelements"

individuals and empty set: From the introduction of the membership relation, we extract this statement:

> **Axiom 1b (axiom of sethood):** If $x \in A$, then A is a set. Equivalently, if A is an individual and x is any object, $x \notin A$ (in which form we would call it the axiom of individuals).
>
> We call this axiom 1b because it is a footnote to extensionality, and also because usual treatments of set theory do not allow for individuals at all (but this tends not to be convenient in undergraduate discrete math classes).

A special case of the identity criterion is that if A and B are sets, and both have no elements at all, then they are equal. There is at most one set with no elements (we will introduce an axiom that says there is one in a moment). In addition, any objects in our world that are not sets have no elements. These objects can be called by various names: we call them individuals.

An important point in mathematics pedagogy is that official foundations of mathematics usually say that everything is a set, but common sense allows for individuals. Moreover, familiar mathematical objects (such as the natural numbers) have implementations as sets, but there is nothing inevitable about these implementations, and it is often natural in an undergraduate discrete math text to treat items not explicitly given as sets as individuals [In this document we will discuss implementations (more than one of them) of the natural numbers as sets, but our official position will be that the natural numbers are individuals.]

An axiom which tempts me which is not found in usual treatments is the assertion that

> **Axiom A (preliminary form):** for any set, there is a set of individuals which is the same size as that set.
>
> This doesn't get a number because it is not found in usual treatments at all.

I'll formalize this and talk about why it is tempting at a later point if this seems natural. The idea is undergraduate accessible: if we have a mathematical structure which we do not think of as a collection of sets,

but we have implemented it as a collection of sets, we can find a collection of individuals the same size to use as our implementation of that structure, so that there are no distractions caused by the "members" of elements of that mathematical structure.

properties define sets: If we have a set A given and a statement $P[x]$ about objects x in general, there is a set $\{x \in A : P[x]\}$ for which this is true: for any object a, $a \in \{x \in A : P[x]\}$ if and only if $a \in A$ and $P[a]$.

This says: Given a set A and a property P, we can extract the collection of all elements of A which have the property P as a new set.

Notice that for any A, $\{x \in A : x \neq x\}$ is an empty set, and there is only one, for which we adopt the notation $\emptyset$.

We codify this as an axiom.

Axiom 2 (separation): For any sentence $P[x]$ about general objects x, and any set A, we have an object $\{x \in A : P[x]\}$, which is a set, and the axiom "for any a, $a \in \{x \in A : P[x]\}$ if and only if $a \in A$ and $P[a]$".

subset relation and power set: We define $A \subseteq B$, read "A is a subset of B" as "A is a set and B is a set and for any x, if x is an element of A, x is an element of B" **or** "A is a set, B is a set, and anything which is not an element of B is not an element of A". The contrapositive formula has value because the implicit quantifier is never vacuous: there is always something not in B, and it's clear from this definition that $\emptyset \subseteq B$: anything not in B is not in $\emptyset$.

Note that $\emptyset \subseteq A$ and $A \subseteq A$ always hold.

We assert as a basic assumption that

Axiom 3 (power set): for any set A, there is a set $\mathcal{P}(A)$, called the power set of A, whose elements are exactly the subsets of A.

Notice that the axiom of extensionality actually says that if $A \subseteq B$ and $B \subseteq A$, then $A = B$. That is exactly what it says.

list notation for finite sets: The notation $\{x\}$ denotes the set whose only element is x.

The notation $\{x, y\}$ denotes the set whose only elements are x and y.

The notation $\{x, y, z\}$ denotes the set whose only elements are x, y, z.

And so forth. List notation $\{x_1, x_2, \ldots, x_n\}$ defines a set whose only elements are the x_i's. The general definition of this (familiar) list notation is technically exacting to state (we will do it eventually).

Notice that order and repetitions of items do not make any difference in the reference of this notation. $\{x, x\}$ is the same set as $\{x\}$. Notice that this means that you cannot tell that a set written $\{x, y\}$ has two elements unless you are given the information that x and y are distinct.

$\{x, y\}, \{y, x\}, \{x, y, y\}$ etc. are all the same set.

assumptions behind list notation: Behind the ability to write this notation, there are two basic assumptions.

Axiom 4 (singletons): For any object a, there is a set $\{a\}$ such that for any x, $x \in \{a\}$ exactly if $x = a$.

It is fun to notice that if a is a set, $\{a\} = \{x \in \mathcal{P}(a) : x = a\}$ is already given by axioms previously stated. But we are not presuming that every object is a set, so we need the axiom of singletons. [2]

Axiom 5 (binary union): For any sets A and B, there is a set $A \cup B$ such that for any x, $x \in A \cup B$ if and only if either $x \in A$, or $x \in B$, or both. This set is called the union of A and B.

Now $\{x, y\} = \{x\} \cup \{y\}$ and more generally $\{x_1, x_2, \ldots, x_n\} = \{x_1\} \cup \{x_2\} \cup \ldots \cup \{x_n\}$.

[2] An inverse operation of sorts to the singleton operation is definite description. It would be handy to have an operator θ such that $\theta(\{x\}) = x$ and $\theta(u) = \emptyset$ if x is not a singleton set. Then $\theta(\{x \in A : P[x]\}$ would represent the unique x in A such that $P[x]$ if there is one. We note the desirability of this without postulating it. We would read $\theta(A)$ as "the unique element of A", and use of this notation would usually presume that A has exactly one element.

