

# Set Theory

Notes on H Enderton's *Elements of Set Theory*

## I. Introduction

### A. What are sets?

#### 1. The idea

Sets are “collections”. The objects “in” the collection are its members. E.g., we are all members of the set of all humans.

Anything can be in a set. There are sets of numbers, people, other sets.

Notation: we write names of sets like this:

|  |  |
|--|--|
| $\{0,1\}$                                      | The set containing 0 and 1   |
| $\{\text{Ted}, \{\text{New York City}\}, 57\}$ | The set containing me, the set containing just NYC, and the number 57                    |
| $\{0, 1, 2, \dots\}$                           | The set containing all the natural numbers (note: this set has infinitely many members.) |
| $\{x: x \text{ is an even number}\}$           | The set containing the even numbers (i.e., $\{0, 2, 4, \dots\}$ )                        |

It's often intuitive to speak of “collecting” things together and “putting them into” a set, but the idea is that the sets exist whether or not anyone has ever done any collection. Moreover, notice that a set's members don't need to be similar, or thought of by people, or anything like that. Except for certain limitations we'll discuss later, *any* objects form a set.

#### 2. M&E

Sets raise lots of interesting metaphysical questions. What are they? What is the membership relation — in what sense are the members of the set “in” it? How do you get *one* thing, the set, from *many* things? (This is, in essence, the question of the nature of the membership relation restated.)

Sets also raise epistemological questions: how do we know about them? Etc.

But we won't take up any of these; we'll just study the mathematics of set theory.

### B. What is set theory?

Set theory is a theory — a bunch of principles — about what sets are like. Why study set theory? Because sets and their theory are important foundational tools in mathematics, linguistics and philosophy. The idea of collecting objects into one turns out to be extremely useful.

In linguistics, for example, one can think of the meaning of a predicate, ‘is red’ for instance, as a set — the set of all red things. Here’s one hint of why sets are so useful: we can apply this idea to predicates that apply to other predicates. For instance, think of ‘red is a color’. ‘Is a color’ is a predicate, and it looks like it holds of things like the meaning of ‘is red’. So we can think of ‘is a color’ as meaning a set containing all the colors. And these colors in turn are sets: the set of all red things, the set of all green things, etc. This works because i) a set collects many things into one, and ii) many sets themselves can be put together to form a new set.

In mathematics, sets are convenient because all mathematical structures can be regarded as sets. Later in this class, for instance, we’ll look at how the natural numbers can be defined as sets.

Furthermore, the mathematical theory of sets is very interesting in its own right. The idea of a set is a very clear and simple one. But coming up with a consistent theory of sets turns out to be pretty hard. Moreover, once we articulate some natural assumptions about sets, lots of interesting results can be proved about them. For example, we can develop a rigorous theory of infinite collections. Infinity has always fascinated philosophers. Zeno’s paradoxes (e.g., Achilles and the Tortoise.) Hilbert’s hotel. The nature of space and time. Set theory lets us show that the idea of the infinite isn’t contradictory.

### C. Baby set theory

**Extensionality:** if sets have exactly the same members then they are identical

(Intuitive idea: sets are just defined by their members. So, intensional differences are irrelevant. E.g.: the set of teachers of this class = the set of 38 year old people in this room)

**Empty set, Singletons, Pair sets, 3-membered sets, Power set, etc.**

(Note: don’t get careful yet. Just talk about these sets, don’t discuss the need to justify their existence.)

Examples:  $\emptyset$ ,  $\{\emptyset\}$ ,  $\{\emptyset, \{\emptyset\}\}$ .  $\{\{\emptyset\}, \emptyset\} = \{\emptyset, \{\emptyset\}\}$ .  $\emptyset \in \{\emptyset\}$  and  $\{\emptyset\} \in \{\{\emptyset\}\}$ , but  $\emptyset \notin \{\{\emptyset\}\}$ .  $\wp(\{0,1\}) = ?$

**Operations: Union, Intersection**

**Relations: disjoint, subset, proper subset.**

- Difference between subset and membership.
- Null set subset of anything.

**Abstraction:**  $\{x: \phi(x)\}$  E.g.,  $\{x: x \text{ is an even number}\}$

Examples: write names of null, singleton, pair, power, in abstraction notation.

D. Paradoxes

Now we run into trouble.

Berry's paradox:  $\{x: x \text{ is the least integer not definable in one line of type}\}$  (Note that this problem isn't particular to set theory; you get a problem with just the description 'the least integer not definable in one line of type'. So what was Enderton thinking?)

Russell's paradox:  $\{x: x \notin x\}$

What has gone wrong? Compare a much simpler "paradox", involving the Barber who shaves all and only those that don't shave themselves. The solution is simply that there can be no such Barber. (The statement that he exists is a logical contradiction.) What we have done here is blithely *assume* that **there exists** a set of a certain sort — something is a member of it iff it is not a member of itself. But this statement too — namely,

$$\exists y \forall x (x \in y \leftrightarrow x \notin x)$$

is also a logical contradiction (instantiate  $\forall x$  to  $y$ ).

We were in effect assuming a lot of principles about sets. The assumption that generated the problem was the assumption that the abstraction notation always defines a set. What we've learned is that sometimes it doesn't.

So we need to get much more careful about the assumptions we make about sets. We need a theory of what sets exist, a theory that we're confident doesn't lead to trouble. We're going to do this by being very careful about what assumptions we make about the existence of sets. Specifically, we're going to use the **axiomatic method**. This means, we're going to write down a very few secure principles about what sets exist and what they are like — the axioms. And we won't assume anything about sets — in particular, we won't assume that a set of a certain sort exists — unless the assumption can be proved from the axioms.

E. Iterative hierarchy

The theory we're going to consider is Zermelo-Frankel set theory. It is based on a certain picture of sets — the “iterative hierarchy”. That is the picture of the iterative hierarchy will guide our choices of axioms. Here I'll describe the picture that motivates the theory.

<draw the picture — with atoms(=nonsets)>

Let me explain this picture. First, we start with the atoms at the bottom (also called urelements).  $V_0$

Then we get another level, by adding to  $V_0$  all the sets of atoms. I.e.,  $V_1 = V_0 \cup \wp(V_0)$ . Then  $V_2$  is what we get by adding the power set of  $V_1$ . Etc. The levels are getting bigger and bigger: each is a subset of the rest, but new sets are added each time.

That gets us the finite  $V$ 's. But we also want infinite sets. So we form a new level,  $V_\omega$ , by taking the *union* of *all* the previous levels. So  $V_\omega$  has all the finite  $V_i$ 's in it.

Then we construct  $V_{\omega+1}$  by adding to  $V_\omega$  its power set. Etc. But we don't stop. There's an infinite sequence of steps where we add the power set to the previous level; but then put all the sets in that sequence into another set, and start over again. And — in a sense we'll explore later, we do this sort of thing as much as possible. We'll discuss later what this means exactly.

So the idea of the picture is that we have a series of levels of sets, each containing all members of the earlier ones, and they're constructed in a certain way (some by taking power sets of the previous member, others by taking the union of all the previous members.)

And now, here's the idea of ZF:

**Intuitive idea of ZF:** every set is a member of  $V_\alpha$  for some  $\alpha$

i.e., every set shows up somewhere in the iterative hierarchy.

Two examples. First, there will always be a pair set, on this picture. For take any  $a$  and  $b$ . Each is at some level. But the levels are cumulative, so each is in whichever level is higher. But then, their pair set is in the next level (since the next level adds all sets that can be formed from the previous level.) Second, there's no set of all sets, since that wouldn't ever show up in the hierarchy. At each level, new sets are always added.

NOTE: we're going to be doing this with no atoms. So our picture is <draw it - V-shaped>

## F. Classes vs sets

Some set theories work differently from ZF. They say there are two kinds of entities, classes and

sets. There is a class that contains all the sets as members, it's just not a set. This is the Von-Neuman, Bernays approach. The ZF approach just says there is nothing that contains all the sets.

Note, though, that it's sometimes harmless to speak of, e.g., the "class of all things", as loose talk, either for its defining condition 'x=x', or for the things plurally (the self-identical things).

## II. Axioms and Operations

### A. First axioms

**Extensionality axiom, Empty set axiom, Pairing axiom, Union axiom** (for any a and b, there exists a set of a's and b's members), **Power set axiom**

### B. Justifying singular terms

Standardly (in logic and mathematics), use of names presuppose a unique referent. So, the use of '∅' for "the null set", i.e., "the set with no members" should be used only if we can prove from the axioms that i) there exists at least one set with no members, and ii) there exists no more than one such set. Well, the empty set axiom gives us i), and extensionality gives us ii).

Likewise, we can justify these singular terms:

|                       |  |
|-----------------------|--|
| {a,b}                 | the set containing just a and b  |
| $a \cup b$            | the union of a and b — i.e., the set whose members are only those sets that are members of a or members of b |
| $\wp a$               | the power set of a — the set of a's subsets  |
| $\{x_1, \dots, x_n\}$ | the set containing just $x_1, \dots, x_n$ (get this by a series of singletons and unions)                    |

### C. Abstraction and subset axioms

What led to Russell's paradox? The assumption that any old condition determines a set. I.e., that for any formula  $\phi$ , we're always guaranteed that there is such a set as the set of all the  $\phi$ s. I.e., that the name  $\{x: \phi\}$  is legit.

Let's state this assumption more carefully. Really, we have not just one assumption; we have many assumptions, for different choices of  $\phi$ . Now, sometimes we want the assumption to say that the set consisting of the  $\phi$ s is a function of other sets. E.g., the assumption in the case of pair sets looks like this:

$$\forall t_1 \forall t_2 \exists B \forall x (x \in B \leftrightarrow (x=t_1 \vee x=t_2))$$

So here's what the assumptions look like in general:

Naive comprehension schema:  $\forall t_1 \dots \forall t_n \exists B \forall x (x \in B \leftrightarrow \_)$

where the blank  $\_$  can be filled in with any formula including the free variables  $t_1, \dots, t_n$ , and  $x$ .

**Naive comprehension:** the result of putting in any formula (with no more than  $t_1, \dots, t_n$  and  $x$  free) into the naive comprehension schema is true

Russell's paradox refutes NC. So we can't include it as an axiom. *But:* we still want to be able to use the  $\{x \mid \dots\}$  notation. If we could never pick out the set of all the  $\phi$ s, then our set theory would be crippled.

Here's the way we do it. Think of the iterative hierarchy. We can't choose a set of *all* the  $\phi$ s, since that may not show up at any one point in the hierarchy. But if we *begin* with some set,  $c$ , in the hierarchy, we should be able to consider the set of all the  $\phi$ s *that are in*  $c$ . For this set will be in the same level of the hierarchy as  $c$ ; it will be a subset of  $c$ .

So here's our new principle:

**Subset axioms:** the result of filling in the blank in the following schema with any formula (including no free variables other than  $t_1 \dots t_n$  and  $x$ ) is an axiom:

$\forall t_1 \dots \forall t_n \forall c \exists B \forall x (x \in B \leftrightarrow (x \in c \& \_))$

That is, if we start with a  $c$ , the subset of  $c$  picked out by  $\phi$  is guaranteed to exist. (We can think of this in terms of classes: for any class and any set, the intersection of the class and the set is guaranteed to be a set.)

We can name sets thus guaranteed to exist this way:  $\{x \in c \mid \_\}$ . These singular terms are acceptable because i) a subset axiom tells us that there is at least one such set, and ii) extensionality tells us that there's no more than one.

We've blocked the Russell paradox. Since we don't have naive comprehension, we don't have any guarantee that there is a set consisting of all the sets that aren't members of themselves. True, for any set,  $c$ , we do learn that there is the following set  $B$ :  $\{x \in c \mid x \notin x\}$ . So we know that something is a member of  $B$  iff it is in  $c$ , and not a member of itself. But now, can we get a contradiction from asking whether  $B \in B$ ? No. We can consistently suppose that  $B \notin B$ . (In fact, we'll later introduce axioms that say that sets are never members of themselves. This gels with the picture of the iterative hierarchy, according to which sets are always built out of earlier members in the hierarchy.) This supposition would lead to contradiction if  $B$  were a member of  $c$ , since  $B$  would then satisfy the defining condition of  $B$ . But we have no way of deriving the result that  $B \in c$ . (And we'd expect that it isn't, given the iterative picture. The set of members of  $c$  that aren't members of themselves should be all of  $c$ . So  $B=c$ . But then, given the iterative picture,  $B \notin c$ .)

#### D. Consequences of the subset axioms

Let's work with the subset axioms a bit, to get the feel of them.

*Intersections.* One thing is that we don't need a separate axiom assuring us that intersections exist. For the intersection  $a \cap b$  is  $\{x \in a \mid x \in b\}$ . (It's also  $\{x \in b \mid x \in a\}$ .) So a subset axiom tells us that for every  $a$  and  $b$ , there exists a set whose members are just the elements of both  $a$  and  $b$ ; and extensionality tells us that there's no more than one such set.

*Complements.* Given any sets  $A$  and  $B$ , the complement  $A - B$  means the set of things that are in  $A$  but not  $B$ . It must exist, given the subset axioms; it is  $\{x \in A \mid x \notin B\}$ .

**Theorem 2A:** there is no set to which every set belongs

Pf: suppose there is such a set,  $A$ . We can use a subset axiom to construct the following set  $B$ :  $B = \{x \in A \mid x \notin x\}$ . Now, since  $A$  contains every set, we can simply write:  $\forall x (x \in B \leftrightarrow x \notin x)$ . But instantiate  $x$  to  $B$  and then we get a contradiction.

#### E. Arbitrary unions and intersections

So far, our notions of union and intersection apply to just two sets at a time. The natural generalization is to not just two, but any set of sets. E.g.,  $\bigcup A$  = the set of things that are in at least one member of  $A$ .

We'll need a new axiom for the new unions operation:

**Union axiom:** for any set,  $A$ , there exists a set,  $B$ , whose members are exactly the members of members of  $A$ .

As for generalized intersections, we don't need a new axiom: the intersection of a (nonempty) set is guaranteed to exist by the subset axioms.

(Annoying special case where we take the intersection of  $\emptyset$  — the original definition would yield a universal set. So we just set  $\bigcap \emptyset = \emptyset$ .)

#### F. Examples

Exercise 4, p. 26: If  $A \subseteq B$ , then  $\bigcup A \subseteq \bigcup B$ .

Exercise 6a, p. 26,  $\bigcup \emptyset A = A$

Exercise 10, p. 26, if  $a \in B$ , then  $\emptyset a \in \emptyset \emptyset \bigcup B$

This one is a little more involved, though straightforward:

We're given that  $a \in B$ . We must show  $\emptyset a \in \emptyset \emptyset \bigcup B$ . Now,

$$\begin{aligned}\emptyset a \in \emptyset \emptyset \bigcup B &\leftrightarrow \emptyset a \subseteq \emptyset \bigcup B \\ &\leftrightarrow \forall y (y \in \emptyset a \rightarrow y \in \emptyset \bigcup B) \\ &\leftrightarrow \forall y (y \subseteq a \rightarrow y \subseteq \bigcup B)\end{aligned}$$

So, take any subset,  $y$ , of  $a$ . We must show that  $y \subseteq \bigcup B$ . So, let  $z$  be any member of  $y$ . We must show that  $z$  is a member of  $\bigcup B$  — that is, we must show that  $z$  is a member of a member of  $B$ . Well, since  $y \subseteq a$  and  $z \in y$ ,  $z \in a$ . But  $a \in B$ . So  $z$  is a member of a member of  $B$ .

## G. Algebra of sets

One of the objects studied by mathematicians are Boolean algebras: (from Stanford encyclopedia article by J. Donald Monk, <http://plato.stanford.edu/entries/boolalg-math/>):

A Boolean algebra (BA) is a set  $A$  together with binary operations  $+$  and  $\cdot$  and a unary operation  $-$ , and elements  $0, 1$  of  $A$  such that the following laws hold: commutative and associative laws for addition and multiplication, distributive laws both for multiplication over addition and for addition over multiplication, and the following special laws:

$$\begin{aligned}x + (x \cdot y) &= x \\ x \cdot (x + y) &= x \\ x + (-x) &= 1 \\ x \cdot (-x) &= 0\end{aligned}$$

Keep in mind that  $+$ ,  $\cdot$ , and  $-$  are not the usual operations on numbers. They're any old operations on our set  $A$ . And if they satisfy these constraints then the set plus the operations is called a boolean algebra.

Let's choose any old set,  $A$ , and, where  $x, y$  are members of  $A$ , define  $+$ ,  $\cdot$ ,  $-$ ,  $0$ , and  $1$  this way:

$$\begin{aligned}
x+y &= x \cup y \\
x \cdot y &= x \cap y \\
-x &= A-x && \text{(on the right, “-” means set-theoretic complement.)} \\
0 &= \emptyset \\
1 &= A
\end{aligned}$$

Then we can show that we’ve got a Boolean algebra. (There are lots of other structures that satisfy the definition of a boolean algebra, e.g., the algebra of propositions, conjunction, disjunction, etc.) Enderton establishes the commutation, distribution and associativity laws at the end of chapter 2. I’ll look at just one simple one of those proofs, and then establish the special laws:

Commutativity of +: That means we must show that  $x \cup y = y \cup x$ . Given extensionality, we must show that something’s in one iff it’s in the other:

$$\forall z (z \in x \cup y \leftrightarrow z \in y \cup x)$$

That’s easy (write out definition.)

For the next couple, it will be useful to have these facts (they’re easy to prove):

$$\begin{aligned}
\text{if } A \subseteq B, \text{ then } & \text{i) } A \cup B = B \\
& \text{ii) } A \cap B = A
\end{aligned}$$

$$x + (x \cdot y) = x, \text{ i.e., } x \cup (x \cap y) = x$$

$$x \cap y \subseteq x, \text{ and so } x \cup (x \cap y) = x.$$

$$x \cdot (x + y) = x, \text{ i.e., } x \cap (x \cup y) = x$$

$$x \subseteq x \cup y, \text{ and so } x \cap (x \cup y) = x$$

$$x + (-x) = 1, \text{ i.e., } x \cup (A-x) = A \quad \text{(obvious)}$$

$$x \cdot (-x) = 0, \text{ i.e., } x \cap (A-x) = \emptyset \quad \text{(obvious)}$$

In any Boolean algebra, one can define a relation  $\leq$  thus:  $x \leq y$  iff  $x+y=y$ . In this context, this means  $x \leq y$  iff  $x \cup y = y$ , which is equivalent to  $x \subseteq y$ . (If  $x \subseteq y$  then  $x \cup y = y$  by the above result, clause i). Now suppose that  $x \cup y = y$ , and let  $z \in x$ .  $z$  is then in  $x \cup y$ , and so is in  $y$ .) So our  $\leq$  relation is  $\subseteq$ .

Enderton proves various results in this vicinity. Here’s one, which introduces some new notation:

$$A \cup \bigcap B = \bigcap \{A \cup X \mid X \in B\} \quad (\text{assuming } B \neq \emptyset)$$

First, the notation. The expression  $\{A \cup X \mid X \in B\}$  means the set of sets of the form  $A \cup X$ , where  $X \in B$ . That is, it's the set,  $D$ , such that  $a \in D \leftrightarrow$  for some  $X \in B$ ,  $a = A \cup X$ .

We need to show that this notation really does pick out a set (assuming that  $A$  and  $B$  are sets to begin with.) This is a bit tricky, and goes by quickly in the book, but it's worth getting straight on, because it's a good illustration of a main technique for showing things to be sets.

So we must show that  $D$  is a set. The way we show something to be a set is by finding some set we know exists, and then showing the thing to be a subset of it. So, what's  $D$  a subset of? That is, how can we fill in the blank in the following:

If  $a \in D$ , then  $a \in \underline{\hspace{2cm}}$

where we can easily show that the set mentioned in the blank exists? Well, we know that if  $a \in D$  then for some  $X \in B$ ,  $a = A \cup X$ . But that doesn't have quite the right form. We need a particular set  $a$  is in. But let's work with what we have, and try to massage it into the right form:

for some  $X \in B$ ,  $a = A \cup X$ , so  
 if  $b \in a$  then for some  $X \in B$ ,  $b \in A \cup X$ , so  
 if  $b \in a$  then for some  $X \in B$ ,  $b \in A \vee b \in X$ , so  
 if  $b \in a$  then  $b \in A \vee$  for some  $X \in B$ ,  $b \in X$ , so  
 if  $b \in a$  then  $b \in A \cup \bigcup B$ , so

$a \subseteq A \cup \bigcup B$ , so

$a \in \wp(A \cup \bigcup B)$

(This illustrates a general strategy for finding a set that  $a$  is in: finding a set of which  $a$  is a subset. For then  $a$  is a member of that set's power set.)

OK, now we know that  $D \subseteq \wp(A \cup \bigcup B)$ . So we can use a subset axiom to pick it out:

$$\exists D \forall x (x \in D \leftrightarrow [x \in \wp(A \cup \bigcup B) \ \& \ \exists X (X \in B \ \& \ x = A \cup X)])$$

(The argument we just gave was that  $\exists X (X \in B \ \& \ x = A \cup X)$  implies  $x \in \wp(A \cup \bigcup B)$ ; hence, the left conjunct is redundant, and we have our set  $D$  such that  $x \in D$  iff  $\exists X (X \in B \ \& \ x = A \cup X)$ .)

## H. The axiomatic approach

Now that we've got a concrete sense of how the axiomatic approach works, I want to clarify the axioms and talk about what's significant about the approach.

The axiomatic approach identifies clearly the pillars on which set theory rests. That has always been regarded as an advantage, from Euclid on. So long as the axioms are sound, anything that follows from them is sound as well, since we only infer things that follow from the axioms. Also, it helps us avoid errors that rest on natural but fallacious assumptions. Without the axiomatic approach, one just reasons along in ways that seem intuitive. But with the axiomatic approach, we're not allowed to smuggle in any assumptions about sets in doing proofs unless they're axioms, however intuitively correct those assumptions are. We're not allowed to use our knowledge of what the (nonlogical) expressions in the axioms *mean*; we're only allowed to use what the axioms strictly *say*.

In the contemporary setting, this notion of *following from the axioms* can be made very precise. It is only since Frege that a good mathematical theory of logic has been developed. We now know how to define the notion of a language, and the notion of logical consequence relative to that language. So we can lay out the *language of set theory*, and talk about logical consequences in that language.

That's particularly important when it comes to the subset axioms. We said that every way of filling in the blank in the subset axiom schema counts as an axiom. But what are acceptable ways of filling in the blank? A precise answer can be given only if we have a precise definition of what counts as a formula in the language of set theory.

What is that language? Well, this isn't a logic class, but all we need is the primitive two-place predicate  $\in$ . No names or function symbols. Then we define the wffs in the usual way.

It is this insistence on having a precisely define language that blocks Berry's paradox. For that paradox uses an expression to pick out a set ('definable in less than one line of type') that isn't an expression in the language of set theory. Nor can you come up with a set such as  $\{x: x \text{ is the favorite set of a tall person}\}$ .

There's another thing that's nice about doing the axiomatic approach with contemporary logic. It makes it possible to precisely investigate what is *not* a consequence of given axioms. E.g., one can prove that various axioms of set theory are independent of each other. That's one of the main bits of research in contemporary set theory.

So, the axiomatic approach+mathematical logic is an excellent approach to the discipline. But it's also important to realize that the fact that the language of set theory itself has been formalized does not mean that set theory is just an inquiry into language. This formal approach is not different in kind to just writing down a bunch of stuff we believe about sets and reasoning in intuitive ways: in each case, we're trying to articulate facts about sets. It's just that the first approach has turned out to be a good way to avoid error, and to identify and tease apart the

various assumptions.

### III. Relations and functions

#### A. Ordered pairs

##### 1. Idea of ordered pairs

Our pair sets are *unordered*;  $\{x,y\}=\{y,x\}$  given extensionality. But sometimes it's handy to have a set that "concerns" two things,  $x$  and  $y$ , in which it makes sense to speak of  $x$  being the *first* thing the set concerns, and  $y$  being the *second* thing the set concerns. Our notation will be this:

$$\langle x,y \rangle$$

This denotes the *ordered pair* of  $x$  and  $y$ ; we'll call  $x$  the "first coordinate" of  $\langle x,y \rangle$ , and  $y$  the "second coordinate" of  $\langle x,y \rangle$ . But we need to define  $\langle x,y \rangle$ , and the relation of being the first coordinate and the second coordinate. Our definition will be guided by the goal of making this turn out true:

$$\langle x,y \rangle = \langle u,v \rangle \text{ iff } x=u \text{ and } y=v.$$

That way, only two things matter to the identity of an ordered pair: i) what its coordinates are, and ii) what its members are. If this law holds, then we can "recover" the information of what the first and second coordinates is. For the law says that if you change either the first or the second coordinate, you change the ordered pair: if either  $x \neq u$  or  $y \neq v$ , then  $\langle x,y \rangle \neq \langle u,v \rangle$ .