Relations of part and whole on sets?: There is a temptation to say that a set is a whole made up of its elements. A classic textbook written by a man who certainly knew better gave packs of wolves and bunches of grapes as examples of sets.

This temptation should be firmly resisted. If A is part of B and B is part of C, then A is part of C, for any objects A, B, C on any reasonable understanding of the relation of part to whole. But if a, b are any two distinct objects, $a \in \{a, b\}, b \in \{a, b\}$, and $\{a, b\} \in \{\{a, b\}\}$. If membership were transitive as the relation of part to whole is, it would follow that $a \in \{\{a, b\}\}$ and $b \in \{\{a, b\}\}$. But $\{\{a, b\}\}$ has only one element, the set $\{a, b\}$: a and b are not both in it (even if one of them were weirdly the same as $\{a, b\}$)

Related temptations are the common desire to say that $\emptyset$ belongs to every set, or to write $\emptyset$ as $\{\emptyset\}$ (the latter is a set with one element while $\emptyset$ has no elements).

It is important to notice that none of this has to do with famous worries about infinite sets: these issues arise from the simplest construction of sets by finite listing. They do have to do essentially with allowing sets to be elements of sets, which is an important move for the uses of sets intended in mathematics.

The natural relation of part to whole on sets is the subset relation. We say "A is included in B" for $A \subseteq B$, not $A \in B$. Note that if $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$.

It may seem peculiar that every set has the empty set as a part: a part of a set could be defined as a *nonempty* subset of the set, which would preserve the idea that disjoint sets have no common part (though they do have the common subset $\emptyset$)

Notice that $x \in A$ is equivalent to $\{x\} \subseteq A$: the elements of A correspond to but are not identical with atomic parts (smallest possible nonempty subsets) of A.

Notice that the examples given in the famous textbook are objects with disjoint parts of an understood kind: a pack of wolves does have a natural association with a set of wolves, and a bunch of grapes with a set of grapes. But the mass of all human cells is roughly speaking the same as the mass of all human beings, while the set of human cells is a lot larger then the set of human beings.

Other interesting binary operations on sets: We define $A \cap B$ as

$$\{x \in A : x \in B\}.$$

This is called the intersection of A and B.

Exercise: prove that $\{x \in A : x \in B\} = \{x \in B : x \in A\}$

We define $A - B$ as $\{x \in A : x \notin B\}$. Here, $x \notin B$ simply means "x is not an element of B". This is called the set difference of A and B or the complement of B relative to A.

These, along with union, are the basic operations for the parlor game of Venn diagrams, which we will play (which does have its uses, mostly to illustrate very simple properties of two or three sets).

Notice that the two operations introduced here require no new axioms, because the set defined is included in a set already given.

The abstraction of ordered lists: In addition to considering sets, which are not ordered ($\{x, y\}$ is the same set as $\{y, x\}$) we want to consider ordered pairs (x, y) or ordered lists $[x_1, x_2, \ldots, x_n]$, which are the same exactly if they have the same number of items appearing in the same order (repetitions being significant). Our reasons for using different delimiters for lists will be revealed later.

This is an independent, if related idea. It is interesting that it can be implemented entirely in terms of the set concept. But please notice that there is nothing inevitable or unique about this implementation.

Concrete examples of the general use of the ordered list concept are not hard to come by (permutations versus combinations).

Basic properties of the ordered pair: The basic properties of the ordered pair concept are...

for any objects x, y there is a pair (x, y)

$(x, y) = (z, w)$ if and only if $x = z$ and $y = w$.

We can define the ordered triple $[x, y, z]$ (for example) as $(x, (y, z))$ (in terms of ordered pairs) but we could equally well define it as $((x, y), z)$ and there is not a bad case for defining it as $\{(1, x), (2, y), (3, z)\}$ (the last definition is easier to generalize to ordered lists of arbitrary length).

We won't just now commit ourselves to one of these definitions of the triple (or of longer lists).

We do note that what we require of the general list concept is

for any $x_1, \ldots, x_n$ there is a list $[x_1, \ldots, x_n]$

$[x_1, \ldots, x_n] = [y_1, \ldots, y_n]$ iff $x_i = y_i$ where $1 \leq i \leq n$.

A definition of the ordered pair as a set (historical, easier than the usual one): This definition belongs at this point in the discussion, but is not important in the first lecture.

We could (but do not) use this definition:

For this paragraph only define (x, y) as $\{\{\{x\}, \emptyset\}, \{\{y\}\}\}$. This is the first definition of the pair as a set, given by Norber Wiener in 1914.

Our axioms ensure that (x, y) exists for any x, y.

Suppose $(x, y) = (z, w)$.

This means, $\{\{\{x\}, \emptyset\}, \{\{y\}\}\} = \{\{\{z\}, \emptyset\}, \{\{w\}\}\}$.

We need to show that this implies that $x = z$ and $y = w$.

Notice that $\{\{\{x\}, \emptyset\}, \{\{y\}\}\}$ has two elements, a set with two distinct elements, $\{\{x\}, \emptyset\}$, and a set with one element, $\{\{y\}\}$. The element with two elements has as its elements a singleton set and the empty set. The element of the singleton set is x. The element with one element has as its sole element a singleton set, whose element is y.

Now if $\{\{\{x\}, \emptyset\}, \{\{y\}\}\} = \{\{\{z\}, \emptyset\}, \{\{w\}\}\}$, this means that the element of $\{\{\{z\}, \emptyset\}, \{\{w\}\}\}$ with two elements has $\emptyset$ and $\{z\}$ as its only elements, but also has $\emptyset$ and $\{x\}$ as its only elements, so $x = z$. Similarly, the element of $\{\{\{z\}, \emptyset\}, \{\{w\}\}\}$ with one element has $\{w\}$ as its only element, but also has $\{y\}$ as its only element, so $w = y$.

The usual definition of the ordered pair: This definition belongs at this point in the discussion, but is not important in the first lecture.

It is usual (since Kuratowski, 1920) to define (x, y) as $\{\{x\}, \{x, y\}\}$. The definition of the pair as a set is simpler, but the proof that if $(x, y) = (z, w)$ then $x = z$ and $y = w$ is rather more complicated.

Notice that x is the only object which belongs to every element of $\{\{x\},\{x,y\}\}$, and y is the only object which belongs to exactly one element of $\{\{x\},\{x,y\}\}$. This fact can be used to show that this definition of the pair has the basic properties we expect.

Notice that as a discrete math student, you do not need to know the details of these proofs to work with the pair. The basic properties of pairs and lists are the interface you need to work with these types of object. But of course the proofs are proofs, and proofs in the abstract at an elementary level are also discrete math content.

The definition of the Cartesian product: Given sets A, B, we define the Cartesian product of A and B as the collection of all ordered pairs (a, b) with $a \in A$ and $b \in B$.

The existence of the Cartesian product (no matter which of the two definitions we use) is a consequence of the axioms we have already given.

We define $\mathcal{P}^2(A)$ as $\mathcal{P}(\mathcal{P}(A))$, and more generally $\mathcal{P}^{n+1}(A)$ as $\mathcal{P}(\mathcal{P}^n(A))$ for $n \geq 2$.

It is then straightforward to observe that for any $a \in A$, $b \in B$, both $\{a\}$ and $\{a, b\}$ belong to $\mathcal{P}(A \cup B)$ so $\{\{a\},\{a, b\}\}$ belongs to $\mathcal{P}^2((A \cup B)$, and so we can define $A \times B$ as the set of all x in $\mathcal{P}^2((A \cup B)$ such that for some $a \in A, b \in B$, $x = \{\{a\},\{a, b\}\}$.

The same strategy would work for the historical definition of the pair, but using $\mathcal{P}^3(A \cup B)$ instead.

The proofs that Cartesian products exist should not be important directly to you; the mere assertion that they exist should be enough, as you really shouldn't work directly with the details of ordered lists as sets very much (what you do with them should actually be quite independent of their implementation). What should have some interest is that an implementation is possible!

2 Informal remarks about cardinality and counting principles

At this point in the class, our application for sets is in presenting counting principles. Actual definitions of the concepts I introduce here will appear in future lectures but require a bit more work, but I give an informal summary here.

Informal definition: For any set A, $|A|$ is informally defined as the number of elements in A. $|\emptyset| = 0$ of course. Some sets are infinite, and for these, for the moment, we don't define $|A|$.

We will state a formal definition of $|A|$ later.

Disjoint sets: We say that sets A and B are disjoint iff there is no x such that $x \in A$ and $x \in B$. Equivalently, A and B are disjoint iff $A \cap B = \emptyset$. Note that "iff" is an abbreviation for "if and only if" common in mathematical text.

Additive principle: If A and B are finite sets (meaning, $|A|$ and $|B|$ are defined) and $A \cap B = \emptyset$ (in English, if A and B are disjoint) then $|A \cup B| = |A| + |B|$.

Additive principle with compensation for overcounting: If A and B are finite sets, $|A \cup B| = |A| + |B| = |A \cap B|$.

Multiplicative principle: If A and B are finite sets, $|A \times B| = |A| \cdot |B|$. Notice that an element of $A \times B$ can be used to represent a process of choosing an element of A, then choosing an element of B.

Notice that $|A \times B| = |B \times A| = |A| \cdot |B|$, because multiplication is commutative, but it is not true that $A \times B = B \times A$ (I'll do an example).

"Exponential principle": If we define A^n as the set of lists $[a_1, \ldots, a_n]$ of n elements taken from the set A, then $|A^n| = |A|^n$. So for example the set of six letter "words" made from the letters A,B,C,D,E, repetitions allowed, pronouncability not a value, which we can write $\{A, B, C, D, E\}^6$, has 5^6 elements.

This is a consequence of repeated application of the multiplicative principle.

A general list counting principle: If we are counting any set of lists $[a_1, a_2, \ldots, a_n]$ where there are k_1 choices for a_1, and given any choice of $a_1, \ldots, a_i$ there are k_{i+1} choices of values for a_{i+1}, the number of lists in the set is $k_1 \cdot k_2 \cdot \ldots k_n = \prod_{i=1}^{n} k_i$.

For example if we are counting words from the alphabet $\{A, B.C, D, E\}$ in which no letters are repeated, the number of such words is $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$, because there are 5 choices for the first letter, 4 for the second (given the first letter, not the same 4 choices in every case), 3 for the third letter (given the previous two choices), and so forth.