##### 2. Definition of ordered pairs

OK, how to define the ordered pair? We can't define  $\langle x,y \rangle$  as  $\{x,y\}$ , because then  $\langle 2,1 \rangle = \langle 1,2 \rangle$ , in violation of the law. We can't define it as  $\{\{x\},y\}$ , because then  $\langle \{\emptyset\},\{\emptyset\} \rangle = \langle \emptyset,\{\{\emptyset\}\} \rangle$ .

Here's a definition that works. (It's not the only one; it's from Kuratowski.)

$$\langle x,y \rangle =_{\text{df}} \{\{x\},\{x,y\}\}$$

Note: this set is always guaranteed to exist. Just use pairing a few times.

##### 3. Law of the ordered pair

Second, we need to establish the law; i.e., we need to prove that:

$$\{\{x\},\{x,y\}\} = \{\{u\},\{u,v\}\} \text{ iff } x=u \text{ and } y=v$$

Right to left is trivial; left to right is straightforward but involves checking some special cases. So, suppose that  $\{\{x\},\{x,y\}\} = \{\{u\},\{u,v\}\}$ . Now, every member in the one set is in the other set. That means that:

$$\{x\}=\{u\} \text{ or } \{x\}=\{u,v\}$$

and

$$\{x,y\}=\{u\} \text{ or } \{x,y\}=\{u,v\}$$

From the first conjunct we know that  $x=u$ . From the second conjunct, we must argue that  $y=v$ . If the first disjunct is true, then  $x=y=u$ , so  $\{\{x\},\{x,y\}\} = \{\{x\}\} = \{\{u\},\{u,v\}\}$ , so  $\{u,v\}=\{x\}$ , so  $v=x$ , so  $v=y$ . Suppose, on the other hand, that the second disjunct is true — i.e.,  $\{x,y\}=\{u,v\}$ . And suppose that  $v \neq y$ . Then  $v=x$ . But  $x=u$ , so  $v=u$ , so  $\{x,y\}=\{v\}$ , so  $v=y$ .

#### 4. Coordinates

Given the law, we can give the following definitions:

the first coordinate of an ordered pair,  $z$ , is the object  $x$ , such that for some  $y$ ,  $z=\langle x,y \rangle$   
the second coordinate of an ordered pair,  $z$ , is the object  $y$ , such that for some  $x$ ,  $z=\langle x,y \rangle$

They're guaranteed to pick out unique objects by the law. For suppose, e.g., that there were two objects,  $x_1$  and  $x_2$ , as described by the first definition:

for some  $y$ ,  $z=\langle x_1,y \rangle$   
for some  $y$ ,  $z=\langle x_2,y \rangle$

call the first  $y$ ,  $a$ , and the second  $y$ ,  $b$ . We have  $\langle x_1,a \rangle$  and  $\langle x_2,b \rangle$ , but then  $x_1=x_2$  by the law. Similarly for the second definition.

#### B. Cartesian product

For any sets  $A$  and  $B$ , we define the Cartesian product  $A \times B$  thus:

$$A \times B = \{\langle a,b \rangle \mid a \in A \text{ and } b \in B\}$$

We have to be sure that this set exists. Any of its members has this form:  $\{\{a\},\{a,b\}\}$ . Both  $\{a\}$  and  $\{a,b\}$  are subsets of  $A \cup B$ , so members of  $\wp(A \cup B)$ . So  $\{\{a\},\{a,b\}\}$  is a subset of  $\wp(A \cup B)$ , so it's a member of  $\wp \wp(A \cup B)$ . So a subset axiom says that there exists a set  $Z$  such that:

$$x \in Z \leftrightarrow x \in \wp \wp(A \cup B) \ \& \ \exists a \exists b (a \in A \ \& \ b \in B \ \& \ x = \langle a,b \rangle)$$

## C. Relations

### 1. Relations defined

Mathematics usually deals with structures, in which what is important is how the elements of the structure are related to each other. For instance, the structure of the natural numbers:

0, 1, 2, ...

part of what is important is the *order* of these things: 0 is before 1, 1 is before 2, etc. This notion of being “before” is a relation --- it specifies how one number is vis a vis another. The relation of **before** is an *entity* that encodes the information about which things are before which other things.

Relations also apply to nonmathematical entities. E.g., being married is a relation that one person bears to another. The relation **marriage** is an entity that encodes the information of who is married to whom.

In set theory we can view relations as sets of ordered pairs. E.g., the marriage relation is:

$$\{ \langle x, y \rangle \mid x \text{ is married to } y \}$$

e.g., the less than relation on natural numbers is:

$$\{ \langle n, m \rangle \mid n, m \in \mathbb{N} \text{ and } n < m \}$$

(Since sometimes  $x$  can stand in a relation to  $y$  but not vice versa, it's important that we're using *ordered* pairs, not pair sets.)

So, we define a relation as a set of ordered pairs. Where  $R$  is a relation, we write  $xRy$  instead of  $\langle x, y \rangle \in R$ .

### 2. Domain, range, field

**Domain of  $R$ :**  $\{x: \exists y xRy\}$   
**Range of  $R$ :**  $\{y: \exists x xRy\}$   
**Field of  $R$ :**  $\text{dom}(R) \cup \text{ran}(R)$

Note: these sets are guaranteed to exist, for the objects from the domain, range, and field are all

in  $\cup\cup R$ .

### 3. n-tuples

We define ordered triples as pairs of ordered pairs and things:

$$\langle x_1, x_2, x_3 \rangle =_{df} \langle \langle x_1, x_2 \rangle, x_3 \rangle$$

similarly for ordered quadruples, and in general, an  $n+1$ -tuple is an ordered pair of an  $n$ -tuple and a thing. Define the one-tuple  $\langle x \rangle$  to be just  $x$ .

An  $n$ -place relation is a set of  $n$ -tuples.

### D. Functions

Remember the function-argument notation:  $f(x)=y$ .  $f$  is a function, or rule, that “takes in” an argument  $x$ , and spits back a value  $y$ . We want to develop set-theoretic definitions for this idea.

Relations correlate objects with other objects, which suggests defining a function as a relation. But not any relation. We use “ $f(x)$ ” as a name, thus assuming that there’s only one thing that  $f$  associates with  $x$ . Functions can look like this

$$\begin{array}{l} A \rightarrow X \\ B \rightarrow Y \\ C \nearrow \end{array}$$

But never like this:

$$\begin{array}{l} A \rightarrow X \\ B \rightarrow Y \\ \quad \searrow Z \end{array}$$

So we give the following definition:

**Function:** a relation,  $R$ , such that if  $xRy$  and  $xRz$  then  $y=z$

Whenever  $x \in \text{dom}(F)$  and  $F$  is a function, we name the unique  $y$  such that  $xFy$  thus:  $F(x)$ .

Note that functions can have ordered pairs in their domains. E.g.,  $+(\langle 2,4 \rangle)=6$ . These can be thought of as binary operations. Notation: write  $2+4$  instead of  $+(\langle 2,4 \rangle)$ . Similarly, functions whose domains are ordered triples are ternary operations.

## 1. Terminology for functions

**F is from A into B (“ $F:A \rightarrow B$ ”)**:  $F$  is a function,  $\text{dom}(F)=A$ ,  $\text{ran}(F) \subseteq B$

**F is from A onto B**:  $F$  is a function,  $\text{dom}(F)=A$ ,  $\text{ran}(F)=B$

**F is one-to-one (an injection)**: if  $uFx$  and  $vFx$  then  $u=v$ . (This and various subsequent definitions are well-defined even when  $F$  isn't a function. If  $F$  is not a function, we use the phrase “single-rooted” rather than “one-to-one”.)

**Inverse of F,  $F^{-1}$** :  $\{\langle u,v \rangle \mid vFu\}$  (Note that  $F^{-1}$  needn't be a function even if  $F$  is a function)

**Composition of F and G,  $F \circ G$** :  $\{\langle u,v \rangle \mid \exists x (uGx \ \& \ xFv)\}$

**Restriction of F to A,  $F \upharpoonright A$** :  $\{\langle u,v \rangle \mid uFv \ \& \ u \in A\}$

**Image of F under A,  $F[A]$** :  $\{v \mid \exists u \in A \ uFv\}$  (equivalently:  $\text{ran}(F \upharpoonright A)$ )

**Indexed unions and intersections**: where  $F$  is a function whose domain includes  $I$ , “ $\bigcup_{i \in I} F(i)$ ”

is defined as  $\bigcup \{F(i) \mid i \in I\}$ . Similarly for  $\bigcap_{i \in I} F(i)$ .

**B-pre-A ( ${}^A B$ )**:  $\{F \mid F:A \rightarrow B\}$

In the case of the last six, we need to justify that the sets exist, by finding sets of which they are subsets. That's easy;  $F^{-1} \subseteq \text{ran}(F) \times \text{dom}(F)$ , etc.

Example: Work through the example of the parabola (p. 45) on the board.

Example:  $F(x)=x^2$ ,  $G(x)=3x$ , then  $F \circ G(x)=9x^2$ ,  $G \circ F(x)=3x^2$ .

Example: relation  $M$  of (biological) mother-of:  $\{\langle u,v \rangle \mid u \text{ is } v\text{'s mother}\}$ .  $M$  is not a function, but  $M^{-1}$  is. (Mothers can have more than one child; each person has exactly one mother.)

## 2. Theorems about functions

Enderton proves a number of theorems here. I'll do some of them.

**Theorem 3E**:  $\text{dom } F^{-1} = \text{ran } F$ ,  $\text{ran } F^{-1} = \text{dom } F$ ; if  $F$  is a relation then  $(F^{-1})^{-1} = F$ .

**Theorem 3F**:  $F^{-1}$  is a function iff  $F$  is single-rooted. If  $F$  is a relation then  $F$  is a function iff  $F^{-1}$  is single-rooted

**Theorem 3G**: If  $F$  is a one-to-one function and  $x \in \text{dom } F$ , then  $F^{-1}(F(x))=x$ . If  $y \in \text{ran } F$  then

$$F(F^{-1}(y))=x$$

Proof: let  $x \in \text{dom } F$ . Then  $\langle x, F(x) \rangle \in F$ . So  $\langle F(x), x \rangle \in F^{-1}$ . Single rooted=one-to-one for functions, so  $F$  is single-rooted, so by 3F  $F^{-1}$  is a function, so  $F^{-1}(F(x))=x$ . On the other hand, suppose  $y \in \text{ran } F$ . By 3E,  $y \in \text{dom } F^{-1}$ . By 3F,  $F^{-1}$  is a function, so  $F^{-1}(y)$  is well-defined. So  $\langle y, F^{-1}(y) \rangle \in F^{-1}$ . By 3E,  $(F^{-1})^{-1}$  is  $F$ , so  $\langle F^{-1}(y), y \rangle \in F$ , so  $F(F^{-1}(y))=y$ .

**Theorem 3H:** where  $F$  and  $G$  are functions,  $F \circ G$  is a function with domain  $\{x \in \text{dom } G \mid G(x) \in \text{dom } F\}$ ; and for any  $x$  in this domain,  $(F \circ G)(x) = F(G(x))$

**Theorem 3I:** for any sets  $F$  and  $G$ ,  $(F \circ G)^{-1} = G^{-1} \circ F^{-1}$

**Theorem 3J** Let  $F: A \rightarrow B$  and  $A$  be nonempty.

- (a) iff  $F$  is one-one, there exists a function  $G: B \rightarrow A$  (a “left inverse”) such that  $G \circ F$  is  $I_A$  ( $I_A$  is the identity function on  $A$ , i.e.,  $\{\langle x, x \rangle \mid x \in A\}$ )
- (b) iff  $F$  maps  $A$  onto  $B$ , there exists a function  $H: B \rightarrow A$  (a “right inverse) such that  $F \circ H$  is  $I_B$

Proof: (<draw picture of all this, as on p. 49 of Enderton>)

- (a): Suppose  $F$  is one-one. By 3F,  $F^{-1}$  is a function. Given 3G, for any  $x \in A$ ,  $F^{-1}(F(x)) = x$ . But  $F^{-1}$  isn't the function  $G$  we want, because it may not be defined on all of  $B$ . However, we are given that  $A$  is nonempty. So there is some  $a \in A$ . So, there is some function  $G$  that gives us what we want, namely, the function that behaves like  $F^{-1}$  over  $\text{ran } F$ , but maps everything in  $B - \text{ran } F$  to  $a$ . (i.e.,  $G = F^{-1} \cup (B - \text{ran } F \times \{a\})$ ). This is what we want because it is a function (obviously) from  $B$  into  $A$ , and because for any  $x \in A$ ,  $G(F(x)) = F^{-1}(F(x)) = x$ .

On the other hand, suppose such a function  $G$  exists, and suppose that  $F(u)=y$  and  $F(v)=y$ . Since  $G \circ F = I_A$ ,  $G(F(u))=u$ , and  $G(F(v))=v$ ; and so  $G(y)=u$  and  $G(y)=v$ , and so  $u=v$ . So  $F$  is one-to-one.

- (b): Suppose that such a function  $H$  exists. Let  $b \in B$ . Then  $F(H(b))=b$ .  $b \in \text{ran } F$ .

On the other hand, suppose  $F$  maps  $A$  onto  $B$ . We must claim that there exists a function  $G$  as described. We need to have it that for each  $b \in B$ ,  $F(G(b))=b$ . Now, for every  $b \in B$ , for some  $a \in A$ ,  $F(a)=b$ . So all  $G$  has to do is map  $b$  to some such  $a$ . But there may be many such  $a$ 's. So we need to find a  $G$  that *chooses* some such  $a$  for each  $b$ .

But actually, we can't argue that there is such a  $G$  without a further axiom:

**Axiom of Choice (first form):** for any relation,  $R$ , there is a function,  $H \subseteq R$ , with the same domain as  $R$

Given this axiom, we argue as follows. Define  $R$  thus:  $\{ \langle b, a \rangle \mid F(a) = b \ \& \ b \in B \}$ . Since  $F$  is onto  $B$ , we know that  $\text{dom } F = B$ . By the axiom of choice, there's some function  $H$  that's a subset of  $R$ , and whose domain is  $B$ . This is our desired  $H$ . For since  $H$  is a subset of  $R$ , we know that if  $b \in H$  then  $F(H(b)) = b$ , and so by definition of  $R$ ,  $F(H(b)) = b$ .

Let's stop to think about why we needed the axiom of choice in the proof of (b), but not in the proof of (a). The proof of (a) contained reasoning that has this form:

There exists a  $\phi$   
Call any such  $\phi$  "a"  
Since a exists and is  $\phi$ , there exists a function  $G$  that is  $\psi$   
So, there exists a function  $G$  that is  $\psi$

(Remember that the construction of  $G$  depended on  $a$ .) Now, in this reasoning, we used the "temporary name" "a", as is customary in logic and mathematics, but really the inference is just this:

There exists a  $\phi$   
For every  $\phi$  there exists a  $\psi$   
So, there exists a  $\psi$

This pattern of reasoning is *logically* valid. No need for further axioms.

But the reasoning in (b) did not have that form. Nor does it have any other logically valid form. You really do need a special assumption about sets to construct a function  $H$ .

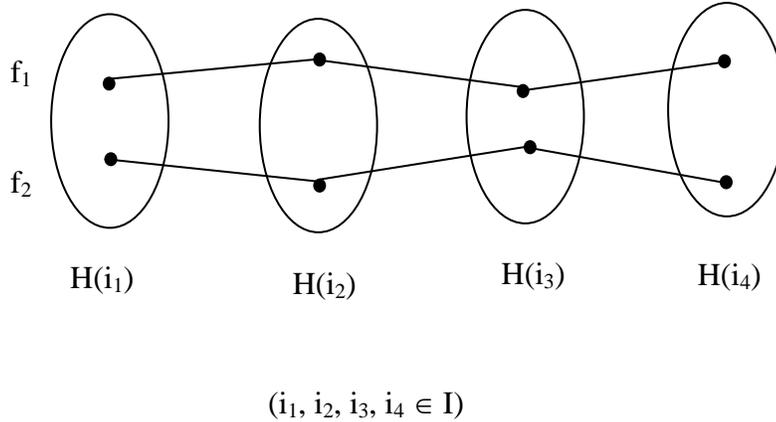
It's really easy to not notice that you're using the axiom of choice, so it's worth looking very carefully at this proof, to get an intuitive sense of when you need it and when you don't.

### 3. Infinite cartesian products

The cartesian product  $A \times B$  is the set of ordered pairs  $\langle a, b \rangle$  drawn from  $A$  and  $B$ , respectively. Similarly for the Cartesian product of three sets  $A \times B \times C$  is the set of ordered triples  $\langle a, b, c \rangle$  drawn from  $A$ ,  $B$  and  $C$ . Where  $I$  is any set, define an *I-tuple* as a function whose domain is  $I$ . This generalizes the notion of *n-tuples* in a natural way ( $I$  could be infinite.)

Now let's introduce the notion of an infinite Cartesian product. We need to introduce the sets of which we are taking the infinite cartesian product. We do this with a function,  $H$ , whose domain includes some set  $I$ . The various sets  $H(i)$ , for  $i \in I$ , are the sets of which we are taking the

Cartesian product; and the members of the cartesian product of the  $H(i)$ s are  $I$ -tuples whose “members” are drawn from the  $H(i)$ s – i.e., functions  $f$ , whose domains are  $I$ , which assign to each  $i \in I$  some member of  $H(i)$ . Here’s the picture:



The definition, then, is this:

The cartesian product  $\times_{i \in I} H(i) = \{f \mid f \text{ is a function with domain } I \text{ and } \forall i \in I, f(i) \in H(i)\}$

Now, suppose that for each  $i \in I$ ,  $H(i)$  is nonempty. Intuitively,  $\times_{i \in I} H(i)$  ought then to be nonempty.

Suppose  $I$  is some particular finite set we’re interested in. (Not that we have ‘finite’ defined yet.) Then, whether or not the  $H(i)$ s are finite, we can prove that  $\times_{i \in I} H(i)$  is nonempty. E.g., suppose that  $I = \{a, b\}$ , and that  $H(a)$  and  $H(b)$  are nonempty. Then, we can reason thus:

$\exists x \ x \in H(a)$   
 $\exists y \ y \in H(b)$   
 So,  $\exists z \ z = \{ \langle a, x \rangle, \langle b, y \rangle \}$   
 So,  $\times_{i \in I} H(i)$  is nonempty

But this reasoning isn’t available when  $I$  is infinite. In that case, we need a new form of the axiom of choice:

**Axiom of Choice II:** the cartesian product of nonempty sets is nonempty

You can prove this version of the axiom from the first version. We can define a relation thus:  $R = \{ \langle i, x \rangle \mid i \in I \ \& \ x \in H(i) \}$ . (It exists, for it is a subset of  $I \times \bigcup (\text{dom } H)$ .) The original axiom of

choice then gives us a function,  $f$ , with the same domain as  $R$ , that's a subset of  $R$ . Since each  $H(i)$  is nonempty,  $\text{dom } R = I$ , so  $\text{dom } f$  is  $I$ . And since  $f$  is a subset of  $R$ , for each  $i \in I$ ,  $f(i) \in H(i)$ .

### E. Equivalence relations

Do some pictures. Move between passing between the cells in a partition and a relation that determines the partition. E.g., the relation *being the same height as, being in the same state as* (over cities).

Definitions:

**B is a partition of A** iff  $B$  is a set of nonempty subsets of  $A$  that is *disjoint* and *exhaustive*, in that i) no two members of  $B$  overlap, and ii) every member of  $A$  is in some member of  $B$

**R is an equivalence relation on A** iff  $R$  is a binary relation on  $A$  that is reflexive on  $A$ , transitive and symmetric.

One can easily prove that partitions induce unique equivalence relations and equivalence relations induce unique partitions

Exercise 37: "Every partition induces an equivalence relation": where  $\Pi$  is a partition of set  $A$ , then the relation  $R_\Pi$  defined as below is an equivalence relation on  $A$ :

$$R_\Pi xy \text{ iff } \exists p \in \Pi, x \in p \text{ and } y \in p$$

Pf: Take any  $x \in A$ . Since  $\Pi$  is a partition, for some  $p \in \Pi$ ,  $x \in p$ . So  $R_\Pi xx$ .

Next take any  $x, y \in A$ . Suppose  $R_\Pi xy$ . Then for some  $p \in \Pi$ ,  $x \in p$  and  $y \in p$ . So, for some  $p \in \Pi$ ,  $y \in p$  and  $x \in p$ , so  $R_\Pi yx$ .

Next take any  $x, y, z \in A$  such that  $R_\Pi xy$  and  $R_\Pi yx$ . So  $x$  and  $y$  are both in some  $p \in \Pi$ , and  $y$  and  $z$  are in some  $p' \in \Pi$ . So  $y$  is in both  $p$  and  $p'$ ; but since no two members of  $\Pi$  overlap, that means that  $p=p'$ , and so  $z \in p$ , so  $R_\Pi xz$ .

Now we'll prove that equivalence relations induce partitions. Let  $R$  be any relation, and define:

$$[x]_R =_{\text{df}} \{t \mid xRt\}$$

( $[x]_R$  is called "the equivalence class of  $x$ " if  $R$  is an equivalence relation.)

**Lemma 3N** If  $R$  is an equivalence relation on  $A$ ; and  $x, y \in A$ ; then:  $[x]_R = [y]_R$  iff  $xRy$

Pf: Suppose  $[x]_R = [y]_R$ . Since  $x \in A$ ,  $xRx$ , so  $x \in [x]_R$ , so  $x \in [y]_R$ , so  $yRx$ , so  $xRy$  (symmetry).

Now suppose  $xRy$ , and take any  $t \in [x]_R$ . By symmetry,  $yRx$ ; since  $t \in [x]_R$ ,  $xRt$ ; by transitivity,  $yRt$ , so  $t \in [y]_R$ . Given symmetry of  $R$ , the same argument establishes that anything in  $[y]_R$  is also in  $[x]_R$ .

**Theorem 3P** If  $R$  is an equivalence relation on  $A$ , then the set  $\{[x]_R \mid x \in A\}$  of equivalence classes of members of  $A$  is a partition of  $A$ .

Pf: since  $R$  is reflexive on  $A$ , each  $[x]_R$  is nonempty. Clearly, each  $[x]_R \subseteq A$  since  $R$  is a relation on  $A$ . Exhaustive: let  $a \in A$ . by reflexivity,  $aRa$ , so  $a \in [a]_R$ . Disjoint: suppose  $a \in A$  is in both  $[x]_R$  and  $[y]_R$ , for  $x, y \in A$ . Thus,  $xRa$  and  $yRa$ . So,  $xRy$  (transitivity and symmetry), and so  $[x]_R = [y]_R$  by Lemma 3N.

Equivalence relations are useful for “factoring out irrelevant differences”. Example: suppose we don’t think that propositions that are necessarily equivalent are identical, but in some context we don’t care about differences between necessarily equivalence propositions (“hyperintensional differences”). In this context, we just want to treat all necessarily equivalent propositions as equal. So, instead of speaking of propositions themselves, we speak instead of equivalence classes of propositions under the relation of necessary equivalence. Example: we may wish to speak of the *directions* of lines. What would a direction be? An equivalence class of lines under the relation *being parallel to*.

One more theorem. The set of all equivalence classes of  $R$  on  $A$  is called the *quotient set*:  $A/R$ . The next theorem concerns the relationship between functions defined on  $A$  and functions defined on  $A/R$ . The basic idea is this. Suppose we have a function  $F: A \rightarrow A$ . It won’t in general be true that we can find a function  $G$  on  $A/R$  that “meshes” with  $F$  in this sense:

$$G([x]) = [F(x)]$$

(put up the picture on 59). For let  $R =$  same height as, and let  $F(x) =$  the father of  $X$ . There can’t be any such function  $G$ . For what will it assign to the set of all  $6'$  people? It must be a set of people of some one height. But different members of the set of  $6'$  people have fathers of different heights; these members will be mapped by  $F$  to people that are *not* in the same equivalence class, and so the equation above won’t hold for all values of  $x$ . The following theorem states the conditions under which there will be such a function  $G$ .

Definition:  $F$  and  $R$  are *compatible* iff for all  $x, y \in A$ , if  $xRy$  then  $F(x)RF(y)$

**Theorem 3Q** Let  $R$  be an equivalence relation on  $A$ , and  $F:A \rightarrow A$ . If  $F$  is compatible with  $R$ , then there exists a unique function  $G:A/R \rightarrow A/R$  such that:

$$(*) G([x]_R) = [F(x)]_R \text{ for all } x \in A$$

If  $F$  is not compatible with  $R$ , no such  $G$  exists.

Pf: Suppose  $F$  is *not* compatible with  $R$ . Then for some  $x, y \in A$ ,  $xRy$  but not  $F(x)RF(y)$ . Suppose such a  $G$  exists. Then  $G([x]) = [F(x)]$ , and  $G([y]) = [F(y)]$ . Since  $xRy$ ,  $[y] = [x]$  (Lemma 3N), so  $G([x]) = G([y])$  (since  $G$  is a function), and so  $[F(x)] = [F(y)]$ , and so by Lemma 3N,  $F(x)RF(y)$ .

Now suppose that  $F$  is compatible with  $R$ . We must now construct  $G$ . Let  $G = \{ \langle p, q \rangle \mid p, q \in A/R, \text{ and } \exists x \in p \exists y \in q: F(x) = y \}$ .

$G$  is a function: suppose  $pGq$  and  $pGr$ . Then for some  $x, x' \in p$  and some  $y \in q$  and some  $y' \in r$ ,  $F(x) = y$ ,  $F(x') = y'$ . Since  $x, x' \in p$ ,  $xRx'$ , and so by compatibility,  $yRy'$ , so  $q = r$  by Lemma 3N.

Dom  $G$  is  $A/R$ : every  $p \in A/R$  is nonempty, so let  $x \in p$ . Since  $F:A \rightarrow A$ ,  $x \in \text{dom } F$ , so there exists some  $y$  such that  $y = F(x)$ ; but then  $\langle p, [y] \rangle \in G$ .

$G$  obeys (\*): take any  $[x]$ .  $\langle [x], [F(x)] \rangle \in G$  by construction, so  $G([x]) = [F(x)]$ .

$G$  is unique: suppose  $G'$  obeys (\*), and consider any  $p \in A/R$ ; we will show that  $G(p) = G'(p)$ .  $p$  is nonempty, so consider any  $x \in p$ . By (\*),  $G([x]) = [F(x)] = G'([x])$ . But  $[x] = p$ .

## F. Orderings

One final concept that we can now formulate, now that we have ordered pairs at our disposal. We often want to talk of *orderings*. If you tell people to get in line, you are asking them to get in some order. You could ask them to stand in order of birthdate, except that you will need to figure some way to break ties in birthdates. Here's the definition:

**Linear order:** a binary relation,  $R$ , that is *transitive* and satisfies *trichotomy*: for any  $x, y$  in  $R$ 's field, exactly one of the following alternatives holds:  $x = y$ ,  $xRy$ ,  $yRx$

Note this theorem:

**Theorem 3R** If  $R$  is a linear order on  $A$ , then  $R$  is irreflexive and connected on  $A$

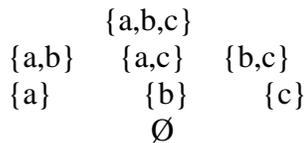
Irreflexive: for all  $x$ , it's not true that  $xRx$

Connected on  $A$ : for all  $x, y \in A$ , if  $x \neq y$  then either  $xRy$  or  $yRx$

(prove this)

What are linear orders like? You never have “ties”, you never have loops (given transitivity, they would lead to  $xRx$ ). So they “string” things out in a line. Note, though, that when we’re dealing with infinite sets, there don’t have to be first or last members in the line. E.g., the  $<$  relation over the set of all integers. Nor do there have to be “next” members --- e.g., the  $<$  relation on the set of all reals.

Linear orders are different from “partial orders”: transitive and irreflexive relations. The main difference is that partial orders don’t need to be connected: there can be objects  $x$  and  $y$  in their fields such that neither  $xRy$  nor  $yRx$ . E.g., take the relation of being a proper subset of:



Lower items on this diagram are subsets of higher ones, but things on the same “level” are such that neither is a proper subset of the other. Partial orders order their fields, but they don’t string them out into lines.

#### IV. Natural numbers

Enderton has a nice distinction between an axiomatic and a constructive approach to a new domain. On the axiomatic approach, you don’t *define* the core concepts of your new theory; you just state principles about them. That’s how we’ve approached set theory. On the constructive approach, we assume already given an old theory of some sort (in our case, set theory) and we then *define* or *construct* the concepts of the new theory in terms of the concepts of the old theory. That will be our approach to numbers.

The natural numbers are:

$$0, \quad 1, \quad 2, \quad \dots$$

One example of how we might try to construct them in terms of set theory is this:

$$\emptyset, \quad \{\emptyset\}, \quad \{\{\emptyset\}\}, \quad \dots \quad (\text{Zermelo})$$

But actually we'll do it a bit differently:

$$\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \dots \quad (\text{von Neumann})$$

That is, in each case, we take the set of all the previous numbers.

What do we want out of a “construction”? Arithmetic is the theory of the natural numbers. It has some basic *concepts*: number, successor, addition, multiplication, etc. And it has some *assumptions* about how those work, e.g.:

Where  $n$  and  $m$  are numbers,  $n+m=m+n$

Where  $n$ ,  $m$  and  $p$  are numbers,  $n \cdot (m+p) = n \cdot m + n \cdot p$

Here's what we want out of our construction: i) set-theoretic definitions of the core arithmetic concepts, and ii) proofs that the assumptions of arithmetic, when translated into the language of set-theory via the definitions in i), turn out to be theorems of set theory.

We'll now make the von Neumann construction precise. The “...” in the list of the von Neumann numbers above, as elsewhere so far, indicates a failure to have successfully defined anything. It's clear that for any finite  $n$ , we could work out the construction of the  $n^{\text{th}}$  natural number (given time). E.g., I can now define 0 as  $\emptyset$ , and 1 as  $\{\emptyset\}$ . But I still don't have a *general* definition of the phrase ‘set  $S$  is a natural number’.

#### A. Inductive sets

**Definition of successor:** the successor of  $a$ ,  $a^+$ , is  $a \cup \{a\}$

This is a definition that applies to *any* set; but when restricted to the natural numbers (we'll define those soon), this operation will be our set-theoretic version of the familiar successor operation on natural numbers.

**Definition of inductive:**  $A$  is inductive iff it contains  $\emptyset$  and is closed under successor --  
- i.e.,  $(\forall a \in A) a^+ \in A$

It should be intuitively clear that the Von Neuman numbers are then the following:

$$\emptyset, \emptyset^+, (\emptyset^+)^+, \text{ etc.}$$

For to get the successor of something, you just add it to everything in it – that way, each number is the set of all previous ones. But that's no proof.

We have no way of proving that any inductive sets exist. So we add a new axiom:

**Axiom of infinity: there exists at least one inductive set**

It's called the "axiom of infinity" because, intuitively, inductive sets must be infinite. But we don't have any independent definition of 'infinity' on the table at this stage, so just take the name informally. This is the first axiom we have given that makes an infinite existence assumption. And it is an intuitively correct axiom, given the iterative conception of set: just keep doing the successor operation forever, and gather the results into a set.

**Definition of number**  $n$  is a natural number iff  $n$  is a member of every inductive set

$\omega$ : the set of all natural numbers

We need to prove that  $\omega$  is a set. That's easy. The axiom of infinity says that there exists at least one inductive set,  $S$ ; and obviously  $\omega \subseteq S$ ; we can then use a subset axiom.

**Theorem 4B**  $\omega$  is inductive, and is a subset of every other inductive set

Pf: We must show that  $\omega$  contains  $\emptyset$  and is closed under successor.  $\emptyset$  is a natural number since by definition  $\emptyset$  is in any inductive set. Next, suppose  $n \in \omega$ . Now let  $S$  be any inductive set. Since  $n \in \omega$ ,  $n$  is in any inductive set, and so is in  $S$ ; since  $S$  is inductive,  $n^+ \in S$ . So  $n^+$  is in every inductive set, and so is in  $\omega$ . (As for being a subset of every other inductive set, that's obvious.)

**Induction principle for  $\omega$**  Any inductive subset of  $\omega = \omega$

This is an immediate consequence of 4B. But it's very useful in this form, because it allows us to prove facts about  $\omega$  "by induction". The basic idea is this. Suppose you are interested in showing that all natural numbers have a certain feature,  $\phi$ . Just form the set,  $B$ , of all natural numbers with  $\phi$ :

$$B = \{n \in \omega \mid \phi(n)\}$$

Then show that  $B$  is inductive. Conclude that  $B = \omega$ , by the induction principle. So all members of  $\omega$  are  $\phi$ . We'll look at interesting proofs by induction below (4C in the book gives a simple example.)

This version of proving things by induction corresponds to the usual method of proving things by

induction in mathematics, where you establish i) that 0 has some property, P, in which we're interested, and ii) that for any natural number n, if n has P then so does n+1; and then conclude that all natural numbers have P. Establishing that the desired set B of numbers with P is inductive involves exactly establishing i) and ii), for an inductive set must contain  $\emptyset$  (zero), and must be closed under successor.

Remember that the idea of constructing arithmetic inside set theory is to show that assumptions we make about numbers are provable when those assumptions are translated into set-theory. Showing that the induction principle holds is thus crucial, because proofs by induction are essential in arithmetic. In fact, the ability to prove things by induction might be regarded as definitive of the natural numbers. What's distinctive about the natural numbers is that they have this structure:

first element, second element, ...

There's a first element, and then each element is the *next* element after a previous element. They do *not* have any of the following structures:

·  
...  
·

←-----→ (real numbers)

... -3, -2, -1, 0, 1, 2, 3, ...

(0, 1, ...,) (a<sub>1</sub>, a<sub>2</sub>, ...) ...

You can't do induction on the first structure, i.e., an amorphous point soup, since there's no such thing as "the next" element. You can't do induction on the real numbers because even though there's an order, there's no such thing as "the next" real number. You can't do induction on the third structure, i.e., positive and negative integers, since even if 0 had a feature and successors inherited the feature then the negative numbers still might not have the feature. (of course, there are tricks we could use to get around this; the point is just that you can't do induction in the most straightforward way.) And you can't do induction in the final structure, even though there's a first element, and even though each thing has a successor. What fails here is that in addition to all those elements that are a finite number of jumps away from the first element (0), there are some extra elements.

## B. Recursion on $\omega$

One of the other main assumptions we make about the natural numbers is the legitimacy of recursive definitions. E.g.:

$$\begin{aligned} 0! &= 1 \\ (n+1)! &= n! \cdot (n+1) \end{aligned}$$

The idea is that you specify a function,  $h$ , by saying what  $h$  maps 0 to, and then you specify what  $h$  maps  $n+1$  to in terms of what it maps  $n$  to.

More carefully, we specify i) an *object* that is  $h(0)$ , and ii) we specify a *function*  $F$ , such that  $h(n+1)=F(h(n))$ . And in the general case,  $h$ 's range needn't be  $\omega$ ;  $h$  could be a function from  $\omega$  into any set  $A$ .

OK. For this kind of thing to count as an adequate definition of function  $h$ , it must be that specifying  $h(0)$ , and specifying the function  $F$  that gives  $h(n+1)$  in terms of  $h(n)$ , specifies a unique function  $h$  --, i.e., there is at least one function, and there is no more than one function.

Intuitively, if the naturals didn't have the structure they do, then recursive definitions wouldn't work. (e.g., specifying what a function does to 0 and to  $n+1$  won't tell us what it does on the negative integers.)

So, we need to show that recursive definitions work.

**Recursion theorem on  $\omega$**  Let  $A$  be a set,  $a \in A$ , and  $F:A \rightarrow A$ . Then there exists a unique function  $h:\omega \rightarrow A$  such that:

$$\begin{aligned} h(0) &= a \\ (\forall n \in \omega) \quad h(n^+) &= F(h(n)) \end{aligned}$$

Pf. (this is long). Call function  $v$  *acceptable* iff  $\text{dom } v \subseteq \omega$ ,  $\text{ran } v \subseteq A$ , and  
 i) if  $0 \in \text{dom } v$  then  $v(0)=a$   
 ii) if  $n^+ \in \text{dom } v$  (where  $n \in \omega$ ) then also  $n \in \text{dom } v$  and  $v(n^+) = F(v(n))$

Basically, the acceptable functions behave like  $h$  is supposed to behave on at least an initial segment of  $\omega$ . We now set  $h = \bigcup \{v \mid v \text{ is an acceptable function}\}$ . So:

$$(*) \quad \langle n, y \rangle \in h \text{ iff } v(n)=y, \text{ for some acceptable } v$$

We must now prove that  $h$  is our desired function:

1.  $h$  is a function
2.  $h$  is acceptable
3.  $\text{dom } h$  is  $\omega$
4.  $h$  is unique

Part 1: we need to show that for every  $n \in \omega$ , there's no more than one  $y$  such that  $\langle n, y \rangle \in h$ . We'll do this by induction, by showing that the following set is inductive:

$$S = \{n \mid \text{there's no more than one } y \text{ such that } \langle n, y \rangle \in h\}$$

Base: show that  $0 \in S$ . Suppose  $\langle 0, y \rangle \in h$  and  $\langle 0, z \rangle \in h$ . Then  $v(0)=y$  and  $v'(0)=z$ , for acceptable functions  $v$  and  $v'$ ; but then  $y=a$  and  $z=a$ , so  $y=z$ .

Induction: suppose  $n \in S$ , show  $n^+ \in S$ . Let  $n \in S$ . To show:  $n^+ \in S$ , that is, that there's at most one  $y$  such that  $\langle n^+, y \rangle \in h$ . Let  $\langle n^+, y \rangle \in h$  and  $\langle n^+, z \rangle \in h$ . Then for some acceptable  $v, v', v(n^+)=y$  and  $v'(n^+)=z$ . By definition of acceptability,  $v(n)$  and  $v'(n)$  are defined, and (%):  $y=F(v(n))$  and  $z=F(v'(n))$ . Since  $v$  and  $v'$  are acceptable,  $\langle n, v(n) \rangle \in h$  and  $\langle n, v'(n) \rangle \in h$ . But since  $n \in S$ , there's no more than one  $x$  such that  $\langle n, x \rangle \in h$ . So  $v(n)=v'(n)$ , and hence by (%),  $y=z$ .

So, by induction,  $S=\omega$ .

Part 2: show that  $h$  itself is acceptable. First, suppose that  $0 \in \text{dom } h$ . Then  $h(0)=b$ , for some  $b$ . So for some acceptable  $v, v(0)=b$ . But then  $b=a$ . So  $\langle 0, a \rangle \in h$ ; and, since we showed in part 1 that  $h$  is a function, we can write:  $h(0)=a$ .

Second, let  $n^+ \in \text{dom } h$  (where  $n \in \omega$ ); we must show that  $n \in \text{dom } h$  and  $h(n^+) = F(h(n))$ . Since  $n^+ \in \text{dom } h$ , then  $h(n^+)=b$ , for some  $b$ . So for some acceptable  $v, v(n^+)=b$ , and so, by definition of acceptability,  $n \in \text{dom } v$  and  $v(n^+)=F(v(n))$ . But then, by construction of  $h, \langle n, F(v(n)) \rangle \in h$ . So  $n^+ \in \text{dom } h$ , and (given part 1)  $h(n^+)=F(v(n))$ ; similarly,  $h(n)=v(n)$ , so  $h(n^+)=F(h(n))$ .

Part 3: show that  $\text{dom } h = \omega$ . Do this by showing that  $\text{dom } h$  is inductive.

Base: to show:  $0 \in \text{dom } h$ . The set  $\{\langle 0, a \rangle\}$  can be seen to be acceptable, so  $0 \in \text{dom } h$ . (Why is  $\{\langle 0, a \rangle\}$  acceptable? It's obvious that it satisfies the first condition of acceptability; to show that it satisfies the second condition, we need to argue that  $0$  is not the successor of any natural number. That is, for no number,  $n$ , is  $\emptyset = n \cup \{n\}$ . This holds because  $n \in n \cup \{n\}$ , whereas  $n \notin \emptyset$ .)

Induction: (I do this a bit different from how Enderton does it; also I fill in a few steps.) Suppose that  $n \in \text{dom } h$ . Thus, for some acceptable  $v, n \in \text{dom } v$ . Now, either  $n^+ \in \text{dom } v$  or it isn't. If it is then  $n^+ \in \text{dom } h$ . If it isn't, then we can construct another acceptable  $v'$  thus:  $v' = v \cup \{\langle n^+, F(v(n)) \rangle\}$ . To show it's acceptable: first, since  $v$  is acceptable,  $v'$ 's domain  $\subseteq \omega$  and its range is  $\subseteq A$ , so obviously the same holds for  $v'$ .  $v'$  is obviously a function since  $v$  was a function and  $n^+ \notin \text{dom } v$ . Since  $n^+ \neq 0$ ,  $v'$  assigns to  $0$  whatever  $v$  did, and so assigns to it a if it assigns it anything. Finally, consider any  $m^+ \in \text{dom}(v')$ . We must show that  $v'(m^+)=F(v'(m))$ . There are two cases:

If  $m^+ = n^+$  then  $v'(m^+)=F(v(n))$ , which (since  $^+$  on  $\omega$  is one-to-one – see below) means that

$v'(m^+) = F(v(m))$ ; but since no number is its own successor (see below),  $m \neq m^+$ , so  $v(m) = v'(m)$ , so  $v'(m^+) = F(v'(m))$ .

If on the other hand  $m^+ \neq n^+$  then  $m^+ \in \text{dom } v$ , so  $v(m^+) = F(v(m))$ . Now,  $m \neq n^+$  since  $m \in \text{dom } v$ , so  $v(m) = v'(m)$ . Also, since  $m^+ \neq n^+$ ,  $v'(m^+) = v(m^+)$ . So  $v'(m^+) = F(v'(m))$ .

(Lacuna: no number is its own successor. Let's prove this by induction. Obviously  $0 \neq 0^+$  since  $0 \in 0^+$ . Next, suppose that  $n \neq n^+$ ; then, since successor is one-to-one, it follows that  $n^+ \neq n^{++}$ .)

So,  $v'$  is acceptable. And since  $n^+ \in \text{dom } v'$ ,  $n^+ \in \text{dom } h$  after all.

Part 4:  $h$  is unique. Let  $h_1$  and  $h_2$  both satisfy the theorem. Since the domain of each is  $\omega$ , to show them identical it suffices to show that  $h_1(n) = h_2(n)$ , for all  $n \in \omega$ . We do this by induction. Clearly,  $h_1(0) = a = h_2(0)$  since  $h_1$  and  $h_2$  satisfy the theorem. Next suppose that  $h_1(n) = h_2(n)$ , and consider  $n^+$ . Since  $h_1$  and  $h_2$  satisfy the theorem,  $h_1(n^+) = F(h_1(n))$ , and  $h_2(n^+) = F(h_2(n))$ . Hence,  $h_1(n^+) = h_2(n^+)$ .

### C. A look at the big picture

Time for a quick look at the big picture. We've now proven that, from the axioms of set theory so far, we can construct the set,  $\omega$ , and do proofs about  $\omega$  by induction, and define functions on  $\omega$  by recursion. Where did this ability come from in our axioms?

Of course, a big part of it came from the Axiom of infinity. However, the axiom of infinity did not on its own guarantee a *minimal* inductive set, for that we needed to use other things, e.g., subset axioms. And it's the fact that  $\omega$  is a minimal inductive set that enables inductive proof and definition by recursion. The axiom of infinity only says that there exists at least one inductive set – i.e., there's at least one way of “continuing forever” with the operation of successors. So the work is divided between the axiom of infinity and the rest of the axioms.

Notice also that the axiom of infinity easily motivated by the iterative conception of set. Remember that we constructed the iterative hierarchy by starting with  $V_0 = \emptyset$  and then letting  $V_{i+1} = \wp V_i$ . Now,  $\emptyset \in V_1$ . And since  $a^+ = a \cup \{a\}$ , that means (though I can't prove it yet) that if  $a \in V_i$ , then  $a^+ \in V_{i+1}$ . So the successors are all showing up in the hierarchy. But we also said, in our intuitive description of the hierarchy, that whenever it seemed like we were done, we should take the union of all the levels we've constructed so far, let that be a new level, and then begin again. So there ought to be a level in the hierarchy containing  $\emptyset, \emptyset^+, \emptyset^{++}, \dots$

#### D. An adequacy condition for our construction of $\omega$

We've developed the natural numbers in set theory. Next we'll show that the natural numbers, as we've developed them, pass one adequacy test; roughly, that they obey the Peano axioms.

##### 1. Peano systems

Here are the "Peano axioms", commonly taken to be a solid basis for the structure of natural numbers:

1. 0 is a natural number.
2. Every natural number has one successor, which is also a natural number.
3. No natural number has 0 as a successor.
4. Different natural numbers have different successors.
5. If 0 has some property, and if when a number has that property then its successor also does, then all natural numbers have that property.

We can think of these axioms as saying what it takes for a given set of things ("numbers"), an operation ("successor"), and a given element ("zero"), to have the structure we demand of the natural numbers. Within set theory, this may be put thus:

A *Peano system* is an ordered triple  $\langle N, S, e \rangle$ , where  $N$  is a set,  $S: N \rightarrow N$ ,  $e \in N$ , and:

- (i)  $e \notin \text{ran } S$
- (ii)  $S$  is one-to-one
- (iii) any subset  $A$  of  $N$  that contains  $e$  and is closed under  $S$  equals  $N$  itself

(verify that this definition does indeed capture the axioms. Note that 2 is captured by the fact that  $S: N \rightarrow N$ .)

##### 2. $\omega$ is a Peano system

Let  $\sigma$  be the restriction of  $+$  to the  $\omega$ :  $\sigma = \{ \langle n, n^+ \rangle \mid n \in \omega \}$ . We want to show that  $\langle \omega, \sigma, 0 \rangle$  is a Peano system.

We'll need a bit of apparatus.

**Definition:** a *transitive set* is a set,  $A$ , such that if  $a \in b \in A$ , then  $a \in A$ .

That is, members of members of a transitive set are themselves members of the set. Transitive sets are thus “cumulative”. Here are some equivalent ways of defining transitive sets:

$$\bigcup_{A \subseteq A} A$$

$$a \in A \rightarrow a \subseteq A$$

$$A \subseteq \bigcup A$$

**Theorem 4E** If  $a$  is transitive then  $\bigcup(a^+) = a$  (for any set  $a$ )

$$\begin{aligned} \text{Pf: } \bigcup(a^+) &= \bigcup(a \cup \{a\}) && \text{(def of } ^+ \text{)} \\ &= \bigcup a \cup \bigcup \{a\} && \text{(exercise 21 of chapter 2)} \\ &= \bigcup a \cup a \\ &= a && \text{(since } \bigcup a \subseteq a \text{ if } a \text{ is transitive)} \end{aligned}$$

Before we asserted, informally, that each natural number is the set of all the earlier ones. This was an informal statement because a) we didn’t prove it, and – more immediately – b) we didn’t even define what “earlier” means. We’ll now make a start at making this statement precise and proving it, by showing that each natural number is transitive:

**Theorem 4F** Every natural number is a transitive set

Pf. Induction. 0 is vacuously transitive. Now suppose that  $n$  is transitive, and let’s show  $n^+$  to be transitive by showing that  $\bigcup n^+ \subseteq n^+$ . Since  $n$  is transitive, by 4E,  $\bigcup n^+ = n$ ; but since  $n^+ = n \cup \{n\}$ ,  $n \subseteq n^+$ ; hence,  $\bigcup n^+ \subseteq n^+$ .

We’ll now get even closer to “each natural number is the set of all earlier ones”, by proving that  $\omega$  is transitive. What this means is that each member of  $\omega$  (= natural number) is itself a member of  $\omega$  (= natural number). i.e.: each natural number is itself a set of natural numbers. (Before showing that each natural number is the set of all *smaller* natural numbers, we still need a definition of ‘smaller’. That will come soon.)

**Theorem 4G**  $\omega$  is a transitive set

Pf. By induction. We must show that every member of a member of  $\omega$  is itself a member of  $\omega$ . So, let  $T = \{n \mid n \in \omega \text{ and every } a \in n \text{ is } \in \omega\}$ . We must show that  $T$  is inductive.

Vacuously,  $0 \in T$ . Now suppose that  $n \in T$ ; we must show that  $n^+ \in T$ , i.e., that each  $a \in n^+$  is in  $\omega$ . Since  $n^+ = n \cup \{n\}$ , either  $a \in n$  or  $a = n$ . If the former then by the inductive hypothesis,  $a \in \omega$ ; and if the latter than  $a \in \omega$ . either way,  $a \in \omega$ .