3 The rest of the material, not yet processed! Some of this we will use, some is advanced or off on a tangent

The definition of a relation: Let A, B be sets. A relation from A to B is a triple $R = (A, B, G)$ where G is a subset of $A \times B$, and where we define (x, y, z) (for this specific purpose) as $((x, y), z)$.

We call A the domain of R ($\texttt{dom}(R)$), B the codomain of R ($\texttt{cod}(R)$) and G the graph of R ($\texttt{graph}(R)$). There is another school of thought (which by temperament I prefer) which identifies a relation with its graph, but there are technical problems with this, because in general the domain and codomain cannot be determined from the graph, and it is common to speak of the domain and codomain as features of the relation.

We define $x\,R\,y$ as the assertion $(x, y) \in \texttt{graph}(R)$.

We define the image or range of R as the set of y in the codomain of R such that there is x such that $x\,R\,y$.

We define the preimage of R as the set of x in the domain of R such that there is y such that $x\,R\,y$.

Many but not all transitive verbs in mathematics can be read as relations. The most general ones, such as $x = y$, $x \in y$, $x \subseteq y$ cannot, because there are no sets large enough to serve as domain or codomain of these "logical relations".

No universal set, so logical relations are not always implemented as set relations:

We prove a theorem (we do not want to give credence to there being a "paradox" here, as people thought during a crisis of foundations at the beginning of the last century).

Let A be a set. Define $\texttt{Russell}(A)$ as $\{x \in A : x \notin x\}$.

We prove that $\texttt{Russell}(A) \notin A$.

We prove this by contradiction. Suppose $\texttt{Russell}(A) \in A$.

Now consider the status of the sentence $\texttt{Russell}(A) \in \texttt{Russell}(A)$.

This expands to $\texttt{Russell}(A) \in \{x \in A : x \notin x\}$

which is equivalent to $\texttt{Russell}(A) \in A$ and $\texttt{Russell}(A) \notin \texttt{Russell}(A)$

which is equivalent to $\texttt{Russell}(A) \notin \texttt{Russell}(A)$ if (as we have assumed) $\texttt{Russell}(A) \in A$ is true.

But this is absurd. So we have shown that $\texttt{Russell}(A) \in A$ cannot be true (and so that $\texttt{Russell}(A) \notin \texttt{Russell}(A)$), which has the more general consequence that there is no set V such that every object x is a member of V.

It follows that there can be no set relation implementing equality, membership or the subset relation, as the domain of any such relation would have to be V, the nonexistent universal set.

It also follows that for no set A can there be a set of all x not belonging to A (a true complement of A). If such a set existed, its union with A would be V. If a working universe U is understood in a particular context, $U \setminus A$ will play the role of the complement of $A \subseteq U$; but it doesn't contain everything that is not in A.

the definition of functions: A function is a relation $F = (A, B, G)$ with the property that for each $x \in A$, there is exactly one $y \in B$ such that $x\,F\,y$.

We expand the language a little: it is equivalent and perhaps easier to follow to say that for each $x \in A$, there is $y \in B$ such that $x\,F\,y$ (so the preimage of F is the domain) and for any x, y, z, if $x\,F\,y$ and $x\,F\,z$, then $y = z$.

If F is a function and $x \in \texttt{dom}(F)$, we define $F(x)$ as the unique y such that $x\,F\,y$. [If we had the definite description operrator, we could write $F(x) = \theta(\{y \in \texttt{cod}(F) : x\,F\,y\})$].

It is worth noting that there are "logical functions" which are not implementable as sets. For example, there can be no function F such that $F(x) = \mathcal{P}(x)$ for all x, since the domain of such a function would be the collection of all sets, which can be shown not to exist by the Russell argument.

You learned a quite different definition of the function concept in high school and in college calculus. We illustrate that our formalization is adequate to support that informal definition.

We set out to define the function $y = 2x + 5$ from real numbers to real numbers. We suppose that we have the set $\mathbb{R}$ of real numbers handy.

We then have the graph of f as the set G of all u in $\mathbb{R} \times \mathbb{R}$ such that there is $x \in \mathbb{R}$ such that $u = (x, 2x + 5)$.

Some logical notation and an upgrade to set builder notation: We define the sentence $(\forall x \in A : P(x))$ as $\{x \in A : P(x)\} = A$. This is read, for all $x \in A$, $P(x)$.

We define the sentence $(\exists x \in A : P(x))$ as $\{x \in A : P(x)\} \neq \emptyset$. This is read, for some $x \in A$, $P(x)$, or there exists $x \in A$ such that $P(x)$.
[3]

Defining quantifiers in terms of sets might be taken as an odd maneuver. It is equally odd in discrete math texts (I think odder) that quantifiers are often introduced before sets are introduced...but with set bounds as here, so they depend on informal understanding of sets anyway.

We can then define $\{(x, y) \in A \times B : P(x, y)\}$ as

$$\{u \in A \times B : (\exists x \in A : (\exists y \in B : u = (x, y) \wedge P(x, y)))\}.$$

This can be used as a general model for how to treat complicated expressions appearing left of the colon in set builder notation.

And then we can say in general that a function definition $y = F[x]$ where $F[x]$ stands in for some complicated expression in x is equivalent to

$$f = (A, B, \{(x, F[x]) \in A \times B : x \in A\}),$$

where A is the intended domain (often implicit in a definition of this kind), B is the intended codomain, and the definition only succeeds if for every $x \in A$ it is the case that $F[x] \in B$ (though this can be qualified: in calculus you often work with partial functions, which may be undefined at some elements of the implicitly understood domain: the calculus definition of domain is more analogous to what we call preimage above).

[3]Further, $(\forall x \in A : P(x) \to Q(x))$ is definable as $\{x \in A : P(x)\} \subseteq \{x \in A : Q(x)\}$ ($P \to Q$ can be defined as $(\forall x \in A : P \to Q)$, x not occurring in P, Q and A nonempty). This is not as absurd as it looks: quantified implication was defined first in the actual history! Similar maneuvers can define the other propositional connectives: we content ourselves with observing that $(\forall x : \neg P(x))$ is definable as $\{x \in A : P(x)\} = \emptyset$ and that all the propositional connectives can be defined in terms of negation and implication.

The notation $(x \in A \mapsto F[x] \in B)$ is convenient for this. The appearance of $\in A$ here is unusual, and the appearance of $\in B$ is very unusual. We may omit either the domain or codomain specification if they can be understood from context.

A very strong axiom of set theory (usually called the Axiom of Replacement) says that $(x \in A \mapsto F[x])$ always exists as a set (is $(x \in A \mapsto F[x] \in B)$ for some codomain B) for any set A and any method $F[x]$ of defining an objects with parameter x. Obviously this is convenient: it is also an enormously strong assumption, having far more strength than its introduction as a mere convenience here would suggest.

We will call it

Axiom F (function abstraction): For any notation $F[x]$ for an object defined in terms of a parameter F, and any set A, there is a set B such that $(x \mapsto A : F[x] \in B)$ exists.

If we allow the notation $(x \in A \mapsto F[x])$ (which we can, even if we do not assume that it always exists) we provide that the codomain is the image (take the smallest codomain if it is not explicitly given).

We do not assume this axiom in general, and will always mention it if we assume it in a particular context.

Some kinds of function which are commonly considered: A function f is an injection, or one-to-one, if for any $x, y \in \mathtt{dom}(f)$, if $f(x) = f(y)$ then $x = y$.

A function f is a surjection, or onto, if for any $y \in \mathtt{cod}(f)$, there is x such that $y = f(x)$.

A function f is a bijection iff it is one-to-one and onto (i.e., an injection and a surjection).

We can now state our

Axiom A (sets of individuals): For any set B, there is a set A all of whose elements are individuals such that there is a bijection (A, B, G).

We don't necessarily assume this officially, but we have an intellectual use for it.

Reasons why we should only identify functions with their graphs with care:

Suppose we did identify functions with their graphs. Two problems would arise for our presentation.

The notion of surjection would not make sense (injection still would). The difficulty is that the codomain of a function cannot be deduced from its graph. Many discrete math books actually define relations and functions as sets of ordered pairs and then try to define onto/surjection as above, which is an error. It is a disgrace that this mistake is so common, I cannot understate this.

One would have to define "f is onto B" as "for any y in B there is x such that $y = f(x)$", for any set B including the range of f (which can be computed from the graph), and define surjection from A to B rather than surjection in any absolute sense.

The above is a general problem and something to watch out for in math texts.

The second problem is particular to our implementation, and is the reason I chose to use explicit domains and codomains here, though by temperament I prefer to identify a function with its graph.

Suppose that G is a set of ordered pairs. How can we show that there are sets A, B such that (A, B, G) is a relation? The quick answer is that we cannot in general with the axioms for set theory which we have presented. We briefly indicate what is needed.

A stronger presentation of the axioms allows a more general construction of unions: given any set A, we can construct the union of all of its elements, which can be written $\bigcup A$, the set of all x such that there is $a \in A$ such that $x \in a$. You can check that $\bigcup\{A, B\}$ is our $A \cup B$: the usual axiomatization of set theory asserts that all unordered pairs $\{x, y\}$ exist and that $\bigcup A$ exists for every set A. This is a much more sophisticated notion of union than the notion of "binary union" which I have used, which should actually already be familiar to you. And understanding the notion of domain and codomain of a relation really does not depend on this more sophisticated notion of union.

This can be codified in a stronger axiom:

Axiom 5b (set union): For any set A, there is a set $\bigcup A$ such that for all x, $x \in \bigcup A$ if and only if there is $a \in A$ such that $x \in A$.

Given this concept, the preimage and the image of a function with graph G an arbitrary set of ordered pairs can be defined as subsets of $\bigcup(\bigcup G)$ (three unions if we used the historical definition of the pair), so we can present functions with that graph.

It is useful to notice that the foundations we are presenting here are actually not mathematically weaker than the usual ones. If $X \subseteq \mathcal{P}(A)$ for some A, then $\bigcup X$ is definable as $\{x \in A : (\exists Y \in X : x \in Y\}$, which exists by our axioms. The stronger union construction works as long as the collection of sets we consider is known to be a subset of a fixed power set, and this is actually the case in most mathematical applications.

If we have the occasion to use the stronger form of the axiom of union, we will mention this explicitly (as the axiom of Set Union) rather than Binary Union, our basic axiom).