OK, now back to the stated goal of this section:

**Theorem 4D**  $\langle \omega, \sigma, 0 \rangle$  is a Peano system

Obviously,  $0 \in \omega$ , and  $\sigma: \omega \rightarrow \omega$ . (The successor of a member of  $\omega$  is itself in  $\omega$  because  $\omega$  is inductive – theorem 4B.)  $0 \notin \text{ran } \sigma$  -- i.e.,  $0$  isn't the successor of anything – because  $0$  has no members. The induction principle for  $\omega$  immediately implies that any subset of  $\omega$  containing  $0$  and closed under  $\sigma$  is all of  $\omega$ . What remains is that  $\sigma$  is one-to-one.

Suppose that, for  $n, m \in \omega$ ,  $n^+ = m^+$ . Then  $\bigcup n^+ = \bigcup m^+$ ; but since  $n^+, m^+ \in \omega$ , they're transitive sets (4F), so  $n = m$  by 4E.

(Note: we've also discharged the obligation in the proof of the recursion theorem to show that successor (on  $\omega$ ) is one-to-one.)

### 3. $\omega$ is isomorphic to all Peano systems

Now we want to show that, in a sense,  $\omega$  is the unique Peano system. Now, all a structure has to do to count as a Peano system is have a first element, a one-one successor function, and have “nothing more than you can get to by successors”. So, e.g., since  $0, 1, 2, \dots$  is a Peano system, so is  $0, 2, 4, \dots$ , as is  $0, 1, 3, 2, 4, 5, 6, \dots$ . Still, these other systems “look like”  $\omega$ , in that they're isomorphic to  $\omega$ : there's a one-one function between them and  $\omega$  that preserves the successor function and the zero element (draw the picture). Here's a theorem that shows that *any* Peano system is isomorphic to  $\langle \omega, \sigma, 0 \rangle$ :

**Theorem 4H** Any Peano system  $\langle N, S, e \rangle$  is isomorphic to  $\langle \omega, \sigma, 0 \rangle$  in that there exists a function  $h$  mapping  $\omega$  one-one onto  $N$  such that:

$$h(\sigma(n)) = S(h(n)) \qquad h(0) = e$$

Pf: Since  $S: N \rightarrow N$ , the recursion theorem guarantees that a function,  $h$ , obeying these equations exists. We must show that  $h$  is one-one and onto  $N$ .

To show the latter, we're going to use the Peano induction condition on Peano systems, to show that  $\text{ran } h = \mathbb{N}$ . We know by the recursion theorem that  $\text{ran } h \subseteq \mathbb{N}$ ; and the second equation tells us that  $e \in \text{ran } h$ . The Peano induction condition on Peano systems then tells us that if  $\text{ran } h$  is closed under  $S$  then  $\text{ran } h = \mathbb{N}$ . So, suppose  $a \in \text{ran } h$ . Then for some  $n \in \omega$ ,  $h(n) = a$ . The first equation above then tells us that  $h(\sigma(n)) = S(a)$ , so  $S(a) \in \text{ran } h$ .

To show the former, we're going to use induction. Let  $T = \{n \in \omega \mid \text{for all } m \in \omega, \text{ if } h(n) = h(m) \text{ then } n = m\}$ . First we'll show that  $0 \in T$ : suppose that  $h(0) = h(m)$ . By the first equation,  $h(0) = e$ , so  $h(m) = e$ . We must show that  $m = 0$ . Suppose otherwise. Then, by 4C,  $m = k^+$  for some  $k \in \omega$ . So then, by the first equation,  $h(k^+) = S(h(k))$ . But by the definition of Peano systems,  $e \notin \text{ran } S$ , so  $h(k^+) \neq e$ . Contradicts the bold line (given  $m = k^+$ ).

Now the inductive step. Suppose  $n \in T$ . We must show that  $n^+ \in T$ . So suppose for some  $m \in \omega$ ,  $h(n^+) = h(m)$ ; we must show  $n^+ = m$ . By the first equation,  $h(n^+) = S(h(n))$ . Since  $h(m) = h(n^+)$ ,  $h(m) = S(h(n))$ , so  $h(m) \neq e$  ( $e$  isn't the  $S$  of anything in a Peano system.) That means that  $m \neq 0$  (since  $h(0) = e$  and  $h$  is a function.) So  $m = k^+$  for some  $k \in \omega$ . Hence  $h(m) = S(h(k))$ . But now we have:  $h(m) = S(h(n))$ , and  $h(m) = S(h(k))$ ; so,  $h(n) = h(k)$  (since  $S$  is one-to-one); but since  $n \in T$ ,  $n = k$ ; so  $n^+ = k^+$ , so  $n^+ = m$ .

## E. Arithmetic

### 1. Addition and multiplication defined

Our next goal is to define addition and multiplication on  $\omega$  -- i.e., define functions on  $\omega$  that have the properties we normally associate with addition and multiplication.

We're going to do it recursively. The usual recursive definitions of addition in terms of successor and multiplication in terms of addition are these:

- (A1)  $m + 0 = m$
- (A2)  $m + n^+ = (m + n)^+$
- (M1)  $m \cdot 0 = 0$
- (M2)  $m \cdot n^+ = m \cdot n + m$

What we want to do now is to turn these recursive definitions into definitions that we can prove to be successful given the recursion theorem. These definitions are of two-place functions, whereas the recursion theorem only tells us that one-place functions defined on  $\omega$  exist. So here's the trick:

Take any  $m \in \omega$ . The recursion theorem guarantees a function,  $A_m$ , which we can think of as the “add  $m$ ” function, obeying these equations:

$$\begin{aligned} A_m(0) &= m \\ A_m(n^+) &= A_m(n) + m \end{aligned}$$

We can then define the more familiar addition function,  $+$ , as  $\{\langle \langle m, n \rangle, k \rangle \mid A_m(n) = k\}$ . Since the recursion theorem guarantees a unique function  $A_m$  for each  $m$ ,  $+$  is a function.

We call  $+$  a “binary operation” on  $\omega$ , meaning that it’s a function from  $\omega \times \omega$  into  $\omega$ . And we write  $k = m + n$  instead of  $k = +(\langle m, n \rangle)$ .

Given the construction of  $+$ , it’s immediately obvious that A1 and A2 hold.

We can use the same trick to define multiplication. The recursion theorem guarantees that for each  $m \in \omega$ , there exists a unique function  $M_m$  (“multiply by  $m$ ”) such that:

$$\begin{aligned} M_m(0) &= 0 \\ M_m(n^+) &= M_m(n) + m \end{aligned}$$

And we can then define the binary operation  $\cdot$  as  $\{\langle \langle m, n \rangle, p \rangle \mid p = M_m(n)\}$ . It follows immediately from this construction that M1 and M2 hold.

## 2. Properties of addition and multiplication verified

We now will verify that various properties that addition and multiplication have are had by our constructed surrogates.

**Theorem 4K** For any natural numbers:

- Addition is associative:  $m + (n + p) = (m + n) + p$
- Addition is commutative:  $m + n = n + m$
- Multiplication distributes over addition:  $m \cdot (n + p) = m \cdot n + m \cdot p$
- Multiplication is associative:  $m \cdot (n \cdot p) = (m \cdot n) \cdot p$
- Multiplication is commutative:  $m \cdot n = n \cdot m$

These are all proved by induction. Since more than one numbers is involved in each case, sometimes more than one induction is needed. E.g., let’s take commutativity:

*Proof that  $+$  is commutative:*

First we need to prove an initial fact:  $0+n=n$ , for all  $n$ . (Note that this is *not* the same as A1. A1 tells us that adding 0 on the right to  $n$  gives us  $n$ . We need to show that adding it on the left to  $n$  gives us  $n$ .) We use induction. We know that  $0+0=0$  by A1. Now suppose that  $0+n=n$ , and show that  $0+n^+=n^+$ . By A2,  $0+n^+=(0+n)^+$ , and so by the inductive hypothesis,  $0+n^+=n^+$ .

The second preliminary fact is that  $m^++n=(m+n)^+$ . (Again, this is not the same as A2.) Induction: First, we must show that  $m^++0=(m+0)^+$ . This follows from A1. Second, we assume that  $m^++n=(m+n)^+$ , and show that  $m^++n^+=(m+n^+)^+$ . By A2,  $m+n^+=(m+n)^+$ , the latter of which  $=m^++n$  by the inductive hypothesis. So  $m+n^+=m^++n$ , and hence  $(m+n^+)^+=(m^++n)^+$ . But by A2,  $(m^++n)^+=m^++n^+$ .

OK, here is the overall inductive proof. Let  $T=\{m \in \omega \mid \text{for all } n \in \omega, m+n=n+m\}$ . We must show that  $T$  is inductive.

First, show that  $0 \in T$ . Take any  $n$ . By the first preliminary fact,  $0+n=n$ . By A1,  $n+0=n$ . So  $0+n=n+0$  – hence,  $0 \in T$ .

Next, suppose  $m \in T$ . We must then show that  $m^+ \in T$ , i.e., that for any  $n \in \omega$ ,  $m^++n=n+m^+$ . By A2,  $n+m^+=(n+m)^+$ . Since  $m \in T$ ,  $n+m=m+n$ ; so we know that  $n+m^+=(m+n)^+$ . By the second preliminary fact,  $(m+n)^+=m^++n$ . QED.

### 3. Ordering on $\omega$

In addition to the operations of  $+$  and  $\cdot$ , there is an ordering the natural numbers:  $<$ . We have an easy definition of this for  $\omega$ :

$$\begin{aligned} n < m &\text{ iff } n \in m \\ n \leq m &\text{ iff } n \in m \text{ or } n = m \end{aligned}$$

Note that  $p \in k^+$  iff  $p \in k$  or  $p = k$  (since  $k^+ = k \cup \{k\}$ ). Call this (\*).

Finally, we can say that a natural number is the set of all smaller natural numbers, since we've finally defined 'smaller'. We showed earlier that  $\omega$  is a transitive set, and hence that each member of a natural number is a natural number. By the new definition, each member of a natural number,  $n$ , is a smaller natural number. Moreover, every natural number smaller than  $n$  is now by definition a member of  $n$ . So  $n$  is the set of all smaller natural numbers.

If  $\in$  is to play the role of  $<$ , it must be a linear order on  $\omega$ . Thus, we must show that

$$\in_\omega = \{ \langle m, n \rangle \mid m, n \in \omega \text{ and } m \in n \}$$

is transitive and satisfies trichotomy on  $\omega$ .

Transitivity: suppose  $m \in n$  and  $n \in p$ . Then, since each natural number is a transitive set,  $m \in p$ .

Trichotomy is harder.

**Lemma 4L** (a) for any  $m, n \in \omega$ ,  $m \in n$  iff  $m^+ \in n^+$   
 (b) no natural number is a member of itself

Pf: (a), right to left. Suppose  $m^+ \in n^+$ . So  $m^+ \in n$  or  $m^+ = n$  (by (\*)). If the latter then since  $m \in m^+$ ,  $m \in n$ . If the former then since  $m \in m^+$ , by transitivity of  $n$ ,  $m \in n$ . So  $m \in n$ .

(a), left to right. Induction on  $n$ . Let  $T = \{ n \mid \text{for any } m, \text{ if } m \in n \text{ then } m^+ \in n^+ \}$ . It's vacuously true that  $0 \in T$ , since no  $m \in 0$ . Now suppose  $n \in T$ , and show that  $n^+ \in T$ . Consider any  $m \in n^+$ . We must show  $m^+ \in n^{++}$ . Since  $m \in n^+$ , by (\*),  $m \in n$  or  $m = n$ . If the latter then since  $n^+ \in n^{++}$ ,  $m^+ \in n^{++}$ . If the former, then by the inductive hypothesis,  $m^+ \in n^+$ ; but since  $n^+ \in n^{++}$ , by transitivity of  $n^{++}$ ,  $m^+ \in n^{++}$ . So  $T$  is inductive

(b): induction.  $0 \notin 0$ . Next suppose that  $n \notin n$ ; by (a),  $n^+ \notin n^+$ .

**Trichotomy for  $\omega$ :** for any natural numbers  $m, n$ , exactly one of the following holds:

$$m \in n \quad n = m \quad n \in m$$

Pf: First, at most one can hold. Suppose that  $m \in n$  and  $n = m$ . Then  $n \in n$ , violating 4L (b). And if  $m \in n$  and  $n \in m$ , then  $n \in n$  by transitivity of  $n$ , again violating 4L(b).

Now to show that at least one must hold. Use induction:  $T = \{ n \mid \text{for all } m, \text{ either } m \in n \text{ or } n = m \text{ or } n \in m \}$ .

$0 \in T$ : Let's show by induction that for each  $m$ , either  $0 = m$  or  $0 \in m$ . Clearly true for  $m = 0$ . Now suppose that  $0 = m$  or  $0 \in m$ . Now,  $m \in m^+$ . From this and the first disjunct, since  $0 \in m^+$ . From this and the second disjunct,  $0 \in m^+$  by transitivity of  $m^+$ . Either way,  $0 = m^+$  or  $0 \in m^+$ . Induction complete. It follows, then, that for each  $m$ , either  $m \in 0$  or  $0 = m$  or  $0 \in m$  – i.e.,  $0 \in T$ .

Now suppose that  $n \in T$ , and consider  $n^+$ . We must show that for all  $m$ , either  $n^+ \in m$ ,

$n^+=m$ , or  $m \in n^+$ . By the inductive hypothesis, either  $m \in n$ , or  $n \in m$  or  $n=m$ .

If  $m \in n$  then since  $n \in n^+$ ,  $m \in n^+$  by transitivity of  $n^+$ .

If  $n=m$  then since  $n \in n^+$ ,  $m \in n^+$

If  $n \in m$  then  $n^+ \in m^+$  by 4L(a). So by (\*),  $n^+ \in m$  or  $n^+=m$ .

In each case, one of the disjuncts,  $n^+ \in m$ ,  $n^+=m$ , or  $m \in n^+$  holds.

So:  $\in$  is our ordering on  $\omega$ . Since each number is a transitive set, we can also, equivalently, use the proper subset relation,  $\subset$ , (i.e., subset of and not identical to) as the ordering:

**Corollary 4M:** for any  $n, m \in \omega$ ,  $m \in n$  iff  $m \subset n$ , and  $(m \in n$  or  $m=n)$  iff  $m \subseteq n$

Pf: Suppose  $m \in n$ . If some  $o \in m$ , then by transitivity of  $n$ ,  $o \in n$ ; so  $m \subseteq n$ . Since  $m \in n$ ,  $m \neq n$  by trichotomy. So  $m \subset n$ . Conversely, suppose  $m \subset n$ . Then  $m \neq n$ . Moreover,  $n \notin m$  (if  $n \in m$  then by transitivity of  $m$ ,  $n \subseteq m$ ; but then since  $m \subseteq n$ ,  $n=m$ .) So by trichotomy,  $m \in n$ .

The second half follows from the first half by elementary logic (we added the same disjunct to both sides of the first half).

Next we have a theorem that's crucial to mathematics, concerning the interaction of  $+$ ,  $\cdot$  and  $<$ :

**Theorem 4N:** For any  $m, n, p \in \omega$ ,

$$\begin{aligned} m \in n &\text{ iff } m+p \in n+p \\ m \in n &\text{ iff } m \cdot p \in n \cdot p \quad \text{if } p \neq 0 \end{aligned}$$

Pf: In each case by induction on  $p$ . + one first. Let  $T = \{p \mid \forall m \forall n \ m \in n \text{ iff } m+p \in n+p\}$ .  $0 \in T$  obviously, given A1. Now suppose  $p \in T$ , consider  $p^+$  and any  $m, n$ . By the i.h.,  $m \in n$  iff  $m+p \in n+p$ . By 4L (a),  $m+p \in n+p$  iff  $(m+p)^+ \in (n+p)^+$ , and by A2 we have  $(m+p)^+ = m+p^+$  and  $(n+p)^+ = n+p^+$ ; hence we have  $m \in n$  iff  $m+p^+ \in n+p^+$ . (Note: Enderton did this differently.)

Now  $\cdot$ . (Here I do it Enderton's way.) First left to right. Let  $m \in n$ . And let  $T = \{q \mid m \cdot q^+ \in n \cdot q^+\}$ . (We consider such a  $T$  since our theorem is to apply only to  $p \neq 0$ , and every  $p$  other than zero is  $q^+$  for some  $q$ .)

First let's show  $0 \in T$ . i.e., we must show that  $m \cdot 0^+ \in n \cdot 0^+$ . Now,

$$\begin{aligned} m \cdot 0^+ &= m \cdot 0 + m && \text{(by M2)} \\ &= 0 + m && \text{(by M1)} \\ &= m + 0 && \text{(by commutativity)} \end{aligned}$$

$$= m \quad (\text{by A1})$$

Likewise,  $n \cdot 0^+ = n$ . So since  $m \in n$ ,  $m \cdot 0^+ \in n \cdot 0^+$ .

Next suppose that  $q \in T$ ; consider  $q^+$ . Since  $q \in T$ ,  $m \cdot q^+ \in n \cdot q^+$ . We must show that  $m \cdot q^{++} \in n \cdot q^{++}$ ; i.e., (by M2) that  $m \cdot q^+ + m \in n \cdot q^+ + n$ :

Since  $m \cdot q^+ \in n \cdot q^+$ , by the first half of this theorem,  $m \cdot q^+ + m \in n \cdot q^+ + m$ .

Since  $m \in n$ ,  $m + n \cdot q^+ \in n + n \cdot q^+$ , again by first half of this theorem

So  $n \cdot q^+ + m \in n \cdot q^+ + n$  (commutativity)

So, from steps 1 and 3 by transitivity,  $m \cdot q^+ + m \in n \cdot q^+ + n$

So,  $T$  is inductive, and so  $= \omega$ .

So since  $T$  is  $\omega$ , and since every  $p$  other than  $z = q^+$  for some  $q$ , we know that if  $p \neq 0$ , then  $m \cdot p \in n \cdot p$ .

Now right to left. Suppose  $p \neq 0$  and  $m \cdot p \in n \cdot p$ .  $n \neq m$ ; otherwise  $m \cdot p \in m \cdot p$ , in violation of 4L (b). Moreover,  $n \notin m$  (otherwise by the left to right direction,  $n \cdot p \in m \cdot p$ , violating trichotomy.) Thus, by trichotomy,  $m \in n$ .

The significance of theorem 4N is that we can get the usual cancellation laws for  $+$  and  $\cdot$ :

**Corollary 4P**      If  $m+p = n+p$  then  $m=n$   
                           If  $m \cdot p = n \cdot p$  and  $p \neq 0$  then  $m=n$

Pf: this is easy given trichotomy and 4N. e.g., take the first case: suppose  $m+p = n+p$ . By trichotomy,  $m+p \notin n+p$  and  $n+p \notin m+p$ ; so by 4N,  $m \notin n$  and  $n \notin m$ ; and so by trichotomy,  $m=n$ . Similarly for the second case.

One final bit of our reconstruction of arithmetic: we should show that proofs by *strong induction* work. What we have been induction (namely, establish that the result holds for 0, and holds for  $k^+$  if it holds for  $k$ ) is called *weak* induction. To prove that every  $n \in P$  by strong induction, what you do is prove that (\*) for every  $m$ , if all smaller numbers are  $\in P$ , then so is  $m$ . This method of proof relies on the principle that every set of integers has a least member. For suppose the principle, and suppose for reductio that not every integer  $\in P$ . Then there is some set,  $S$ , of integers that are not in  $P$ . By (\*),  $S$  has a least element,  $k$ . But that means that every integer smaller than  $k$  *does* have  $P$ , which contradicts (\*).

So, we must prove the least-number principle, or what's called in set theory the *well-ordering principle*. Say that  $S$  has a *least element* iff for some  $m \in S$ ,  $m \in n$  or  $m=n$  for all  $n \in S$ .

**Well ordering of  $\omega$ :** Any nonempty subset of  $\omega$  has a least element.

Pf: (This is a bit tricky.) Let  $A \subseteq \omega$  and suppose that  $A$  has no least member. We'll show that  $A = \emptyset$ .

Let  $B = \{m \in \omega \mid \text{no } n \text{ smaller than } m \text{ is in } A\}$ . (remember: smaller-than =  $\in$ ). We'll show that  $B$  is inductive. (The intuitive idea is this: if no  $n$  smaller than  $m$  is in  $A$ , that means that  $n$  is a lower limit for being in  $A$ . By showing that  $B$  is inductive, we will keep pushing the lower limit for being in  $A$  higher and higher.)

$0 \in B$  vacuously because nothing is a member of  $0$ .

Next suppose that  $m \in B$ , and consider  $m^+$ . We must show that no  $n$  smaller than  $m^+$  is in  $A$ . Suppose for reductio that  $n \in m^+$  and  $n \in A$ . Then  $n \in m$  or  $n = m$ . But  $n \notin m$  because  $m \in B$ . So  $n = m$ . But since  $m \in B$ , that means that  $n \in A$  and no  $o$  smaller than  $n$  is in  $A$ . Hence,  $A$  has a least member – contradiction.

So,  $B$  is inductive. Now, suppose for reductio that  $A$  is nonempty – i.e., that some  $o \in A$ .  $o$  is smaller than  $o^+$ , so  $o^+ \notin B$ . Contradiction.

We can now justify proofs by strong induction:

**Strong induction principle:** Suppose  $B \subseteq \omega$ , and is such that:

for every  $m \in \omega$ , if (for every  $n \in m$ ,  $n \in B$ ) then  $m \in B$

then  $B = \omega$

Proof: Suppose  $B \neq \omega$ . Then  $\omega - B$  is nonempty, and so by the well-ordering principle has a least member  $m$ . So  $m \in \omega$  and  $m \notin B$ . By the supposition in the theorem, for some  $n \in m$ ,  $n \notin B$ , contradicting the fact that  $m$  was the least member of  $\omega - B$ .

## V. Construction of real numbers

In this chapter we're going to build on our construction of  $\omega$  by constructing, in turn: integers (both positive and negative), then rational numbers (fractions), then real numbers.

In each case, we're going to construct the new set of numbers as constructions out of the old set of numbers. We'll need to define the operations we want (e.g., division) on the numbers thus

constructed. And we'll need to verify that the operations have the desired properties.

## A. Integers

### 1. Integers defined as differences

We need to expand our domain of numbers to include negatives. Now, each negative number is the difference between two positive numbers. E.g.,  $-2=2-4$ . So we can define a negative number as a pair of a natural number and a larger natural number, e.g.,  $-2 = \langle 2,4 \rangle$ . Think of an ordered pair of integers,  $\langle n,m \rangle$ , as a *difference* (the difference between  $n$  and  $m$ ).

But why define  $-2$  as  $\langle 2,4 \rangle$  rather than  $\langle 1,3 \rangle$ ? The choice would be arbitrary. So we'll take  $-2$  to be an equivalence class of differences:  $\{ \langle 0,2 \rangle, \langle 1,3 \rangle, \langle 2,4 \rangle, \dots \}$ .

Here's how we do it:

**Definition:**  $\sim$  is the relation on  $\omega \times \omega$  such that  $\langle m,n \rangle \sim \langle p,q \rangle$  iff  $m+q=p+n$

(Comment: the idea is to have the difference  $m-n$  equal the difference  $p-q$ . But we don't have difference defined on integers. However, we anticipate that  $m-n=p-q$  will amount to the same thing as  $m+q=p+n$  – that's what guides our definition. Of course, in the end, the test of whether it's a good definition is whether it lets us prove the right sorts of things about the resulting “negative integers”.)

**Theorem 5ZA**  $\sim$  is an equivalence relation on  $\omega \times \omega$

*Reflexivity:* this amounts to:  $m+n=m+n$

*Symmetry:* this amounts to showing that if  $m+q=p+n$  then  $p+n=m+q$  – follows from commutativity of  $+$

*Transitivity:* this amounts to showing that if  $m+q=p+n$  and  $p+s=r+q$  then  $m+s=r+n$ . Given the first two statements, by 4P,  $(m+q)+(p+s)=(p+n)+(r+q)$ . By commutativity and associativity, that means that  $(m+s)+(p+q)=(r+n)+(p+q)$ , and so by 4P,  $m+s=r+n$ .

**Definition** The set  $\mathbb{Z}$  of integers is the set  $(\omega \times \omega)/\sim$  (the set of equivalence classes of differences)

Note that this definition includes what are intuitively positive integers as well, for instance  $2_{\mathbb{Z}} = \{ \langle 2,0 \rangle, \langle 3,1 \rangle, \dots \}$ .