It is worth noting that in the usual axiomatization of set theory, there is an axiom of pairs (not singletons) providing $\{x, y\}$ for each pair of objects x, y and the axiom of set union, and $A \cup B$ exists because it is $\bigcup\{A, B\}$.

Our official definition of ordered lists: We decouple lists from ordered pairs. We use the notation $[x_1, x_2, \ldots, x_n]$ to make this clear.

We define an ordered list as a graph of a function whose domain is an interval in the integers (assuming familiarity with the integers; our treatment below will at least suggest how the integers themselves could fit into our scheme). This allows variations in how they are indexed. If x is an ordered list, x_i is then simply defined as $x(i)$. The list $[x, y, z]$ is for us the set $\{(x, 1), (y, 2), (z, 3)\}$ (or it might be $\{(x, 0), (y, 1), (z, 2)\}$ if the context tells you our indexing starts at 0).

Note that this definition supports lists of length 0 and 1 (and lists $[x, y]$ of length 2 are not the same as ordered pairs (x, y)).

Note that if a is a set, $\bigcup\{a\} = a$. We can define $\theta(u)$ as $\bigcup\{x : x \in u \wedge |u| = 1\}$ for all sets of sets u: notice that if u does not have exactly one element, this set is empty! To get this operator for general sets would require an axiom.

The natural numbers as an abstraction: We give a very abstract axiom of infinity.

Axiom 6 (infinity): There is a set $\mathbb{N}'$, an element 0 of $\mathbb{N}'$, and an injection σ from $\mathbb{N}'$ to $\mathbb{N}' \setminus \{0\}$.

Notice that $0, \sigma(0), \sigma(\sigma(0)), \ldots$ are all different. The intention here is that σ implement the operation "add one".

The collection of these objects would be an abstract implementation of the natural numbers.

Notice that $\mathbb{N}'$ might have elements not in this set. For example, nothing prevents $m \in \mathbb{N}'$ with $\sigma(m) = m$.

But we can define the subset of $\mathbb{N}'$ we want, though this may not be obvious.

We call a subset I of $\mathbb{N}'$ *inductive* just in case $0 \in I$ and for every $k \in \mathbb{N}'$, if $k \in I$ then $\sigma(k) \in I$. This definition should look suspiciously familiar if you have ever seen a proof by mathematical induction.

We define $\mathbb{I}$ as the set of inductive subsets of $\mathbb{N}'$ (which does exist as a subset of $\mathcal{P}(\mathbb{N}')$).

We then define $\mathbb{N}$ as $\{n \in \mathbb{N}' : (\forall I \in \mathbb{I} : n \in I)\}$.

Notice that if the set $\{0, \sigma(0), \sigma(\sigma(0)), \ldots\}$ exists at all, this has to be it. The set I have just given informally is certainly inductive, and any element of it clearly has to belong to any inductive set.

We then understand $\mathbb{N}$ as the set of natural numbers, 0 as the natural number 0 (we could have started at 1, but we want natural numbers to be usable as cardinalities of finite sets, including the empty set), and σ as the function which adds one to a natural number.

Notice that we have said nothing at all about whether natural numbers are sets or individuals.

We can now define 1 as $\sigma(0)$, 2 as $\sigma(1)$, 3 as $\sigma(2)$, and so forth. A real definition of our usual notation for natural numbers waits on definitions of addition and multiplication.

Iterating application of functions: For any function $f = (A, A, G)$ and $a \in A$, define $\texttt{iter}(f)$ as the intersection of all sets included in $\mathbb{N} \times A$ which contain $(0, a)$ and if they contain (n, x) also contain $(\sigma(n), f(x))$.

It should be reasonably clear intuitively that $(\mathbb{N}, A, \texttt{iter}(f))$ is the function which sends a natural number n to $f^n(a)$, the result of applying

f n times to a. I'll provide a proof that it is in fact a function in an appendix: this is an application of mathematical induction in an extremely abstract context.

Once we have this machinery, we can define for any natural numbers m and n, $m + n = \sigma^n(m)$ (apply the successor operation n times to m) and in a backward application we see that $n + 1 = n + \sigma(0) = \sigma^{\sigma(0)}(n) = \sigma(\sigma^0(n)) = \sigma(n)$.

Further, we can define mn as $(\sigma^m)^n(0)$ (σ^m is the function "add m": so this tells us to add m n times).

Once we have the definitions of addition and multiplication, we can define the usual notation for natural numbers explicitly, something which might again appear in an appendix.

Proofs of the basic results about addition and multiplication (or proofs of some representative ones and statement of others) will be forthcoming. This will require the discussion of the basic Peano axioms of arithmetic, including mathematical induction.

In addition, we will use our result about iteration, which we package as a theorem and plan to prove in an appendix.

Iteration Theorem: Let f be a function from A to A and let $a \in A$. Then there is a uniquely determined function $\texttt{iter}(f, a)$ from $\mathbb{N}$ to A such that $\texttt{iter}(f, a)(0) = a$ and for every $n \in \mathbb{N}$,

$$\texttt{iter}(f, a)(n + 1) = f(\texttt{iter}(f, a)(n)).$$

We use the notation $f^n(a)$ for $\texttt{iter}(f)(a)$ and restate: $f^0(a) = a$; for any n, $f^{n+1}(a) = f(f^n(a))$.