## 2. Integer addition

We need now to define an addition operation on  $\mathbb{Z}$ . We will be guided in our definition by this fact about normal addition:

$$(m-n) + (p-q) = (m+p) - (n+q)$$

So the basic idea will be define the  $+_{\mathbb{Z}}$  operation thus:

$$(+_{\mathbb{Z}}) [\langle m, n \rangle] +_{\mathbb{Z}} [\langle p, q \rangle] = [\langle m+p, n+q \rangle],$$

where  $+$  on the right side is the operation of addition on  $\omega$ .

But for this to be a well-defined operation, we need to be sure that if we choose other representatives from  $[\langle m, n \rangle]$  and  $[\langle p, q \rangle]$ , we will still arrive at a member of the same set  $[\langle m+p, n+q \rangle]$ . This is the point of the following lemma:

**Lemma 5ZB:** if  $\langle m, n \rangle \sim \langle m', n' \rangle$  and  $\langle p, q \rangle \sim \langle p', q' \rangle$ , then  $\langle m+p, n+q \rangle \sim \langle m'+p', n'+q' \rangle$

Pf: What we are given is:

$$m+n' = m'+n \quad \text{and} \quad p+q' = p'+q$$

$$\text{So, by 4P } m+n'+p+q' = m'+n+p'+q$$

$$\text{So by commutativity and associativity, } (m+p)+(n'+q') = (m'+p')+(n+q)$$

The latter of which is what we want.

Now we can define addition thus:

$$+_{\mathbb{Z}} \text{ is the function on } \mathbb{Z} \text{ such that } [\langle m, n \rangle] +_{\mathbb{Z}} [\langle p, q \rangle] = [\langle m+p, n+q \rangle]$$

We can use theorem 3Q to argue that that a unique such function exists. Let  $F$  be the function from  $(\omega \times \omega) \times (\omega \times \omega)$  into  $\omega \times \omega$  defined thus:  $F(\langle m, n \rangle, \langle p, q \rangle) = \langle m+p, n+q \rangle$ . We know  $\sim$  is an equivalence relation on  $\omega \times \omega$ , and 5ZB established that  $F$  is compatible with  $\sim$ . So by 3Q (the “analogous result”, there exists a unique function  $+_{\mathbb{Z}}$  from  $\omega \times \omega / \sim \times \omega \times \omega / \sim$  into  $\omega \times \omega / \sim$  such that for any  $\langle m, n \rangle$  and  $\langle p, q \rangle$  in  $\omega \times \omega$ ,  $+_{\mathbb{Z}}([\langle m, n \rangle], [\langle p, q \rangle]) = [F(\langle m, n \rangle, \langle p, q \rangle)]$  – which is what we want.

We next need to establish that  $+_{\mathbb{Z}}$  has all the properties we want.

**Theorem 5ZC**  $+_{\mathbb{Z}}$  is commutative and associative

(I’ll skip this; it’s boring.)

### 3. Zero, inverses, subtraction

**Definition:**  $0_Z$  is  $[\langle 0, 0 \rangle]$ .

**Theorem 5ZD** (a)  $0_Z$  is an identity element for  $+_Z$ :  $a +_Z 0_Z = a$   
 (b) Additive inverses exist: for any  $a \in \mathbb{Z}$ , there is a  $b \in \mathbb{Z}$  such that  
 $a +_Z b = 0_Z$

Pf. (a) is obvious. As for (b), where  $a = \langle m, n \rangle$ , choose the inverse  $b$  to be  $\langle n, m \rangle$ . Then  
 $[\langle m, n \rangle] +_Z [\langle n, m \rangle] = [\langle m+n, n+m \rangle]$ , but the latter integer is  $0_Z$  since:

$$\begin{aligned} m+n+0 &= 0+n+m \\ \text{so, } \langle m+n, n+m \rangle &\sim \langle 0, 0 \rangle \end{aligned}$$

We can prove that inverses are unique: if  $b$  and  $b'$  are inverses of  $a$  then  $b=b'$ .

Pf: we can't just argue that  $a +_Z b' = 0_Z = a +_Z b$ , then cancel  $a$ 's, since we haven't proved the cancellation laws yet. The proof Enderton gives is cool:

$$b = b + (a +_Z b') = (b +_Z a) +_Z b' = b'$$

Since inverses are unique, we can introduce an inverse operation:  $-a$  is the inverse of  $a$ . And given this, we can introduce a subtraction operation:  $a - b = a +_Z (-b)$

### 4. Integer multiplication

We do a similar thing to the introduction of  $+_Z$ : look for an equation we want  $\cdot_Z$  to obey. Informally:

$$(m-n) \cdot (p-q) = mp + nq - (np + mq)$$

So we want the following equation to define  $\cdot_Z$ :

$$[\langle m, n \rangle] \cdot_Z [\langle p, q \rangle] = [\langle mp + nq, np + mq \rangle]$$

This is OK, as before, if the resulting equivalence class is independent of the choice of  $\langle m, n \rangle$  out of  $[\langle m, n \rangle]$  and the choice of  $\langle p, q \rangle$  out of  $[\langle p, q \rangle]$  (i.e., that the function  $f(\langle m, n \rangle) = \langle mp + nq, np + mq \rangle$  is compatible, in the sense of 3Q, with  $\sim$ .) This lemma establishes compatibility:

**Lemma 5ZE** if  $\langle m,n \rangle \sim \langle m',n' \rangle$  and  $\langle p,q \rangle \sim \langle p',q' \rangle$ , then  
 $\langle mp+nq, np+mq \rangle \sim \langle m'p'+n'q', n'p'+m'q' \rangle$

- Pf:
1.  $m+n' = m'+n$  (given)
  2.  $p+q' = p'+q$  (given)
  3.  $mp+n'p = m'p+np$  (multiplied 1 by  $p$ )
  4.  $m'+n = m+n'$  (from 1)
  5.  $m'q+nq = mq+n'q$  (multiplied 4 by  $q$ )
  6.  $pm'+q'm' = p'm'+qm'$  (multiplied 2 by  $m'$ )
  7.  $p'+q = p+q'$  (from 2)
  8.  $p'n'+qn' = pn'+q'n'$  (multiplied 7 by  $n'$ )
  9.  $mp+n'p+m'q+nq+pm'+q'm'+p'n'+qn' = m'p+np+mq+n'q+p'm'+qm'+pn'+q'n'$   
 (added 3, 5, 6, 8)
  10.  $mp+nq+q'm'+p'n' = np+mq+p'm'+q'n'$  (9, cancellation)

Now we need to establish the needed properties of  $\cdot_Z$ :

**Theorem 5ZF**  $\cdot_Z$  is commutative, associative, and distributive over  $+_Z$

I won't prove all of these, just commutativity. Consider any  $a, b \in \mathbb{Z}$ . Suppose  $a = \langle m,n \rangle$  and  $b = \langle p,q \rangle$ . Then by definition of  $\cdot_Z$  we have:

$$a \cdot_Z b = \langle mp+nq, np+mq \rangle$$

$$b \cdot_Z a = \langle pm+qn, qm+pn \rangle$$

These are identical iff  $\langle mp+nq, np+mq \rangle \sim \langle pm+qn, qm+pn \rangle$ . But in fact, these ordered pairs are identical, given the commutativity of addition and multiplication on  $\omega$ , so they certainly stand in  $\sim$ . (In other cases of proving this sort of thing, one might have had to look into the definition of  $\sim$ , and use other properties of  $+$  and  $\cdot$  on  $\omega$ .)

In both  $\omega$  and  $\mathbb{Z}$ ,  $0$  is an *identity element* for  $+$ , in that  $x+0=x$ , for all  $x$ . An identity element for  $\cdot$ , therefore, would be an object  $b$  such that  $x \cdot b = x$ , for all  $x$ . Call  $1_Z$  the integer  $\langle 1,0 \rangle$ :

**Theorem 5ZG**

- (a)  $1_Z$  is a multiplicative identity element
- (b)  $0_Z \neq 1_Z$
- (c) whenever  $a \cdot_Z b = 0_Z$ , then either  $a = 0_Z$  or  $b = 0_Z$

Pf. (a):  $\langle m,n \rangle \cdot_Z \langle 1,0 \rangle = \langle m \cdot 1 + n \cdot 0, n \cdot 1 + m \cdot 0 \rangle = \langle m,n \rangle$ . Note that this assumes that

1 is a multiplicative identity element for  $\omega$ . We never proved that. But consider any  $m \in \omega$ :

$$\begin{aligned} m \cdot 1 &= m \cdot 0^+ && \text{(definition of 1)} \\ &= m \cdot 0 + m \\ &= 0 + m \\ &= m \end{aligned}$$

(b) we must show that it's not true that  $\langle 0, 0 \rangle \sim \langle 1, 0 \rangle$ , i.e., that  $0+0 \neq 0+1$ , which is true.

(c) see book.

Note that we don't have multiplicative inverses – i.e., we can't do division. For that we need to wait for the rationals.

### 5. Ordering on $\mathbb{Z}$

Now we want an ordering relation on the positive integers. To guide our definition, let's make use of informal mathematical knowledge. When should  $[\langle m, n \rangle]$  be less than  $[\langle p, q \rangle]$ ? When  $m < p - q$ , i.e., when  $m + q < p + n$ , i.e. (since our  $<$  is  $\in$ ), when  $m + q \in p + n$ .

So, we want our definition to have the consequence that for any  $m, n, p, q$ ,

$$[\langle m, n \rangle] <_Z [\langle p, q \rangle] \text{ iff } m + q \in p + n$$

But for this to be a well-defined relation, we need to be sure that it didn't matter which representatives  $m, n, p, q$  we took from the integers. Thus:

**Lemma 5ZH** If  $\langle m, n \rangle \sim \langle m', n' \rangle$  and  $\langle p, q \rangle \sim \langle p', q' \rangle$  then  $m + q \in p + n$  iff  $m' + q' \in p' + n'$

|     |   |   |
|-----|---|---|
| Pf: | $m + n' = m' + n$   | (by hypothesis)                             |
|     | $p + q' = p' + q$   | (by hypothesis)                             |
|     | $m + q \in p + n$ iff $m + q + n' + p' \in p + n + n' + p'$ | ( $4N$ – added $n' + p'$ to both sides)     |
|     | $m + q \in p + n$ iff $m' + n + p + q' \in p + n + n' + p'$ | (by above equations)                        |
|     | $m + q \in p + n$ iff $m' + q' \in p' + n'$                 | ( $4N$ – cancelled $n + p$ from both sides) |

OK, what does this Lemma do for us? (Enderton doesn't really discuss this.) It lets us define  $<_Z$  rigorously as follows:

**Definition:**  $a <_Z b$  iff for some  $m, n, p, q, \langle m, n \rangle \in a, \langle p, q \rangle \in b,$  and  $m+q \in p+n$

This has the form of a proper definition. And it follows from this definition, via Lemma 5ZH, that:

for any  $m, n, p, q, [\langle m, n \rangle] <_Z [\langle p, q \rangle]$  iff  $m+q \in p+n$

Pf:

$[\langle m, n \rangle] <_Z [\langle p, q \rangle]$  iff for some  $m', n', p', q', \langle m', n' \rangle \in [\langle m, n \rangle], \langle p', q' \rangle \in [\langle p, q \rangle],$   
and  $m'+q' \in p'+n'.$

(by definition of  $<_Z$ ). But wherever  $\langle m', n' \rangle \in [\langle m, n \rangle], \langle p', q' \rangle \in [\langle p, q \rangle],$  we have:  
 $\langle m, n \rangle \sim \langle m', n' \rangle,$  and  $\langle p, q \rangle \sim \langle p', q' \rangle.$  And so, 5ZH (and the fact that equivalence classes are nonempty) tells us that:

(for some  $m', n', p', q', \langle m', n' \rangle \in [\langle m, n \rangle], \langle p', q' \rangle \in [\langle p, q \rangle],$  and  $m'+q' \in p'+n')$  iff  
 $m+q \in p+n$

Note that we didn't need to use 3Q to be assured that the relation  $<_Z$  exists; we know it exists because it's a subset of  $\mathbb{Z} \times \mathbb{Z}.$  But we do need the lemma to prove that the definition implies the result that we want.

OK, I won't prove all the important results that are proved in the book; I'll just state many of them.:

**Theorem 5ZI**  $<_Z$  is a linear order on  $\mathbb{Z}$

**Theorem 5ZJ** for any integers,  $a, b, c:$

- (a)  $a <_Z b$  iff  $a+_Z c <_Z b+_Z c$
- (b)  $a <_Z b$  iff  $a \cdot_Z c <_Z b \cdot_Z c$  (if  $0_Z <_Z c$ )

(from this and trichotomy we then get the cancellation laws.)

## B. Rationals

### 1. rational defined as fractions

The book now drops the subscript  $Z$  on  $0_Z, +_Z,$  etc.

We now need to construct the rationals, i.e., fractions. The basic thing here is adding

multiplicative inverses, by analogy with the way the move to  $\mathbb{Z}$  from  $\omega$  added additive inverses. Recall that the additive inverse of  $a$  is a number  $-a$  such that  $a + (-a) = 0$ . And  $0$ , is the additive identity element. Thus, the multiplicative inverse of a number  $a$ , will be a number  $b$  such that  $a \cdot b = 1$ , where  $1$  is the multiplicative identity element.

So, the obvious thing is to define a rational as a fraction – as an ordered pair  $\langle a, b \rangle$  of integers where  $b \neq 0$  (think of  $\langle a, b \rangle$  as standing for  $a/b$ ). But of course, we don't want to count the fractions  $\langle 1, 2 \rangle$  and  $\langle 2, 4 \rangle$  as being different from each other, so we'll need an equivalence relation. The equivalence relation is guided by our normal mathematics of fractions thus:

$$a/b = c/d \text{ iff } ad = cb$$

Let  $\mathbb{Z}' = \mathbb{Z} - \{0\}$

**Definition**  $\sim$  is the binary relation on  $\mathbb{Z} \times \mathbb{Z}'$  such that

$$\langle a, b \rangle \sim \langle c, d \rangle \text{ iff } ad = cb$$

**Theorem 5QA**  $\sim$  is an equivalence relation on  $\mathbb{Z} \times \mathbb{Z}'$

Proof of transitivity: suppose  $\langle a, b \rangle \sim \langle c, d \rangle$  and  $\langle c, d \rangle \sim \langle e, f \rangle$ . We then have:

$$ad = bc \text{ and } cf = de.$$

We want to establish:  $af = be$ . To that end, multiply both sides of the first equation by  $f$ :

$$adf = bcf$$

And both sides of the second equation by  $b$ :

$$cfb = deb$$

so:

$$adf = deb$$

Now, since  $\langle c, d \rangle \in \sim$ , that means that  $d \neq 0$ , so we can cancel  $d$  from both sides:

$$af = be$$

We define  $0_Q$  as  $\langle 0, 1 \rangle$  and  $1_Q$  as  $\langle 1, 1 \rangle$ .

## 2. Rational addition and multiplication

How will we define our operations of  $+_{\mathbb{Q}}$  and  $\cdot_{\mathbb{Q}}$ ? As usual, we're guided by our informal knowledge of the rationals:

$$a/b + c/d = (ad+cb)/bd \quad a/b \cdot c/d = ac/bd$$

### Definitions

$$[\langle a,b \rangle] +_{\mathbb{Q}} [\langle c,d \rangle] = [\langle ad+cb, bd \rangle]$$

$$[\langle a,b \rangle] \cdot_{\mathbb{Q}} [\langle c,d \rangle] = [\langle ac, bd \rangle]$$

As before, we need lemmas to insure us (via 3Q) that these definitions work – that the resulting sums and products are independent of the choices of  $\langle a,b \rangle$  and  $\langle c,d \rangle$ :

### Lemmas 5QB and 5QD

For any  $a, b, c, d, a', b', c', d' \in \mathbb{Z}$ , if  $\langle a,b \rangle \sim \langle a',b' \rangle$  and  $\langle c,d \rangle \sim \langle c',d' \rangle$  then:

$$\langle ad+cb, bd \rangle \sim \langle a'd'+c'b', b'd' \rangle, \text{ and} \\ \langle ac, bd \rangle \sim \langle a'c', b'd' \rangle$$

Enderton also proves theorems insuring that  $+_{\mathbb{Q}}$  and  $\cdot_{\mathbb{Q}}$  work properly:

### Theorem 5QC

- (a)  $+_{\mathbb{Q}}$  is commutative and associative
- (b)  $0_{\mathbb{Q}}$  is an identity element for  $+_{\mathbb{Q}}$
- (c) additive inverses exist:  $\forall r \in \mathbb{Q} \exists s \in \mathbb{Q}: r +_{\mathbb{Q}} s = 0_{\mathbb{Q}}$

As before, the additive inverse is unique; we denote it as  $-r$ .

### Theorem 5QE

$\cdot_{\mathbb{Q}}$  is associative, commutative, and distributive over  $+_{\mathbb{Q}}$

Now we get the new thing: the existence of multiplicative inverses:

### Theorem 5QF

$\forall r \in \mathbb{Q}$ , if  $r \neq 0_{\mathbb{Q}}$  then  $\exists q \in \mathbb{Q}: q \neq 0_{\mathbb{Q}}$  and  $r \cdot_{\mathbb{Q}} q = 1_{\mathbb{Q}}$

Pf: since  $r \neq 0_{\mathbb{Q}}$ ,  $r = [\langle a,b \rangle]$  for some  $a \neq 0$ . (it's immediate that  $\langle 0,c \rangle \sim \langle 0,d \rangle$  for any  $c, d \in \mathbb{Z}$ .) Let  $q$  be  $[\langle b,a \rangle]$ . This is indeed a rational since  $a \neq 0$ . And  $r \cdot_{\mathbb{Q}} q = [\langle ab, ba \rangle]$ , and  $[\langle ab, ba \rangle] \sim [\langle 1, 1 \rangle]$  since  $ab \cdot 1 = 1 \cdot ba$ .

As before we can show that multiplicative inverses are unique. We call  $s$ 's multiplicative inverse:  $s^{-1}$ .

### 3. Ordering on $\mathbb{Q}$

From our informal knowledge of the rationals:

We know that if  $b, d$  are positive then  $a/b < c/d$  iff  $ad < cb$ . So, to define  $r <_Q s$ , we simply need to choose  $\langle a, b \rangle \in r$  and  $\langle c, d \rangle \in s$ , where  $b$  and  $d$  are positive, and ask whether  $ad < cb$ . Will we always be able to choose such  $a, b, c, d$ ? Yes: we can always find a fraction with positive denominator in any rational: for any  $r$ , some  $\langle a, b \rangle \in r$ , where  $b \neq 0$ . If  $b > 0$  then  $\langle a, b \rangle$  is our desired fraction; otherwise,  $\langle -a, -b \rangle$  is (since  $\langle -a, -b \rangle \sim \langle a, b \rangle$ .) (We earlier proved that every integer is exactly one of positive, negative, or zero.)

So, our definition is:

**Definition:**  $r <_Q s$  iff for some  $\langle a, b \rangle \in r$  and some  $\langle c, d \rangle \in s$ , where  $0 <_Z b$  and  $0 <_Z d$ ,  $ad < cb$

We need this lemma to justify the definition's independence of the particular integers chosen:

**Lemma 5QH** If  $\langle a, b \rangle \sim \langle a', b' \rangle$  and  $\langle c, d \rangle \sim \langle c', d' \rangle$ , and  $b, b', d, d'$  are all positive, then  $ad < cb$  iff  $a'd' < c'b'$

**Theorem 5QI**  $<_Q$  is a linear order on  $\mathbb{Q}$

**Theorem 5QJ**

- (a)  $r <_Q s$  iff  $r +_Q t < s +_Q t$
- (b)  $r <_Q s$  iff  $r \cdot_Q t < s \cdot_Q t$  if  $t$  is positive (i.e.,  $0_Q <_Q t$ )

As before, we get cancellation laws (e.g., if  $r +_Q t = s +_Q t$  then  $r = s$ ), only now, since we have both additive inverses and multiplicative inverses, we can prove the cancellation laws without using the order-preserving features of  $\cdot_Q$  and  $+_Q$ , just by adding and multiplying  $-t$  and  $t^{-1}$ , respectively.

## C. Reals

### 1. The need for more numbers

We have been considering various systems of numbers. In each case, newer systems had stronger closure properties. E.g., the (positive) rationals are closed under multiplicative inverses.

There are a couple closure properties that our rationals lack:

*Closure of S under roots:*

For any  $x \in S$  and any  $n \in (S\text{'s "copy" of } \omega)$ ,  $\exists y \in S: y^n = x$  ( $y$  is the  $n^{\text{th}}$  root of  $x$ )

(we didn't talk about copies, but the idea is that the integers and rationals have isomorphic copies of  $\omega$ . See Enderton.)

*Closure of S under least upper bounds.* An *upper bound* of a subset  $T$  of  $S$  is an object  $x$  that is greater than or equal to every member of  $T$ . (This leaves it open whether  $x \in T$ .) A *least upper bound* of  $T$  is an upper bound of  $T$  that is less than every other upper bound of  $T$ . Closure of  $S$  under least upper bounds means that every nonempty subset of  $S$  that has an upper bound has a least upper bound.

Here is a proof sketch that  $\mathbb{Q}$  is not closed under roots:

Suppose that 2 has a rational square root, call it  $p/q$ . Then  $2 = p^2/q^2$ , so  $2q^2 = p^2$ , where  $p, q$  are positive integers. Now, the fundamental theorem of arithmetic (which we didn't prove; but its proof just depends on features of  $\omega$  that we *did* prove) states that every positive integer has a unique factorization into primes. Thus, every positive integer has a unique number of 2s in its prime factorization. Thus,  $p^2$  and  $q^2$  each have an even number of 2s in their prime factorizations. But then  $2q^2$  has an odd number of 2s in its prime factorization, and so  $\neq p^2$ .

I won't prove that  $\mathbb{Q}$  isn't closed under least upper bounds.

## 2. Alternate constructions of the reals

Enderton mentions a couple alternate ways to construct the reals, which are worth mentioning because you may encounter them at some point.

One way, which he says nearly nothing about, is to define a real as an infinite decimal: an ordered pair of an integer (the bit before the decimal) and an infinite sequence of digits (a function from  $\omega$  into 10). There's arbitrariness here (in the base – we chose 10 but could have chosen any other base). And we'll need to identify some sequences (e.g., .9 repeating with 1.0 repeating.)

A more common way is the method of Cauchy sequences. The basic idea is that a real will be a sequence of rationals that converges to a particular value. But we can't really say that, because

the value the sequence converges to will be a real number, and we haven't yet defined what the real numbers are. So we instead define a *Cauchy sequence* as an infinite sequence,  $s$ , of rationals (function from  $\omega$  into  $\mathbb{Q}$ ) where the differences between members of the sequence grow arbitrarily small:

$$(\forall \text{positive } \varepsilon \in \mathbb{Q})(\exists k \in \omega)(\forall m > k)(\forall n > k) |s_m - s_n| < \varepsilon$$

We say that Cauchy sequences  $r$  and  $s$  are *equivalent* iff differences between their members get arbitrarily small:

$$(\forall \text{positive } \varepsilon \in \mathbb{Q})(\exists k \in \omega)(\forall m > k) |r_m - s_m| < \varepsilon$$

We then take the reals to be the equivalence classes under this equivalence relation.

### 3. Dedekind cuts

Our method will be neither of those two. Our basic idea will be to define a real number,  $x$  as the set of rationals less than  $x$ . But of course we can't say precisely that before we define what a real number is. So what we do instead is this:

**Definition** A *Dedekind cut* is a subset  $x$  of  $\mathbb{Q}$  such that:

- a)  $\emptyset \neq x \neq \mathbb{Q}$
- b)  $x$  is "closed downward": if  $q \in x$  and  $r < q$  then  $r \in x$
- c)  $x$  has no largest member

We then define  $\mathbb{R}$  to be the set of all Dedekind cuts.

Note: we need no equivalence relation anymore; the reals are directly the Dedekind cuts.

### 4. Ordering on $\mathbb{R}$

We can just use the relation  $\subset$ :  $x <_{\mathbb{R}} y$  iff  $x \subset y$ .

**Theorem 5RA**  $<_{\mathbb{R}}$  is a linear order on  $\mathbb{R}$

Pf: Enderton says that it's obviously transitive; let's go more slowly. Suppose  $x \subset y$  and  $y \subset z$ . Then  $x \subset y$ ,  $x \neq y$ ,  $y \subset z$ ,  $y \neq z$ , and so,  $x \subset z$ . We must show that  $x \neq z$ . Suppose otherwise. Then we have  $z \subset y$ , so  $z \subset y$ , and so  $z = y$  – contradiction.