We also note (we did this informally above) that we can use f^n as notation for the function $\{(x, f^n(x)) : x \in A\}$.

The historical concrete definition of natural numbers: The original statement of the axiom of infinity in Zermelo's 1908 paper which founded the subject in its modern form as "there is a set Z which contains $\emptyset$ and contains $\{x\}$ for each $x \in Z$".

This gives the situation in our axiom of infinity in a concrete form, with $\mathbb{N}' = Z$, $0 = \emptyset$, and $\sigma = (Z, Z - \{\emptyset\}, \{(x, \{x\}) : x \in Z\})$. That σ is an

injection [if it exists] is evident: if $\{x\} = \{y\}$ it follows that $x = y$, so the function is injective

This definition identifies the natural numbers with particular sets $\emptyset, \{\emptyset\}, \{\{\emptyset\}\} \ldots$.

The usual concrete definition of natural numbers: The definition due to von Neumann is the one used now. The axiom of infinity in our format would assert that there is a set N such that $\emptyset \in N$ and if $x \in N$ then $x \cup \{x\} \in N$.

This is very elegant. 0 is defined as the empty set. 1 is defined as $\{0\}$. 2 is defined as $\{0, 1\}$, 3 is defined as $\{0, 1, 2\}$ and so forth.

This has formal advantages. The one sticky bit is that the proof that for natural numbers x, y if $x \cup \{x\} = y \cup \{y\}$ then $x = y$ is rather tricky. The problem is that without additional assumptions, this isn't necessarily true as a general fact about sets (though a fairly simple assumption enforces it: it seems quite natural to forbid two distinct sets from being members of each other, and that causes this implication to work for general sets. No such axiom is needed to prove the theorem for von Neumann natural numbers, it just makes things tidier).

We do not define the natural numbers: Our view is that it isn't necessary or even appropriate to dictate a concrete definition of the natural numbers. If we stay in the abstract format we have presented, no questions about what members a natural number has can legitimately arise. It is useful to see that there are purely set theoretical axioms which give rise to an implementation, but we prefer to stay more abstract.

We do point out that the von Neumann definition has the very nice feature that the size of the set implementing n is n. This can be quite convenient.

We also point out that it is very useful to define the natural numbers in a context including sets and functions. Our formalization makes it clear why mathematical induction works, for example (the numbers are defined as the set on which induction works!)

Counting: This is what numbers are for!

Let X be a set. For any $A \subset \mathcal{P}(X)$, we define $+_X(A)$ as

$$\{a \cup \{x\} \in \mathcal{P}(X) : a \in A \wedge x \in X \setminus a\}.$$

Visiting a notational issue: the notation

$$\{a \cup \{x\} \in \mathcal{P}(X) : a \in A \wedge x \in X \setminus a\}$$

expands out to

$$\{u \in \mathcal{P}(X) : (\exists a \in A : (\exists x \in X \setminus a : u = a \cup \{x\}))\}.$$

This is a sample of how to interpret such locutions with complex terms left of the colon in set builder notation.

We define $[X]^n$, for $n \in \mathbb{N}$, as $+_X^n(\{\emptyset\})$. We call this "the set of subsets of X with n elements", and from a common sense standpoint that is certainly what it seems to be.

We prove a theorem: if $A \in [X]^n$ and $B \in \mathcal{P}(X)$, then $B \in [X]^n$ if and only if there is a bijection $f = (A, B, G)$ from A to B.

We do this by showing that the set of natural numbers n for which this statement is true is inductive, and so must contain all natural numbers.

0 is in this set: $+_X^0(\{\emptyset\})$ is simply $\{\emptyset\}$, so if $A, B \in +_X^0(\{\emptyset\})$ it follows that $A = B = \emptyset$ and there is a bijection from the empty set to the empty set; moreover, any bijection with domain the empty set has codomain the empty set.

Suppose that k is in the set, i.e., if $A \in [X]^k$ and $B \in \mathcal{P}(X)$, then $B \in [X]^k$ if and only if there is a bijection $f = (A, B, G)$ from A to B. Our goal is to show that under this hypothesis, if if $A \in [X]^{k+1}$ and $B \in \mathcal{P}(X)$, then $B \in [X]^{k+1}$ if and only if there is a bijection $f = (A, B, G)$ from A to B.

Suppose $A \in [X]^{k+1}$. This means that $A = C \cup \{x\}$ for some $C \in [X]^k$ and $x \in X \setminus C$.

Suppose further $B \in [X]^{k+1}$. This means that $B = D \cup \{y\}$ for some $D \in [X]^k$ and $y \in X \setminus D$.

We know by hypothesis that there is a bijection (C, D, G). $(C \cup \{x\}, D \cup \{y\}, G \cup \{(x, y)\})$ is a bijection from A to B.

Now suppose that there is a bijection f from A to E. $(C, E \setminus \{f(x)\}, \texttt{graph}(f \setminus \{(x, f(x))\})$ is a bijection from C to $E \setminus \{f(x)\}$ so by hypothesis $E \setminus \{f(x)\}$ is in $[X]^k$, so E is in $[X]^{k+1}$. So we have shown both directions of the implication: the set of n for which the local instance

of the theorem is true is inductive and so includes all natural numbers, so the theorem is true in general.

cardinality: In general, for any sets A, B, we say that A and B are "the same size" or have the same cardinality just in case there is a bijection $f = (A, B, G)$ from A to B.