Now trichotomy: Suppose that  $x \neq y$  and  $x \not\subseteq y$ . We must show that  $y \subset x$ . Since  $x \neq y$ , we must show that  $y \subset x$ . So consider any  $q \in y$ . Now, since  $x \not\subseteq y$ , some  $r \in x$  but  $r \notin y$ . Now, it cannot be that  $r < q$ , because then by downward closure of  $y$ ,  $r \in y$ . So, by trichotomy for  $<$ :  $r = q$  or  $q < r$ . Obviously in the first case,  $q \in x$ ; and likewise, in the second case  $q \in x$  by downward closure of  $x$ .

**Theorem 5RB** Any bounded nonempty subset of  $\mathbb{R}$  has a least upper bound in  $\mathbb{R}$

Pf: Let  $A$  be any bounded nonempty subset of  $\mathbb{R}$ . We will show that  $\bigcup A$  is a least upper bound of  $A$ . That is, we will show that i)  $\bigcup A$  is an upper bound of  $A$  – i.e., that  $x \subseteq \bigcup A$ , for each  $x \in A$ , and ii)  $\bigcup A$  is a *least* upper bound of  $A$  – i.e., that  $\bigcup A \subseteq z$ , for any upper bound  $z$  of  $A$ , and finally iii)  $\bigcup A \in \mathbb{R}$ :

i) This is immediate given definition of  $\bigcup$

ii) Let  $z$  be any upper bound of  $A$ : and so, for each  $a \in A$ ,  $a \subseteq z$ . Thus,  $\bigcup A \subseteq z$ . (this bit is true for any sets; it follows just from the definition of  $\bigcup$ .)

iii) we must show that  $\bigcup A$  is a Dedekind cut.

First,  $\bigcup A \subseteq \mathbb{Q}$  since it is a union of sets of rationals, and it is nonempty since  $A$  is nonempty and has reals (Dedekind cuts) as members.

Second,  $\bigcup A$  isn't all of  $\mathbb{Q}$ .  $A$  is bounded, by the real number  $x$ , suppose.

Suppose now for reductio that  $\bigcup A$  is all of  $\mathbb{Q}$ ; we'll show that  $x$  is all of  $\mathbb{Q}$ , and so isn't a real: Let  $r \in \mathbb{Q}$ . So  $r \in y$ , for some  $y \in A$ . Since  $x$  is a bound for  $A$ ,  $y \subseteq x$ , so  $r \subseteq x$ , so  $r \in x$ . Thus, every rational is in  $x$ , and so  $x$  is all of  $\mathbb{Q}$ .

Third,  $\bigcup A$  is closed downward: any  $r \in \bigcup A$  is in some  $x \in A$ ; but members of  $A$  are closed downward, so any  $q < r$  is in  $x$ , and so is in  $\bigcup A$ .

Fourth,  $\bigcup A$  has no largest member: for suppose it had a largest member,  $q$ .  $q \in x$ , for some  $x \in A$ . But since  $x$  is a Dedekind cut, it has some member  $r$  larger

than  $q$ , whence  $r \in \bigcup A$ , contradicting the fact that  $q$  was a largest member of  $\bigcup A$ .

The fact that  $\mathbb{R}$  is closed under least upper bounds is one of the most important features of  $\mathbb{R}$ .

I won't go on and investigate all of the properties of  $+_{\mathbb{R}}$  and  $\cdot_{\mathbb{R}}$ ; I'll just note their definitions:

**Definitions**  $x +_{\mathbb{R}} y = \{q + r \mid q \in x \text{ and } r \in y\}$   
 $0_{\mathbb{R}} = \{r \in \mathbb{Q} \mid r < 0\}$

One then has to prove separately that  $0$ , and  $x +_{\mathbb{R}} y$ , thus defined, are themselves real numbers, that  $+_{\mathbb{R}}$  is commutative and associative, that additive inverses exist, that addition preserves order.

How about defining multiplication? We can't just offer the simple definition that  $x \cdot_{\mathbb{R}} y = \{rs \mid r \in x \text{ and } s \in y\}$  because this set will always be the entirety of  $\mathbb{R}$  (since  $x$  and  $y$  will always have arbitrarily negative rationals.) So instead we offer this disjunctive definition. The idea will be to use the simple definition as far as we can, and then just union in the rest of the desired real number. First we'll need the definition of  $|x|$  -- the absolute value of  $x$ . It is simply the larger of  $x$  and  $-x$ , which, given the definition of  $<_{\mathbb{R}}$  as  $\subset$ , means we can say:  $|x| = x \cup -x$ .

**Definition**

- (a) If  $x$  and  $y$  are nonnegative real numbers then  $x \cdot_{\mathbb{R}} y = 0_{\mathbb{R}} \cup \{rs \mid 0 \leq r \in x \text{ and } 0 \leq s \in y\}$
- (b) if  $x$  and  $y$  are both negative real numbers then  $x \cdot_{\mathbb{R}} y = |x| \cdot_{\mathbb{R}} |y|$
- (c) if one of the real numbers  $x$  and  $y$  is nonnegative and the other is negative, then  $x \cdot_{\mathbb{R}} y = -( |x| \cdot_{\mathbb{R}} |y| )$

One then has to establish that products thus defined are real numbers, and behave correctly.  
 Uncle!!

## VI. Cardinal numbers and the axiom of choice

### A. Equinumerosity

Our next topic is comparisons of size between sets. Set theory itself was spurred by Cantor's discovery in 1873 that some there are infinite collections that are "larger" than others. This contradicted old (and commonsensical) ideas about infinity, and showed that there was more complexity in the area than met the eye, and that needed to be studied.

We need to start by making precise exactly what we mean by "has the same size as", "is larger than", etc.

**Definition**  $A \approx B$  (A and B are equinumerous) iff there exists some one-to-one function from A onto B

There are familiar surprising examples of sets being equinumerous to each other, for instance  $\omega$  with the evens,  $\omega$  with  $\omega \times \omega$ ,  $\omega$  with  $\mathbb{Q}$ ,  $\mathbb{R}$  with  $(0,1)$ , etc.

Here's a general fact:

For any set, A,  $\wp A \approx 2^A$ .

Proof: for any  $B \subseteq A$ , let the characteristic function of B within A,  $c_B$ , be the following function from A into 2:

$$\begin{aligned} c_B(a) &= 1 \text{ if } a \in B \\ c_B(a) &= 0 \text{ if } a \notin B \text{ but } a \in A \end{aligned}$$

(it says "yes or no" to each member of A, as to whether it's in B.) We now map  $\wp A$  onto  $2^A$  via the function H, which assigns to any  $B \subseteq A$  its characteristic function. H is one-to-one: suppose  $H(B) = H(C)$ . Then B and C have the same members, since for any  $a \in A$ ,  $a \in B$  iff  $H(B)(a) = 1$ ; and likewise, for any  $a \in A$ ,  $a \in C$  iff  $H(C)(a) = 1$ .

**Theorem 6A** (a)  $\omega \not\approx \mathbb{R}$

(b) for every A,  $A \not\approx \wp A$

Pf: (a) is the standard diagonal argument. As for (b), suppose there exists a one-to-one function, f, from A onto  $\wp A$ . Let's define  $d \subseteq A$  as follows:

$$D = \{a \in A \mid a \notin f(a)\}$$

Now, since  $D \subseteq A$  and  $b$  is onto  $\wp A$ , for some  $d \in A$ ,  $f(d) = D$ . Now we ask: is  $d \in D$ ? If it is then it isn't, and if it isn't then it is. Contradiction.

## B. Finite sets

Now that we have the notion of equinumerosity at our disposal, we can finally define 'finite':

**Definition** A set is finite iff it is equinumerous with some natural number. Otherwise it is infinite.

(This works only because of our construction of the natural numbers. Each one was the set of all the earlier ones; thus, intuitively, each natural number  $n$  has  $n$  elements. So all the finite sized sets are represented.)

There are various other definitions we could use, e.g., "maps one-to-one onto one of its proper subsets." When one adopts one of the definitions, the biconditionals corresponding to the others become theorems (given the right background assumptions! – this is where the axiom of choice will get into the picture.) That's one thing that we're going to want to prove; another is that each finite set is equinumerous with a *unique* natural number. (thus, we can use the natural numbers as *measures* of the sizes of finite sets.)

**Pigeonhole principle:** No natural number is equinumerous with a proper subset of itself

Pf: Let  $T = \{n \in \omega \mid \text{every function from } n \text{ into } n \text{ has range } n\}$ . By proving  $T$  inductive, we will have proved the theorem.

Base:  $0 \in T$ : There's only one function from  $0$  into  $0$ :  $\emptyset$ , and its range is  $\emptyset$ , i.e.,  $0$ .

Induction: suppose  $k \in T$ , show  $k^+ \in T$ . So consider any function  $f$  from  $k^+$  into  $k^+$ ; we must show that  $\text{ran } f = k^+$ . Consider  $f \upharpoonright k$  (the restriction of  $f$  to  $k$ ). We now consider two subcases.

Case 1:  $k$  is closed under  $f$ . In that case,  $f \upharpoonright k$  is a function from  $k$  into  $k$ , and so by the inductive hypothesis,  $\text{ran } f \upharpoonright k$  is  $k$ . Now, the domain of  $f$  is  $k^+$ , and  $k \in k^+$ . So what could  $f(k)$  be? It must be in  $k^+$ , since  $f$  is into  $k^+$ , and  $k^+ = k \cup \{k\}$ , so it must be either in  $k$  or identical to  $k$ . It can't be in  $k$ , since  $f \upharpoonright k$  is already onto  $k$  and  $f$  is one-to-one. So it must be  $k$ . Thus, since  $f = f \upharpoonright k \cup \{ \langle k, k \rangle \}$  and  $f \upharpoonright k$  is onto  $k$ ,  $f$  is onto  $k^+$ .

Case 2:  $k$  isn't closed under  $f$ . I.e., for some  $n \in k$ ,  $f(n) \notin k$ , and so  $f(n) = k$ . We now need a function to which we can apply the inductive hypothesis. Now,  $k$  is mapped by  $f$  to some  $n \in k^+$ , and  $m \neq k$  because  $f(n) = k$  and  $f$  is one-to-one. So let's construct a new function,  $g$ , by "swapping" the values for  $n$  and  $k$ :

$$\begin{aligned} g(n) &= m \\ g(k) &= n \\ g(o) &= f(o), \text{ for all } o \in k^+ \text{ other than } n \text{ and } k. \end{aligned}$$

Now, given its construction,  $g$  is a one-to-one function because  $f$  was. Moreover, since  $g(k) = n$ ,  $k$  is closed under  $g$ , and so by Case 1,  $\text{ran } g = k^+$ . But  $\text{ran } g = \text{ran } f$  (since they differ only by a swap). So  $\text{ran } f = k^+$ .

**Corollary 6C** No finite set is equinumerous to a proper subset of itself

Pf: We here use a common technique, to investigate properties of a set by looking at corresponding properties of the set's image under a one-to-one function. Suppose  $A$  is finite and that  $g$  maps  $A$  one-to-one into  $A$ . Now, since  $A$  is finite, there is some one-to-one  $f$  from  $A$  onto some  $n \in \omega$ . Define a function from  $n$  into  $n$  thus:

$$h(k) = g(f(g^{-1}(k)))$$

(draw the picture!). Note that  $g^{-1}$  does indeed exist and has domain  $n$ , since  $g$  is one-to-one and onto  $n$ . Now,  $\text{dom } h$  is  $n$  ( $\text{dom } g^{-1}$  is  $n$ ,  $\text{ran } g^{-1}$  is  $A$  since  $\text{dom } g$  is  $A$ ;  $\text{dom } f$  is  $A$ ;  $\text{ran } f \subseteq A$ , and  $\text{dom } g$  is  $A$ ). And  $h$  is one-to-one (since it's a composition of one-to-one functions – exercise 17 from chapter 3.) So by the pigeonhole principle,  $g$  is onto  $n$ .

Now consider any  $a \in A$ . And consider  $g(a)$ . Since  $h$  is one-to-one, for some  $k \in n$ ,  $h(k) = g(a)$ . So  $g(f(g^{-1}(k))) = g(a)$ . But that means that something is mapped by  $f$  to  $a$  after all:  $g^{-1}(k)$ . (remember that  $g$  is one-to-one.)

**Corollary 6D** (a) Every set equinumerous to a proper subset of itself is infinite  
(b)  $\omega$  is infinite

Pf: (a) follows immediately from 6C. As for (b), consider this function:  $h(n) = n^+$ , for each  $n \in \omega$ . This is a one-to-one function from  $\omega$  into  $\omega - \{0\}$ . So  $\omega$  is equinumerous with a proper subset of itself, and so is infinite by (a).

**Corollary 6E** Every finite set is equinumerous with a unique natural number

Pf: By definition of finite, every finite set is equinumerous with a natural number. Suppose  $A$  is equinumerous with two natural numbers,  $n$  and  $m$ . By transitivity of  $\approx$ ,  $n \approx m$ . But we know by trichotomy and 4M that since  $n \neq m$ , either  $n \subset m$  or  $m \subset n$ , contradicting the pigeonhole principle.

Given 6E, we can introduce the term “card A” (“the cardinality of A”), standing for the natural number to which A is equinumerous, whenever A is finite. It follows from this definition that when A and B are finite, we have:

$$\text{card } A = \text{card } B \text{ iff } A \approx B$$

The idea here is that we’re using the natural numbers as *measures of size*. Each finite set A gets assigned one of these measures (= natural numbers). It’s a measure of A because it is equinumerous with A.

### C. Infinite cardinal numbers

Now, we want to do this generally. We want to find a way of extending the card notation so that to *any* set A, finite or infinite, we can associate a unique set, card A, such that:

$$\text{card } A = \text{card } B \text{ iff } A \approx B$$

It won’t matter much exactly which sets we choose to be the cards, just as it didn’t matter much exactly which sets we called the natural numbers. We’ll define ‘card’ so that it has this feature in the next chapter. (‘Card’ will obey the displayed equation; and card A will be the natural number equinumerous to A when A is finite.)

Call a set  $\kappa$  a “cardinal number” iff for some set A,  $\kappa = \text{card } A$ . The cardinal numbers will be our “infinite numbers”, since they measure the size of all sets in the way that the natural numbers measure the size of finite sets.

So, the members of  $\omega$  are cardinal numbers. But there are others. Let’s call card  $\omega$  “ $\aleph_0$ ”. Now,  $\aleph_0$  is not a member of  $\omega$  since  $\omega$  is infinite and so not equinumerous with any natural number.  $\aleph_0$  is an infinite cardinal, as are card  $\wp \omega$ , card  $\mathbb{R}$ , etc.

The final thing to do is verify that every subset of a finite set is finite.

**Lemma 6F** if C is a proper subset of some natural number n, then  $C \approx m$  for some natural number m less than n.

Pf: induction. Let  $T = \{n \in \omega \mid \text{every proper subset of } n \text{ is equinumerous to some member of } n\}$ .

$0 \in T$ : vacuous. Induction step: suppose  $k \in T$ , and consider any proper subset C of  $k^+$ . Now we separate cases.

Case 1:  $C \subset k$ . In that case, by the inductive hypothesis,  $C \approx m$ , for some  $m$  less than  $k$ ; since  $m < k^+$ ,  $k^+ \in T$ .

Case 2:  $C = k$ . Then  $C \approx k$  (use the identity map), and since  $k < k^+$ ,  $k^+ \in T$ .

Case 3:  $k \in C$ . We now need a proper subset of  $k$  to apply the inductive hypothesis to. Consider  $C \cap k$ . It is certainly a subset of  $k$ . And since  $k \in C$ , it must be a proper subset of  $k$  (otherwise,  $C$  would be all of  $k^+$ .) By the inductive hypothesis,  $C \cap k \approx m$ , for some  $m < k$ . Let  $f$  be a one-to-one function from  $C \cap k$  onto  $m$ . And define a function  $g$  with domain  $C \cap k \cup \{k\}$  thus:

$$\begin{aligned} g(k) &= m \\ g(x) &= f(x) \text{ if } x \in C \cap k \end{aligned}$$

Clearly  $g$  is one-to-one (since  $f$  was one-to-one and into  $m$ ); and it is onto  $m^+$  since  $f$  was onto  $m$ . But  $C \cap k \cup \{k\}$  is  $C$ . So  $g$  maps  $C$  one-one onto  $m^+$ . Since  $m < k$ ,  $m^+ < k^+$ . So  $k^+ \in T$ .

**Corollary 6G** Every subset of a finite set is finite

Pf. Suppose  $A \subseteq B$  and  $B$  is finite. So  $B \approx n$  for some  $n \in \omega$ . Let  $f$  be a one-one function from  $B$  onto  $n$ . Now consider  $f[A]$ . It is a subset of  $n$ . So it maps one-to-one onto some  $m \leq n$ , by the lemma. And so,  $A$  maps one-to-one onto some  $m \leq n$ , and so is finite.

#### D. Cardinal arithmetic

The notions of addition, multiplication, and addition were defined on the finite cardinals – the members of  $\omega$ . We want to *extend* these notions to the infinite cardinals. What that means is:

The extended notions coincide with the old notions for the finite cardinals

The extended notions behave *somewhat* like the old notions

We can't hope for the extended notions to behave *exactly* like the old notions since infinite sets break lots of rules. Consider, for example, the notion of equinumerosity. We defined it as mapping one-to-one onto. That in itself was an extension of the ordinary notion, and it behaves a lot like the ordinary notion even when applied to infinite sets (e.g., it's transitive, symmetric, reflexive, etc.) But it doesn't behave exactly like the ordinary notion: an infinite set can be equinumerous with a proper subset of itself.

Here are the definitions:

**Definition** where  $\kappa$  and  $\lambda$  are any cardinals,

- (a)  $\kappa + \lambda$  is  $\text{card}(K \cup L)$ , where  $K$  and  $L$  are any disjoint sets of cardinality  $\kappa$  and  $\lambda$  respectively
- (b)  $\kappa \cdot \lambda$  is  $\text{card}(K \times L)$ , where  $K$  and  $L$  are any sets of cardinality  $\kappa$  and  $\lambda$  respectively
- (c)  $\kappa^\lambda$  is  $\text{card } {}^L K$ , where  $K$  and  $L$  are any sets of cardinality  $\kappa$  and  $\lambda$  respectively

The definition of cardinal addition generalizes a very natural way of performing addition on finite sets. If you take two finite sets, of sizes  $n$  and  $m$ , and they don't overlap, then the number  $n+m$  is just the number of things in their union. If they *do* overlap, then just choose two other sets that don't overlap, and have the same numbers of things as the first two sets, respectively, and then take *their* union. Likewise for multiplication: to multiply  $n$  and  $m$ , count  $n$  collections of  $m$  things; but  $N \times M$ , where  $N$  has  $n$  things and  $M$  has  $m$  things, is  $n$  collections of  $m$  things – for each of the  $n$  objects  $a$  in  $N$ , it contains  $\langle a, b \rangle$ , for each of the  $m$  objects  $b$  in  $M$ . Likewise for exponentiation. Squaring is multiplying twice. Cubing is multiplying three times. Raising  $n$  to the  $m$  power is multiplying  $n$  by itself  $m$  times. Given the definition of multiplication, these become the following:

|   |   |
|---|---|
| Squaring: $K \times K$  | $(\kappa^2 = \text{card } K \times K)$          |
| Cubing: $K \times K \times K$   | $(\kappa^3 = \text{card } K \times K \times K)$ |
| Raising to the $n^{\text{th}}$ power: $K \times \dots [n \text{ times}] \dots \times K$ | $(\kappa^n = \text{card } K \times \dots)$      |

But what is this idea of the Cartesian product of  $K$  with itself  $n$  times, or even  $\lambda$  times, where  $\lambda$  is an infinite cardinal? That's the idea of the generalized Cartesian product. An  $L$ -tuple of members of  $K$  is a function from  $L$  into  $K$  (an  $L$ -sequence of members of  $K$ ), so the set  ${}^L K$  is the set of all  $L$ -tuples of members of  $K$ .

Similarly for multiplication and exponentiation.

These are well-defined definitions only if the resulting cardinalities don't turn on the sets  $K$  and  $L$  selected. Thus, we should prove:

**Theorem 6H** Assume that  $K_1 \approx K_2$  and  $L_1 \approx L_2$ . Then:

- (a) If  $K_1 \cap L_1 = K_2 \cap L_2 = \emptyset$ , then  $K_1 \cup L_1 \approx K_2 \cup L_2$
- (b)  $K_1 \times L_1 \approx K_2 \times L_2$
- (c)  ${}^{L_1} K_1 \approx {}^{L_2} K_2$

(Also, for addition, we need to be sure that we always *can* select disjoint sets of cardinality  $\kappa$  and  $\lambda$ . But we know that there always do exist sets  $K$  and  $L$  equinumerous to  $\kappa$  and  $\lambda$  -- namely,  $\kappa$  and  $\lambda$  themselves. And we can turn any  $K$  and  $L$  into disjoint sets equinumerous to  $\kappa$  and  $\lambda$  by taking  $K \times \{0\}$  and  $L \times \{1\}$ .)

*Proof of 6H*

(a). Since  $K_1 \approx K_2$ , some  $f$  maps  $K_1$  one-to-one onto  $K_2$ ; likewise, some  $g$  maps  $L_1$  one-to-one onto  $L_2$ . Now construct  $h: K_1 \cup L_1 \rightarrow K_2 \cup L_2$  thus:

$$\begin{aligned} h(x) &= f(x) & \text{if } x \in K_1 \\ h(x) &= g(x) & \text{if } x \in L_1 \end{aligned}$$

Since  $K_1 \cap K_2 = \emptyset$ , this is a function; since  $L_1 \cap L_2 = \emptyset$  and  $f$  and  $g$  are one-to-one,  $h$  is one-to-one. And since  $f$  and  $g$  are onto  $L_1$  and  $L_2$ ,  $h$  is onto  $L_1 \cup L_2$ . So  $K_1 \cup L_1 \approx K_2 \cup L_2$ .

(b). Again select our  $f$  and  $g$ . Define  $h: K_1 \times L_1 \rightarrow K_2 \times L_2$  thus:  $h(\langle k, l \rangle) = \langle f(k), g(l) \rangle$ , for any  $\langle k, l \rangle \in K_1 \times L_1$ .

(c). Again select our  $f$  and  $g$ . <Draw the picture of the sets, as on p. 140 of the book. > Consider any  $j \in {}^L_1 K_1$ . Which function in  ${}^L_2 K_2$  should we map  $j$  to? The function that behaves as follows: start with any  $l \in L_2$ , go back on  $f^{-1}$ , go down on  $j$ , go across on  $g$  into  $K_2$ . This function is:  $g \circ j \circ f^{-1}$ . So, for any  $j \in {}^L_1 K_1$ , define  $H(j) = g \circ j \circ f^{-1}$ . We must show that  $H$  is one-to-one. So suppose that  $j \neq j'$  (where each is in  ${}^L_1 K_1$ ). Then for some  $l \in L_1$ ,  $j(l) \neq j'(l)$ .  $H(j)$  and  $H(j')$  will differ in what they assign to  $f(l)$ :

$$\begin{aligned} H(j)(f(l)) &= g(j(f^{-1}(f(l)))) = g(j(l)) \\ H(j')(f(l)) &= g(j'(f^{-1}(f(l)))) = g(j'(l)) \end{aligned}$$

Since  $j(l) \neq j'(l)$  and  $g$  is one-to-one, these are different. So  $j$  and  $j'$  are different. We must also show that  $H$  is onto  ${}^L_2 K_2$ . Take any  $j \in {}^L_2 K_2$ .  $H$  maps the following function to  $j$ :  $g^{-1} \circ j \circ f$ .

Some examples: (where  $+$ ,  $\cdot$  and exponentiation express their cardinal versions):

$$n + \aleph_0 = \aleph_0 \quad n \cdot \aleph_0 = \aleph_0 \quad \aleph_0 + \aleph_0 = \aleph_0 \quad \aleph_0 \cdot \aleph_0 = \aleph_0$$

I won't show the first two rigorously (i.e., by induction); but e.g., take the first. Take a two-membered set  $\{a, b\}$ , that does not overlap with  $\omega$ . We can then map  $\{a, b\} \cup \omega$  one-one onto  $\omega$ , by mapping  $a$  to 0,  $b$  to 1, and  $n$  to  $n+2$ . Clearly we can do the same for a three-membered set.

Now take the second. Let's show that  $\{a, b\} \times \omega \approx \omega$ . Well, we can map the ordered pairs  $\langle a, n \rangle$  to 1, 3, 5, ..., and the ordered pairs  $\langle b, n \rangle$  to 2, 4, 6, .... Clearly this strategy generalizes to three-membered sets, etc.