For A, X sets, we define $|A|_X$ as the set of all $B \in \mathcal{P}(X)$ such that B is the same size as A, if this is nonempty. It is not required that A be a subset of X, but it is required that A be the same size as some subset of X for this notation to be meaningful.

Notice that the previous theorem asserts in effect that if $A \in [X]^n$, then $|A|_X = [X]^n$.

Now we can define sums and products of cardinals relative to X.

$|A|_X \cdot |B|_X = |A \times B|_X$ (this is useful for discussions of cartesian product).

$|A|_X + |B|_X = |(A \times \{0\}) \cup (B \times \{1\})|_X$ (this is actually harder: the issue of disjointness of sets for addition and the ability to make disjoint copies of sets come up here.

Then we can prove by induction that $[X]^m + [X]^n = [X]^{m+n}$ and $[X]^m \cdot [X]^n = [X]^{mn}$ where addition and multiplication of natural numbers are defined as above.

This will make verification of basic algebra principles much easier than proving them purely by induction.

a set theoretic approach to "proof": We present a recursive definition (a complex example of structural recursion) of sets $\mathtt{ev}(A)$ of evidence for the truth of a sentence A. The idea is that a sentence is true iff $\mathtt{ev}(A)$ is nonempty. The construction of $\mathtt{ev}(A)$ is based on the logical form of the sentence A. We don't just now talk about what mathematical objects we identify with sentences of our language, but we do note that we are making the strong assumption that every object has a name.

Managing this recursion in our set theory and language is a bit tricky.

$\mathtt{ev}(x = y)$ is the set $\{x\} \cap \{y\}$.

$\mathtt{ev}(x \in A)$ is the set $\mathtt{x} \cap A$.

This handles the base cases. Notice that these sets have elements just in case the sentences are true.

We assume that we know how to define $\mathtt{ev}(P)$ and $\mathtt{ev}(Q)$.

$\mathtt{ev}(P \wedge Q)$ is $\mathtt{ev}(P) \times \mathtt{ev}(Q)$: evidence for "$P$ and Q" is taken to be an ordered pair of an item of evidence for P and an item of evidence for Q.

$\mathtt{ev}(P \vee Q)$ is $\mathtt{ev}(P) \times \{0\} \cup \mathtt{ev}(Q) \times \{1\}$. Evidence for "$P$ or Q" is either evidence for P labelled with 0 or evidence for Q labelled with 1.

$\mathtt{ev}(P \to Q)$ is the collection of functions from $\mathtt{ev}(P)$ to $\mathtt{ev}(Q)$. Evidence for "if P then Q" takes the form of a machine which takes evidence for P as input and gives evidence for Q as output.

$\mathtt{ev}(\neg P)$ is the collection of functions from $\mathtt{ev}(P)$ to $\emptyset$. Notice that this set has exactly one element if there is no evidence for P and otherwise has no elements, which correlates with what we mean by "not P".

$\mathtt{ev}((\forall x \in A : P(x))$ is a function f from A to $\bigcup_{a \in A} \mathtt{ev}(P(a))$ (it is easy to show that this set exists using strong union and replacement, as the set union of the image of $(a \in A : \mathtt{ev}(P(a)))$; it actually exists anyway for all expressions on our language, but proving this is extremely involved) such that $f(a) \in \mathtt{ev}(P(a))$.

$\mathtt{ev}((\exists x \in A : P(a))$ is the set of pairs (a, p) where $p \in \mathtt{ev}(P(a))$: this has as a bounding set $A \times \bigcup_{a \in A} \mathtt{ev}(P(a))$, so the same set existence question arises.

I believe that the existence of $\bigcup_{a \in A} \mathtt{ev}(P(a))$ can be shown by structural induction on formulas [this is a matter for an appendix]. This makes the point that we do not need full union or even replacement, as a strictly formal consideration of the situation might suggest (and this is beyond the students).

To get existence of proofs of obvious things, we actually need a form of the axiom of choice. A convenient axiom is, postulate an operation `choose` such that for any nonempty set A, $\mathtt{choose}(A) \in A$.

Axiom C (choice): We postulate a construction `choose` such that for any nonempty set A, $\mathtt{choose}(A) \in A$.

We can then define a function `best` taking each A to a preferred member of $\texttt{ev}(A)$. We could use $\texttt{choose}(\texttt{ev}(A))$ of course, but the definition of `best` that we give makes it clear that the only issue is with existential quantifiers.

The best evidence for an atomic sentence is the only evidence.

The best evidence for a conjunction is the pair of the best evidences for the conjuncts.

The best evidence for a disjunction is constructed from evidence for the first disjunct if there is any, and otherwise from evidence for the second disjunct.

The best evidence for $P \to Q$ takes each item of evidence for P to the best evidence for Q (if there is any).

The best evidence for $(\forall x : P(x))$ returns, given input x, the best evidence for $P(a)$.

The best evidence for $(\exists x : P(x))$ is $u = (\texttt{choose}(\{x \in A : P(x)\})$ paired with the best evidence for $P(u)$.

Notice that if we defined evidence for $(\exists x : P(x))$ as evidence for $\neg(\forall x : \neg P(x))$, the definition of `best` would not require the use of choice.