Note: these first two examples show that the usual cancellation laws fail for

cardinal arithmetic. E.g.,  $2 + \aleph_0 \neq 3 + \aleph_0$ .

The third: do 2 hilbert hotels

The fourth: we showed earlier that  $\omega \times \omega \approx \omega$

For any A,  $\text{card } \wp A = 2^{\text{card } A}$

The definition of  $2^{\text{card } A}$  is  $\text{card } {}^A 2$ . And we showed earlier that  $\wp A \approx {}^A 2$ .

**Theorem 6I** For any cardinal numbers,  $\kappa$ ,  $\lambda$  and  $\mu$ :

1.  $\kappa + \lambda = \lambda + \kappa$     $\kappa \cdot \lambda = \lambda \cdot \kappa$
2.  $\kappa + (\lambda + \mu) = (\kappa + \lambda) + \mu$     $\kappa \cdot (\lambda \cdot \mu) = (\kappa \cdot \lambda) \cdot \mu$
3.  $\kappa \cdot (\lambda + \mu) = \kappa \cdot \lambda + \kappa \cdot \mu$
4.  $\kappa^{\lambda + \mu} = \kappa^\lambda \cdot \kappa^\mu$
5.  $(\kappa \cdot \lambda)^\mu = \kappa^\mu \cdot \lambda^\mu$
6.  $(\kappa^\lambda)^\mu = \kappa^{\lambda \cdot \mu}$

These are straightforward. I'll follow Enderton in doing just the last one. We must show that for any sets, K, L, and M,  ${}^{M({}^L K)} \approx {}^{L \times M} K$ . Consider any  $j \in {}^{M({}^L K)}$ .  $j$  is a function that maps any  $m \in M$  to a function from L into K. We need to map every such  $j$  to a member of  ${}^{L \times M} K$  – i.e., a function that maps ordered pairs  $\langle l, m \rangle$ , where  $l \in L$  and  $m \in M$ , to members of K. Let's define our mapping thus:

$G(j)$  = the function  $i$  such that  $i(\langle l, m \rangle) = j(m)(l)$ , for any  $m \in M$  and any  $l \in L$

$G$  is one-to-one: consider any  $j, j' \in {}^{M({}^L K)}$ , where  $j \neq j'$ . Then for some  $m \in M$ ,  $j(m) \neq j'(m)$ . This in turn means that for some  $l \in L$ ,  $j(m)(l) \neq j'(m)(l)$ . Thus, the function  $i$  such that  $i(\langle l, m \rangle) = j(m)(l)$ , for any  $m$  and  $l$ ,  $\neq$  the function  $i'$  such that  $i'(\langle l, m \rangle) = j'(m)(l)$ ; hence,  $G(j) \neq G(j')$ .

$G$  is onto  ${}^{L \times M} K$ : consider any  $i \in {}^{L \times M} K$ .  $G$  maps the following function to  $i$ : the function  $j$  such that  $j(m)$  = the function  $h$  such that  $h(l) = i(l, m)$ .

**Theorem 6J** Cardinal and  $\omega$ -addition, multiplication, and exponentiation agree over  $\omega$

Pf: I'll just do the proof for addition. The following statements are true (the operations in each case are the cardinal versions):

- (a1)  $\kappa+0=\kappa$
- (a2)  $\kappa+(\lambda+1)=(\kappa+\lambda)+1$

a1 is trivial; the second follows from 6K. Next, note the following mini lemma: if  $n$  is a finite cardinal, then  $n+1=n^+$  (+ here is cardinal addition). Pf:  $n$  and  $\{n\}$  do not overlap (we proved earlier that no natural number is a member of itself.) Moreover,  $n\approx n$  and  $1\approx\{n\}$ . Thus,  $n+1=\text{card}(n\cup\{n\})$ . But  $n\cup\{n\}=n^+$ , and  $\text{card } n^+ = n^+$ .

Now the proof goes by induction. Let's do it for addition. Let  $T=\{n\in\omega\mid \text{for all } m\in\omega, m+n=m+_{\omega}n\}$ . We'll show that  $T$  is inductive.

$0\in T$ : we must show that  $m+0=m+_{\omega}0$ . This follows from a1 and A1 (from chapter 4).

Now suppose that  $n\in T$ . We must show that  $m+n^+=m+_{\omega}n^+$ :

$$\begin{aligned} m+n^+ &= m+(n+1) && \text{(mini lemma)} \\ &= (m+n)+1 && \text{(a2)} \\ &= (m+n)^+ && \text{(mini lemma)} \\ &= m+_{\omega}n^+ && \text{(A2, chapter 4)} \end{aligned}$$

### E. Ordering cardinal numbers

**Definition**  $\kappa\leq\lambda$  iff for any  $K, L$ , of cardinality  $\kappa$  and  $\lambda$ ,  $K$  maps one-one *into*  $L$  (i.e., " $K\leq L$ ")

We need to be sure that whether  $\kappa\leq\lambda$  does not turn on the choice of  $K$  and  $L$ . So suppose that  $K, K'\approx\kappa$ , and  $L, L'\approx\lambda$ . Thus there is a one-one function  $f$  from  $\kappa$  onto  $K$ , and some one-one function  $f'$  from  $\kappa$  onto  $K'$ ; likewise there is a one-one function  $g$  from  $\lambda$  onto  $L$ , and some one-one function  $g'$  from  $\lambda$  onto  $L'$ . Now suppose that some  $h$  maps  $K$  one-to-one into  $L$ . Then the following function maps  $K'$  one-one into  $L'$  (draw it):  $g'\circ g^{-1}\circ h\circ f\circ f'^{-1}$ .

**Definition**  $\kappa<\lambda$  iff  $\kappa\leq\lambda$  and  $\kappa\not\approx\lambda$

(i.e., some  $K\approx\kappa$  is equinumerous with a subset of some  $L\approx\lambda$ , but is not also equinumerous to  $L$  itself.)

Examples:

If  $A\subseteq B$  then  $\text{card } A\leq\text{card } B$  (obvious – use the identity map.) Also, if  $\kappa\leq\lambda$ , then there are sets  $K$  and  $L$  of cardinality  $\kappa$  and  $\lambda$ , such that  $K\subseteq L$ . For take any  $L$  of cardinality  $\lambda$ .

By definition of  $\leq$ , there exists some one-one  $f:C \rightarrow L$ , from some  $C$  of cardinality  $\kappa$ . Just choose  $K$  to be  $\text{ran } f$ .

For any finite cardinal  $n$ ,  $n < \aleph_0$ . (every finite cardinal is a member of  $\omega$ ; but members of  $\omega$  are subsets of  $\omega$  (transitivity), so the identity map from  $n$  to itself is a map from  $n$  into  $\omega$ . But  $n$  can't map one-one onto  $\omega$ ; otherwise  $\omega$  would be finite (6D says it's infinite.)

Our ordering  $\leq$  coincides with  $\in$ -or-equal-to on  $\omega$ : suppose  $m \notin n$ ; then  $m \subset n$  (proved earlier); but then  $m \leq n$ . Other direction: suppose  $m \leq n$  – then  $m$  maps one-one onto some subset,  $A$ , of  $n$ . Suppose for reductio that  $m$  is not  $\in$   $n$ . Then by trichotomy,  $n \in m$ , and so  $n \subset m$ , and so  $A \subset m$ , violating the pigeonhole principle.

We need to verify that  $\leq$  deserves to be called an ordering. Here are some features it has.

- $\leq$  is reflexive and transitive on cardinals
- $\leq$  is anti-symmetric on cardinals: if  $\kappa \leq \lambda$  and  $\lambda \leq \kappa$  then  $\kappa = \lambda$
- $\leq$  is connected on cardinals: either  $\kappa \leq \lambda$  or  $\lambda \leq \kappa$

The first two are obvious. The last we'll prove in awhile. The third we now prove:

**Schröder-Bernstein theorem** if  $A \preceq B$  and  $B \preceq A$  then  $A \approx B$ . (And so, for cardinals, if  $\kappa \leq \lambda$  and  $\lambda \leq \kappa$  then  $\kappa = \lambda$ )

Pf. Let  $f:A \rightarrow B$  and  $g:B \rightarrow A$  be one-one functions. We must find a one-one function from  $A$  onto  $B$ . Define a series of sets by recursion (draw diagram from book):

$$\begin{aligned} C_0 &= A - \text{ran } g \\ C_{n+} &= g[f[C_n]] \end{aligned}$$

(i.e., bounce  $C_n$  down to  $B$  using  $f$ , then back to  $A$  using  $g$ .) Now define our  $h$  thus:

$$\begin{aligned} h(x) &= f(x) \text{ if } x \in C_n \text{ for some } n \\ &= g^{-1}(x) \text{ otherwise} \end{aligned}$$

We need to be sure that  $g^{-1}(x)$  exists in this case (since  $g$  isn't known to be onto  $A$ .) But if  $x \notin C_n$  for all  $n$ , then  $x \notin C_0$ ; but  $x \in A$ , so  $x \in \text{ran } g$ .

Let's show that  $h$  is one-to-one. Suppose that  $x \neq x'$ , where each is in  $A$ . If  $f(x)$  and  $f(x')$  are both calculated using the same clause of  $h$ 's definition, then they are distinct since  $f$  and  $g^{-1}$  are both one-to-one. The only remaining case (without loss of generality) is when  $x \in C_n$  for some  $n$  and so  $h(x) = f(x)$ ; and when  $x' \notin C_n$  for all  $n$ , and so  $h(x') = g^{-1}(x')$ .

But now, if  $h(x)=h(x')$  then  $f(x)=g^{-1}(x')$ ; so  $g(f(x)) = g(g^{-1}(x')) = x'$ ; so, since  $x \in C_n$ ,  $x' \in C_{n+}$ , which it isn't. So  $h(x) \neq h(x')$ .

Now we must show that  $h$  is onto  $B$ . Consider any  $b \in B$ . Define  $D_n$  as  $f[C_n]$  (i.e., the intermediate term in calculating  $C_{n+}$ .) If  $b \in D_n$  for some  $n$  then clearly  $b \in \text{ran } h$ . Suppose on the other hand that  $b \notin D_n$  for all  $n$ , and consider  $g(b)$ .  $g(b) \notin C_0$  since it's in  $\text{ran } g$ . Nor can  $g(b)$  be a member of  $C_{m+}$  for any  $m$ , since then  $b$  would be in  $D_m$  (remember that  $C_{m+} = g[D_m]$ , and  $g$  is one-to-one.) So  $g(b) \notin C_m$ , for all  $m$ . Thus,  $h(g(b)) = g^{-1}(g(b)) = b$ , and again,  $b \in \text{ran } h$ .

The S/B theorem is important because it lets us “squeeze” cardinal numbers. If we want to show that two sets have the same cardinality, we can just show that each can be mapped one-one into a subset of the other.

Example: Show that  $\mathbb{R} \approx \aleph^\omega$ . To do this, we'll show that  $\mathbb{R} \approx {}^\omega 2$ . We'll use the fact that each real has a decimal expansion in binary notation. First, let's map the open interval  $(0,1) \subseteq \mathbb{R}$  one-one into  ${}^\omega 2$ . For any real,  $x$ , in this interval, let  $H(x)$  be the function from  $\omega$  into  $2$  corresponding to  $x$ 's decimal expansion. (If  $x$  has two decimal expansions, then (fact) one is repeating; choose that one. Note: if we were happy to use the axiom of choice, we could use that to select one in each case.) This is one-to-one. And since  $\mathbb{R} \approx (0,1)$  (use  $\tan$ , or the geometric proof in the book), we have that  $\mathbb{R} \approx {}^\omega 2$ . Next let's map  ${}^\omega 2$  into  $\mathbb{R}$ . Here, simply send each such function  $j$  into the real  $0.d_1d_2\dots$  where  $d_i$  is  $j(i)$ . (Why is this one-to-one? What about repeating decimals? What we can do is this: if a function has another function representing the same real,  $r$ , send the nonrepeating function to  $r$  and the repeating one to  $r+1$ .) Thus, by S/B,  $\mathbb{R} \approx {}^\omega 2$ .

This next theorem shows that the operations of cardinal arithmetic have the right order-preserving properties:

**Theorem 6L** let  $\kappa$ ,  $\lambda$ , and  $\mu$  be cardinals:

- (a)  $\kappa \leq \lambda \Rightarrow \kappa + \mu \leq \lambda + \mu$
- (b)  $\kappa \leq \lambda \Rightarrow \kappa \cdot \mu \leq \lambda \cdot \mu$
- (c)  $\kappa \leq \lambda \Rightarrow \kappa^\mu \leq \lambda^\mu$
- (d)  $\kappa \leq \lambda \Rightarrow \mu^\kappa \leq \mu^\lambda$  (if not both  $\kappa$  and  $\mu$  are zero)

Pf. Suppose  $\kappa \leq \lambda$ . Let's make a bunch of choices at the start. Choose  $K$ ,  $L$ , and  $M$  to be sets of cardinality  $\kappa$ ,  $\lambda$ , and  $\mu$ , subject to the following constraints. (it will be clear that these choices are possible.) Choose  $K \subseteq L$ . And choose  $L \cap M = \emptyset$  (and so  $K \cap M = \emptyset$ ). Now (a) is easy:  $K \cup M$  has cardinality  $\kappa + \mu$ ;  $L \cup M$  has cardinality  $\lambda + \mu$ , and since

$K \cup M \subseteq L \cup M$ , the identity map establishes that  $K \cup M \preceq L \cup M$ . (b) and (c) are exactly similar.

As for (d), read the book.

#### F. Axiom of choice

We now need to say more about the axiom of choice. Here's something that makes it relevant here: we'd like to prove that  $\aleph_0$  is the smallest infinite cardinal. To do this, we'd have to show that for any  $A$ , if  $A$  is infinite then some one-one  $f$  maps  $\omega$  into  $A$ . Here's a natural way to define  $f$ :

$f(0) = \text{some member of } A$   
 $f(n^+) = \text{some member of } A - f[[n^+]]$  (i.e., anything in  $A$  other than  $f(0), f(1), \dots, f(n)$ .)

But this "some member" business means the axiom of choice.

What we're going to do is prove that various claims are equivalent (given the other ZF axioms); thus they can all be thought of as "versions of the axiom of choice".

**Theorem 6M** The following statements are equivalent:

- (1) For every relation, there exists a function  $F$  such that  $F \subseteq R$  and  $\text{dom } F = \text{dom } R$
- (2) The Cartesian product of nonempty sets is always nonempty
- (3) For any set  $A$ , there exists a function  $F$  (a "choice function for  $A$ ") whose domain is the set of nonempty subsets of  $A$ , and which is such that  $F(B) \in B$ , for every nonempty subset  $B$  of  $A$
- (4) Let  $A$  be a set of disjoint nonempty sets. Then there exists a set  $C$  containing exactly one member of each member of  $A$ .
- (5) Cardinal comparability: for any sets  $K$  and  $L$ , either  $K \preceq L$  or  $L \preceq K$ . (so, for any cardinals  $\kappa$  and  $\lambda$ , either  $\kappa \leq \lambda$  or  $\lambda \leq \kappa$ .)
- (6) Zorn's lemma. Call  $B$  a *chain* iff for every  $C, D \in B$ , either  $C \subseteq D$  or  $D \subseteq C$ . (Thus, its members are arranged in a line by subset.) Let  $A$  be a set such that for each chain  $B \subseteq A$ ,  $\bigcup B \in A$ . Then  $A$  has a maximal element  $M$ : for no  $S \in A$  other than  $M$  itself is  $M \subseteq S$ .

Pf: We proved (1)  $\rightarrow$  (2) in class.

(2)  $\rightarrow$  (4): Define  $H: A \rightarrow A$  thus:  $H(a) = a$ . Each  $H(a)$  is nonempty. Thus,  $H$  has a Cartesian product, a function  $F$  with  $\text{dom } F = A$  such that  $F(a) \in H(a) = a$ . Let  $C = \text{ran } F$ . Clearly,  $C$  contains a member of each member of  $A$ , since for any  $a \in A$ ,  $F(a) \in H(a) = a$ . And  $C$  contains no more than one member of any member of  $a$ . For suppose that  $x, y$  are

each in  $C$  and also in some  $a \in A$ . Since  $x, y$  are each in  $C$ ,  $x = F(b)$  and  $y = F(c)$  for some  $b, c \in A$ . So  $x \in b$  and  $y \in c$ . But since  $x \in b$  and  $x \in a$ ,  $b = a$  (since no two members of  $A$  overlap); likewise, since  $y \in c$  and  $y \in a$ ,  $a = c$ . Thus,  $b = c$ ; and so  $x = y$ .

(4)  $\rightarrow$  (3): Define this set:  $D = \{ \{B\} \times B \mid B \subseteq A \text{ and } B \neq \emptyset \}$ . (This in effect replaces each nonempty  $B \subseteq A$  with the set of ordered pairs  $\langle B, b \rangle$ , where  $b \in B$ .) Each member of  $D$  is nonempty; and no two members of  $D$  overlap (since any two members of  $D$  are sets of ordered pairs with distinct left members.) Now apply (4); that means there exists a set,  $F$ , containing exactly one member of each member of  $D$ .  $F$  is our desired function. It's a function: each member of  $D$  was a set of ordered pairs with different first members, so there can't be ordered pairs  $\langle x, y \rangle$  and  $\langle x, z \rangle$  in  $F$  when  $y \neq z$ . And since for each nonempty  $B \subseteq A$ ,  $D$  contained a set of ordered pairs of the form  $\langle B, b \rangle$ , where  $b \in B$ . It follows that  $\text{dom } F = \text{the set of such } B\text{'s}$ , and also that  $F(B) \in B$ .

(3)  $\rightarrow$  (1): Let  $R$  be any relation. Let  $H$  be a choice function for  $\text{ran } R$  as described in (3). Now define the desired function  $F$  as follows: for any  $a \in \text{dom } R$ ,  $F(a) = H(\{y \mid aRy\})$ .  $F$  is a well-defined function, since any  $\{y \mid aRy\}$  where  $a \in \text{dom } R$  is a nonempty subset of  $\text{ran } R$ . And note that  $aRF(a)$ . So  $F \subseteq R$ .

(6)  $\rightarrow$  (1): Let  $A$  be the set of functions that are subsets of  $R$ :  $A = \{f \mid f \subseteq R\}$ . We want to apply Zorn's lemma to  $A$ ; so we need to show that  $A$  is closed under unions of chains.

So let  $B \subseteq A$  be a chain; we must show that  $\bigcup B$  is a function that is a subset of  $R$ .

Clearly,  $\bigcup B \subseteq R$ , since every member of  $B$  is a subset of  $R$ . Now, let  $\langle x, y \rangle, \langle x, z \rangle \in \bigcup B$ .

So  $\langle x, y \rangle \in C$  and  $\langle x, z \rangle \in D$ , for some  $C, D \in B$ . Either  $C \subseteq D$  or  $D \subseteq C$  since  $B$  is a chain; suppose the former without loss of generality. Then  $\langle x, y \rangle$  and  $\langle x, z \rangle$  are both in  $D$ . But  $D \in B \subseteq A$ , so  $D \in A$ , so  $D$  is a function, so  $y = z$ .

Zorn's lemma now tells us that  $A$  has a maximal element  $M$ :  $M \in A$ , and for any other  $N \in A$ , if  $M \subseteq N$  then  $M = N$ . Since  $M \in A$ ,  $M$  is a function and  $M \subseteq R$ . We must show that  $\text{dom } M = \text{dom } R$ . Suppose otherwise – suppose some  $a \in \text{dom } R$  but  $a \notin \text{dom } M$ . Then for some  $b$ ,  $\langle a, b \rangle \in R$ . But then  $M' = M \cup \{\langle a, b \rangle\}$  is also a member of  $A$ , contradicting  $M$ 's maximality.

(6)  $\rightarrow$  (5): Consider any  $K, L$ . Let  $A = \{f \mid f \text{ is a one-one function \& } \text{dom } f \subseteq K \text{ and } \text{ran } f \subseteq L\}$ . First we show that  $A$  is closed under unions of chains. Let  $B \subseteq A$  be a chain; we must show that  $\bigcup B \in A$ . As before,  $\bigcup B$  is a function. Moreover, essentially the same argument shows that it's one-to-one. And since each of  $A$ 's members have a  $\text{dom} \subseteq K$  and a  $\text{ran} \subseteq L$ , the same is true for  $\bigcup B$ .

Zorn's lemma thus informs us of a maximal  $M \in A$ . We must show that either  $\text{dom } M = K$  or  $\text{ran } M = L$ . Suppose neither. Then for some  $k \in K$  and some  $l \in L$ ,  $k \notin \text{dom } M$  and  $l \notin \text{ran } M$ . But then  $M \cup \{ \langle k, l \rangle \}$  would be  $\in A$ , violating  $M$ 's maximality. So either  $\text{dom } M = K$  or  $\text{ran } M = L$ ; and so either  $K \preceq L$  or  $L \preceq K$ .

The rest of the proof of  $6M$  will need to wait for a new axiom. So at this point, we have not really established (5) – cardinal comparability.

Apart from the fact that it implies other forms of the axiom of choice, why is Zorn's lemma "choicy"? Enderton sketches the following "plausibility argument" for Zorn's lemma, which is basically an argument for Zorn's lemma by transfinite induction; as such, we can't do the argument yet since we haven't developed transfinite induction. Here's the argument: let  $A$  be as described; we must construct a maximal element of  $A$ . We know that  $\emptyset \in A$  since  $\emptyset$  is a chain and  $\emptyset \subseteq A$ , and  $\bigcup \emptyset = \emptyset \in A$ . Is  $\emptyset$  a maximal element of  $A$ ? If so then we're done; otherwise, choose (there's the choicy bit) some larger superset,  $a$ , of  $\emptyset$  in  $A$ :  $\emptyset \subset a \in A$ . If  $a$  is maximal then we're done; otherwise choose another  $b$  such that  $a \subset b \in A$ . Keep doing this. If you have done it infinitely many times and still haven't reached a maximal element, then take the union of what you've got so far and see if that's now a maximal element of  $A$  (the sets we've constructed so far form a chain, so their union is in  $A$ ). Continue this (that's the transfinite induction part) and eventually we'll reach a maximal element of  $A$ .

Example: we can use the axiom of choice to prove the following highly intuitive statement:

If there exists a function from  $A$  onto  $B$  then  $B \preceq A$

Let  $f$  be a function from  $A$  onto  $B$ . We now need a one-one function from  $B$  into  $A$ . Intuitively, we just need to delete all the ordered pairs in  $f$  that make  $f$  not one-to-one. So consider  $f^{-1}$ . That's a relation with domain  $B$  and range  $A$ . By the first form of the axiom of choice, there's some function  $g \subseteq f^{-1}$  with domain  $B$ . And its range is a subset of  $A$  (since  $f^{-1}$ 's range was  $A$ .) Moreover, it's one-to-one: suppose  $\langle b, a \rangle$  and  $\langle b', a \rangle \in g$ . Then  $\langle a, b \rangle$  and  $\langle a, b' \rangle \in f$ , so  $b = b'$  since  $f$  is a function.

We can also prove the converse. If  $B \preceq A$ , then there's a one-one function,  $f$ , from  $B$  into  $A$ . Since  $f$  is one-one,  $f^{-1}$  is a function whose domain is a subset of  $A$  and whose range is all of  $B$ . Now, where  $b_0$  is any member of  $B$ , let  $g = f^{-1}$  plus all pairs  $\langle a, b_0 \rangle$  where  $a \notin \text{dom } f^{-1}$ .  $g$  is a function from  $A$  onto  $B$ .

Thus we've established:

$B \preceq A$  iff there exists some function from  $A$  onto  $B$

It is often convenient to use this formulation of  $A \preceq B$  rather than the official one.

OK, now that we have all the forms of the axiom of choice at our disposal, let's go back and fix up that proof that  $\aleph_0$  is the smallest cardinal:

**Theorem 6N** (a) for any infinite set  $A$ ,  $\omega \preceq A$   
 (b) for any infinite cardinal  $\kappa$ ,  $\aleph_0 \leq \kappa$

Pf: (b) follows immediately from (a), so we just must prove (a). So we need a one-one function from  $\omega$  into  $A$ . We're going to repeatedly choose members of  $A$ . Let's use the axiom of choice, form III. That gives us a choice function  $F$ , which assigns to each nonempty subset  $B$  of  $A$  some member of  $B$ . We now want to recursively define a function from  $\omega$  into  $A$  utilizing  $F$ . Here's a first try. We know that  $A$  is nonempty since it's infinite. So there's some  $a \in A$ . So here's our definition:

$$\begin{aligned} g(0) &= a \\ g(n^+) &= F(A - g[n^+]) \end{aligned}$$

Note that on this definition,  $A - g[n^+]$  is just the members of  $A$  other than  $g(0), \dots, g(n)$ ; so the  $F$  function just chooses some member of this subset of  $A$ . The problem with this definition is that our recursion theorem doesn't guarantee that these equations pick out a function. We're only allowed to look at  $g(n)$  in defining  $g(n^+)$ ; we're not allowed to look at  $g[n^+]$ .

So here's a trick. We define an intermediate function,  $h$ , which assigns to integers larger and larger finite subsets of  $A$ . For each new value of  $h$ , we'll add a *new* member of  $A$  to the set:

$$\begin{aligned} h(0) &= \emptyset \\ h(n^+) &= h(n) \cup F(A - h(n)) \end{aligned}$$

This is a well-defined function because of the following.  $h$  is being defined as a function from  $\omega$  into the set of finite subsets of  $A$ . The second recursion equation defines  $h(n^+)$  as a certain function  $G(x) = x \cup F(A - x)$  from the set of finite subsets of  $A$  into the set of finite subsets of  $A$ , as called for by the recursion theorem. And this in turn is a well-defined function because if  $x$  is a finite subset of  $A$  then  $A - x$  is nonempty (since  $A$  is infinite) and so is in  $\text{dom } F$ .

OK, we now define our desired  $g$  function as, informally,  $g(n) =$  the thing we added to  $h(n)$  to get  $h(n^+)$ :

$$g(n) = F(A-h(n))$$

To show:  $g$  is a one-one function from  $\omega$  into  $A$ . Clearly  $g$  is into  $A$  since  $F$  is a choice function and  $A-h(n) \subseteq A$ . And  $\text{dom } g = \omega$  since  $\text{dom } h = \omega$  (which in turn follows from the recursion theorem.) Finally,  $g$  is one-to-one: suppose that  $g(m)=g(n)$ . So:

$$g(m) = F(A-h(m)) = F(A-h(n)) = g(n)$$

and since  $F$  is a choice function,  $g(m) \in A-h(n)$  and  $g(n) \in A-h(m)$ , and so:

$$g(m) \notin h(n) \text{ and } g(n) \notin h(m).$$

Now suppose for reductio that  $m \neq n$ . Then either  $m < n$  or  $n < m$ ; suppose the former (without loss of generality). Now, a quick induction shows that for each  $k, k'$ , if  $k < k'$  then  $g(k) \in h(k')$  (the  $h$ 's are cumulative, and we arrive at  $h(n^+)$  by adding  $g(n)$  to  $h(n)$ ). Contradiction.

We can prove this: Any infinite subset of  $\omega$  is equinumerous to  $\omega$

Pf:  $\aleph_1$  shows that if  $A$  is an infinite subset of  $\omega$  then there's a one-one function from  $\omega$  into  $A$ . But there's obviously a one-one function from  $A$  into  $\omega$ : identity function. By S/B,  $A \approx \omega$ . (we wouldn't really need  $\aleph_1$  here, since the least number principle gives us a way of avoiding choice: duplicate its proof but substitute the least number principle for the axiom of choice.)

We can also prove this (though I won't):

$$\kappa < \aleph_0 \text{ iff } \kappa \text{ is finite}$$

Also, (and this was Dedekind's definition of 'infinite'):

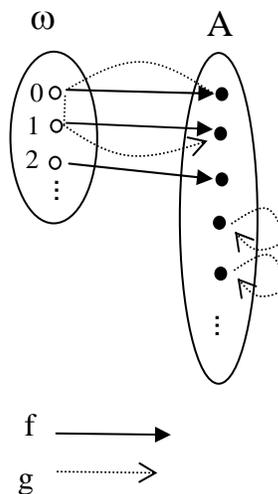
**Corollary 6P** A set is infinite iff it is equinumerous to a proper subset of itself

Pf:  $\aleph_1$  already established the right-to-left direction. Now for the other direction, suppose that  $A$  is infinite. We must construct a one-one function,  $g$ , from  $A$  onto a proper subset of  $A$ . Since  $A$  is infinite, by  $\aleph_1$  there is some one-one function  $f: \omega \rightarrow A$ . We define  $g$  thus:

$$g(a) = a \text{ if } a \notin \text{ran } f$$

$$g(a) = f(f^{-1}(a)^+) \text{ if } a \in \text{ran } f$$

$g$  is one-to-one, and  $f(0) \notin \text{ran } g$ .



### G. Countable sets

**Definition**  $A$  is *countable* iff  $A \leq \omega$  (i.e., iff  $\text{card } A \leq \aleph_0$ )

Equivalently:  $A$  is countable iff  $A$  is finite or  $\text{card } A = \aleph_0$   
iff there is some function from  $\omega$  onto  $A$

Facts: the union of two countable sets is countable ( $\aleph_0 + \aleph_0 = \aleph_0$ ); a subset of a countable set is countable; the cartesian product of a pair of countable sets is countable ( $\aleph_0 \cdot \aleph_0 = \aleph_0$ );  $\wp A$  is not countable if  $A$  is infinite ( $\text{card } \wp A = 2^{\text{card } A} > \text{card } A$ ; if  $A$  is infinite then  $\text{card } A \leq \aleph_0$  ( $6\aleph$ ); so  $\text{card } \wp A > \aleph_0$ .)

**Theorem 6Q** A countable union of countable sets is countable

Pf: Let  $A$  be a countable set of countable sets; it's intuitively obvious that we can define a one-one function  $f$  from  $\omega$  into  $\bigcup A$  – just go through the grid of  $A$ , skipping repeats where necessary. But here's a more careful proof. Since  $A$  is countable,  $A \leq \omega$ ; and so  $\omega$  can be mapped onto  $A$  (by a result proved earlier), so there exists some function  $G$  (not necessarily one-to-one) from  $\omega$  onto  $A$ . Thus, we can think of  $A$  as being  $\{G(0), G(1), \dots\}$ , where there may be repeats on this list. First, set aside a couple special cases, the case where  $A$  is empty (then obviously  $\bigcup A$  is countable), and the case where  $\emptyset \in A$  (it doesn't affect  $\bigcup A$ ). Thus, each  $G(m)$  is a nonempty countable set, and  $\bigcup A =$

$$\bigcup_{m \in \omega} G(m).$$

Let's now use the axiom of choice to construct a one-one function from  $\omega \times \omega$  onto  $\bigcup_{m \in \omega} G(m)$ . Since each  $G(m)$  is countable, for each  $G(m)$ , there exists a function from  $\omega$  onto  $G(m)$ . Define  $H(m) = \{g \mid g \text{ is a function from } \omega \text{ onto } G(m)\}$ . Since each  $G(m)$  is nonempty, each  $H(m)$  is nonempty. By the axiom of choice, second form, there exists a function  $F$  with domain  $\omega$  such that for each  $m$ ,  $F(m) \in H(m)$ . ( $F$  is the Cartesian product of  $H$ .) So what we've got now is this: for each  $m$ , our function  $F$  chooses a function from  $\omega$  onto  $G(m)$ , namely,  $F(m)$ . Now, we can map the members of  $\omega \times \omega$  onto  $\bigcup A$  thus:  $f(m,n) = F(m)(n)$ . For every member,  $x$ , of  $\bigcup A$  is in some  $G(m)$  (since  $G$  is onto  $\bigcup A$ ), and  $F(m)$  is onto  $G(m)$ , and so for some  $n$ ,  $F(m)(n) = x$ .

Thus,  $\bigcup A \preceq \omega \times \omega$ . But  $\omega \times \omega \approx \omega$  (shown earlier). Thus,  $\bigcup A \preceq \omega$ , and so it's countable.

#### H. Arithmetic of infinite cardinals

We're going to prove a theorem that trivializes the addition and multiplication of infinite cardinals, saying that the sum or product of infinite cardinals is just the maximum of the two. To do that, we'll need this lemma:

**Lemma 6R:** for any infinite cardinal  $\kappa$ ,  $\kappa \cdot \kappa = \kappa$

Pf: Let  $B$  be any set of cardinality  $\kappa$ . Let

$$K = \{f \mid f = \emptyset \text{ or for some infinite } A \subseteq B, f \text{ is a one-one function from } A \times A \text{ onto } A\}$$

The proof is going to be a little indirect. We're going to use Zorn's lemma on  $K$  (that's why we included  $\emptyset$  in  $K$ ). So let's first show that  $K$  is closed under unions of chains. Let  $C$  be a chain that's a subset of  $K$ . We must show that  $\bigcup C \in K$ . Special case:  $C = \emptyset$ ; in that case  $\bigcup C = \emptyset \in K$ . With that special case aside, we may assume that  $C$  is nonempty, in which case  $C$  is a set of one-one functions as described above. Then we know that  $\bigcup C$  itself is a one-one function, by reasoning like that in our earlier proofs using Zorn's lemma. To show that  $\bigcup C \in K$ , we must now find some infinite  $A \subseteq B$  such that  $\text{dom } \bigcup C$  is  $A \times A$  and its range is  $A$ . Let  $A = \bigcup \{\text{ran } f \mid f \in C\}$ .

$$\text{ran } \bigcup C = A: \text{ By exercise 8, chapter 3, } \bigcup \{\text{ran } f \mid f \in C\} = \text{ran } \bigcup C$$

A is infinite: since C is nonempty, it has at least one f as a member with an infinite range.

$\text{Dom } \bigcup C = A \times A$ : suppose  $x \in \text{dom } \bigcup C$ . So  $\langle x, a \rangle \in \bigcup C$ , for some  $a \in A$ . So for some  $f \in C$ ,  $f(x) = a$ . So for some  $A' \subseteq B$ ,  $A' = \text{ran } f$ , and  $x = \langle b, c \rangle$ , where  $b, c \in A'$ . But  $A' \subseteq A$ , so  $b, c \in A$ ; i.e.,  $x \in A \times A$ . **Thus,  $\text{dom } \bigcup C \subseteq A \times A$ .** Conversely, suppose  $\langle b, c \rangle \in A \times A$ , so  $b, c \in A$ . So for some  $f, f' \in C$ ,  $b \in \text{ran } f$  and  $c \in \text{ran } f'$ . Since C is a chain, either  $f \subseteq f'$  or  $f' \subseteq f$ ; suppose the latter (without loss of generality). Thus,  $b, c$  are both in  $\text{ran } f$ . So  $\langle b, c \rangle \in \text{ran } f \times \text{ran } f$ , and so  $\langle b, c \rangle \in \text{dom } f$ . So for some  $d \in \text{ran } f$ ,  $\langle \langle b, c \rangle, d \rangle \in f$ ; so  $\langle \langle b, c \rangle, d \rangle \in \bigcup C$ , so  $\langle b, c \rangle \in \text{dom } \bigcup C$ . Hence,  $A \times A \subseteq \text{dom } \bigcup C$

We can now use Zorn's lemma to infer the existence of a maximal element,  $f_0$ , of K.

We need to verify that  $f_0 \neq \emptyset$ . Since B is infinite,  $\omega \preccurlyeq B$ , so for some  $A \approx \omega$ ,  $A \subseteq B$ . But  $\omega \times \omega \approx \omega$ , so  $A \times A \approx A$ . Hence there's a one-one function,  $f$ , from  $A \times A$  onto  $A$ . Since A is infinite ( $A \approx \omega$ ),  $f \in K$ . Since  $\emptyset \subseteq f$ ,  $\emptyset \neq f_0$ . Thus,  $f_0$  is a one-one function from  $A_0 \times A_0$  onto  $A_0$ , for some infinite  $A_0 \subseteq B$ .

Let  $\lambda = \text{card } A_0$ . So  $\lambda$  is infinite, and  $\lambda \cdot \lambda = \lambda$ . Our strategy will be to show that  $\lambda = \kappa$  by first showing that  $\text{card } (B - A_0) < \lambda$ .

Take the second first: suppose for reductio that  $\lambda \leq \text{card } (B - A_0)$ . That means that  $A_0 \preccurlyeq (B - A_0)$ , and so  $A_0 \approx D$ , for some  $D \subseteq (B - A_0)$ . This D has cardinality  $\lambda$ . We're going to use D to show that  $f_0$  isn't maximal after all, by constructing a one-one function from  $A_0 \cup D \times A_0 \cup D$  onto  $A_0 \cup D$  of which  $f_0$  is a proper subset. Let's split  $A_0 \cup D \times A_0 \cup D$  into four nonoverlapping (since  $D \subseteq B - A_0$ ,  $D \cap A_0 = \emptyset$ ) parts as in Enderton's diagram:

$$A_0 \cup D \times A_0 \cup D = A_0 \times A_0 \cup A_0 \times D \cup D \times A_0 \cup D \times D$$

Now, we already know that  $f_0$  maps  $A_0 \times A_0$  onto  $A_0$ . So if we can map the remaining three bits one-one onto D, we're done. But we can. Since  $\text{card } A_0 = \text{card } D = \lambda$ , each has  $\text{card } \lambda \cdot \lambda$ , which is just  $\lambda$ . So their union has cardinality  $\lambda + \lambda + \lambda = 3 \cdot \lambda \leq \lambda \cdot \lambda$  (order preservation) =  $\lambda$ . But obviously,  $\lambda \leq \text{card } (\text{their union})$ . So  $\text{card } (\text{their union}) = \lambda$ . So there's a one-one function  $g$  from their union onto D. Thus,  $f_0 \cup g$  maps  $A_0 \times A_0 \cup A_0 \times D \cup D \times A_0 \cup D \times D (= A_0 \cup D \times A_0 \cup D)$  one-one onto  $A_0 \cup D$ .  $f_0 \cup g$  violates the maximality of  $f_0$ :  $f_0 \subsetneq f_0 \cup g$ , but  $f_0 \cup g \in K$ .

Hence,  $\text{card}(B-A_0) < \lambda$ . Now let's use S/B to show that  $\kappa=\lambda$ :

$$\begin{aligned}\kappa \leq \lambda: \quad \kappa &= \text{card } B = \text{card}(A_0 \cup B-A_0) \leq \lambda + \lambda = 2 \cdot \lambda \leq \lambda \cdot \lambda = \lambda \\ \lambda \leq \kappa: \quad A_0 &\subseteq B\end{aligned}$$

So  $\kappa=\lambda$ , and hence  $\kappa \cdot \kappa = \kappa$ .

Now we have this:

**Absorption law for infinite cardinals:** let  $\lambda$  and  $\kappa$  be infinite cardinals, the larger of which is infinite, and the smaller of which is nonzero. Then  $\lambda + \kappa = \lambda \cdot \kappa = \max(\lambda, \kappa)$ .

Pf: Suppose (without loss of generality) that  $\lambda \leq \kappa$ .

$$\lambda + \kappa \leq \kappa + \kappa = 2 \cdot \kappa \leq \kappa \cdot \kappa = \kappa. \text{ But obviously } \kappa \leq \lambda + \kappa. \text{ So } \lambda + \kappa = \kappa.$$

$$\lambda \cdot \kappa \leq \kappa \cdot \kappa = \kappa. \text{ But obviously } \kappa \leq \lambda \cdot \kappa. \text{ So } \lambda \cdot \kappa = \kappa.$$

Aside: we've used facts like  $\kappa + \kappa = 2 \cdot \kappa$ ,  $\kappa + \kappa + \kappa = 3 \cdot \kappa$ , etc. Why are those true? E.g., take the first. 2 has two members; so take any two-membered set,  $A = \{a, b\}$ ; then  $\{a, b\} \times K = \{a\} \times K \cup \{b\} \times K$ ; the sets don't overlap (since  $a \neq b$ ), and each has the same cardinality as  $\kappa$ .

## I. Continuum hypothesis, independence results

**Continuum hypothesis:** there is no cardinal number between  $\aleph_0$  and  $\text{card } \mathbb{R}$

**Generalized continuum hypothesis:** for each  $\kappa$ , there is no cardinal number between  $\kappa$  and  $2^\kappa$

These were outstanding questions for a long time. Then Godel proved in 1939 that the negation of the CH could not be proved from the axioms (if they're consistent). And Cohen proved in 1963 that the CH could not be proved from the axioms (if they're consistent).

Point 1: these are facts about logic, about the language of set theory. (though, one uses set theory itself to prove them.)

Point 2: there's always the qualification "if the axioms are consistent". Let me explain (this will have to be rough.) Suppose we're trying to show that certain axioms,  $A$ , of set theory do not imply a certain sentence,  $S$ . Now, if axioms  $A$  were in fact inconsistent, then *every* sentence would follow from  $A$ , and so  $A$  would imply  $S$  after all. So showing that  $A$  does not imply  $S$  also shows that axioms  $A$  are consistent.

Now, to prove that  $A$  doesn't imply  $S$ , one needs to make some assumptions of course. And the assumptions in fact are set theoretic in nature: one assumes that certain axioms of set theory, call them  $P$ , are true, and then one proves the result in question. For short, the proof takes place "within" this portion  $P$  of set theory. And in fact, this portion  $P$  contains the very axioms  $A$  that we're trying to show don't imply  $S$ , and so are consistent. (In fact, given certain Godelian results,  $P$  has to be "greater" than  $A$ .) Thus, while we do in fact have a proof that axioms  $A$  are consistent, this is pretty unimpressive, since we had to assume that those very axioms were *true* in order to do the proof. (Obviously, one can prove *any* assumptions to be consistent if one is willing to assume that they are true. Every true set of sentences is ipso facto consistent.)

Might one be able to prove that axioms,  $A$ , are consistent assuming only a more minimal basis, and thus not by assuming that they're all true? Godel stands in the way. He showed that provided you have a minimally strong theory (can embed a certain portion of arithmetic), then there can be no proof of the theory's consistency within the theory itself. Any realistic attempt to show that a significant fragment of set theory is consistent will have to use a bit of set theory,  $B$ , that's sufficiently strong. But the axioms  $A$  that we're interested in include  $B$ . So if you could show using  $B$  that  $A$  is consistent, you could show using  $B$  that  $B$  is consistent, in violation of Godel's result.

So: we can't really *prove* the axioms of set theory to be consistent, except by assuming something (their truth) that in essence amounts to assuming they're consistent. And if the axioms are *not* consistent, then they can't be true, in which case the independence proofs don't go through. This is acknowledged up front by qualifying the proofs in question thus: *if the axioms are consistent*. (Wouldn't it be better to make the qualification: "if the axioms are *true* then..."?)

Point 3: Enderton alludes to the question of whether there's a fact of the matter as to whether the continuum hypothesis is correct. How could there not be? One way: if there are two kinds of sets,  $sets_1$  and  $sets_2$ . Of one, the CH is correct; of the other, it's not. This is naturally connected with a philosophy of mathematics according to which, any (consistent) axioms are true of *some* objects.

Point 4: We can of course explore the addition of CH as an axiom to the other axioms of ZF. But generally we don't add an axiom unless it seems to count as a part of our intuitive conception of set. CH isn't like that. (of course, we may one day find an axiom that *does* correspond to our intuitive conception, and which implies CH.)

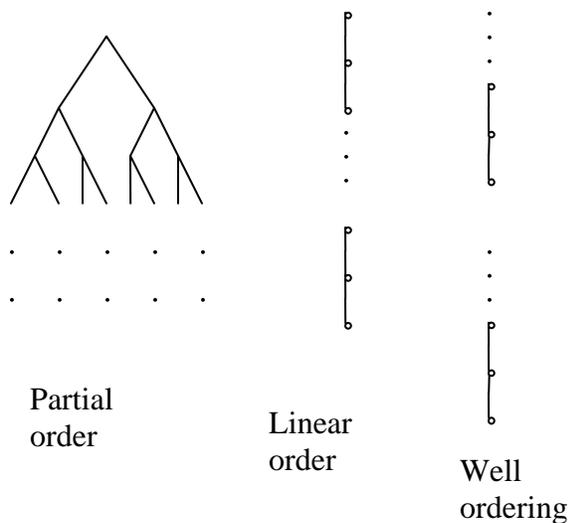
Point 5: axiom of choice: this is also shown by Godel and Cohen to be independent of the other axioms of ZF. But this does seem to accord with our conception of a set, so it gets added as an axiom.

## VII. Orderings and Ordinals

We want to introduce the concepts that will ultimate allow us to do transfinite induction and recursion. These are analogs of what we've already got for  $\omega$ . This stuff is at the very core of set theory.

### A. Partial vs. linear vs well orderings

Orderings are certain kinds of relations on a set. Here are a couple pictures:



All of the orderings we're considering are transitive relations. In the first case, there are "ties" allowed in the ordering. In the second there are no ties allowed; the field of the ordering is arranged into a line. The third case has the special property that every set in the field of the relation has a least element in the ordering. Those are the kind that we're going to focus on.

There are two ways we can set out orderings. We can use a reflexive relation ("weak" ordering) or we can use an irreflexive relation ("strict" ordering). If you do the latter you can recover a reflexive ordering by defining it as the disjunction of the irreflexive relation and identity. If you do the former then you can recover the irreflexive relation as the conjunction of the reflexive

relation and nonidentity. It doesn't matter which approach one takes; we're going to take the irreflexive approach.

## B. Well orderings

OK, here's the official definition:

**Linear order on A:** a transitive binary relation on A that satisfies trichotomy – i.e., for every  $x, y \in A$ , exactly one of the following holds:  $xRy, x=y, yRx$

**Well order on A:** a linear ordering on A such that every nonempty subset B of A has a least element – i.e., some  $b_0 \in B$  such that for all  $b \in B, b_0 \leq b$ .

Note that if  $<$  is a well order on A, then there is a “first” member of A,  $a_0$ .  $a_0$  is simply the least member of A under the ordering  $<$ . Further, for each  $a \in A$ , there is always a “next” member of A. For, define **segments**: Let  $<$  be any ordering on A; let  $t \in A$ . Then:

$$\text{seg } t = \{x | x < t\}$$

I.e., the “initial segment” of A up to but not including t. Now, for any  $a \in A$ , consider  $A - (\text{seg } a \cup \{a\})$ . This is the set of all members of A that are greater than a. This must have a least element, b, which is the “next” element of A after a.

A **structure** is an ordered pair  $\langle A, R \rangle$ , where A is a set and R is a binary relation on A.

(We can speak of linearly ordered structures, well-ordered structures, etc.)

Note that  $\langle \omega, \in_\omega \rangle$  is a well-ordered structure – this follows from the least-number principle proved earlier.

## C. Transfinite induction

We now prove the principle of transfinite induction. (This is a big part of why well-orderings are important.)

**Transfinite induction principle:** If  $<$  is a well ordering on A,  $B \subseteq A$ , and:

$$\text{for every } t \in A, \text{ if } \text{seg } t \subseteq B \text{ then } t \in B$$

then  $B=A$

This principle lets us prove facts about the entirety of a well ordered set,  $A$ , as follows. Simply let  $B$  be the subset of  $A$  whose members all have the desired property. Establish that  $B$  is *<-inductive*, in that whenever every earlier (under  $<$ ) than  $t$  member is in  $B$ , then  $t$  itself is in  $B$ . Conclude by transfinite induction that  $B = A$ , and so every member of  $A$  has the desired property.

Pf: Suppose for reductio that  $B \neq A$ ; so  $A-B$  is not null. Since  $<$  is a well-ordering, for some  $a_0 \in A-B$ , every  $a \in A-B$  is such that  $a_0 \leq a$ . That means that for each  $a \in A$ , if  $a < a_0$  then  $a \in B$ . Thus,  $\text{seg } a_0 \subseteq B$ , so  $a_0 \in B$ ; contradiction.

#### D. Transfinite recursion

We want to develop a method for defining functions recursively when the domain is any well-ordered set, not just  $\omega$ . Let  $<$  be a well-ordering on  $A$ . The basic idea is to provide a *rule* for arriving at  $F(t)$  as a function of *all* of the previous values  $F(a)$  for all  $a < t$ . That is, the rule is allowed to “look at” all the previous values of  $F$  – i.e.,  $F \upharpoonright \text{seg } t$ . Think of  $F \upharpoonright \text{seg } t$  as a sequence of values of  $F$  – for all the arguments  $< t$ . (Doesn’t there need to be a rule for  $a_0$ , the *first* member of  $A$ ? Well, since the rule tells us what  $F(t)$  is as a function of  $F \upharpoonright \text{seg } t$ , for *all*  $t \in A$ , it must do so for  $a_0$ . But  $\text{seg } a_0$  is  $\emptyset$ , and so the rule must in essence tell us what  $F(a_0)$  is, without being told anything at all.)

The notion of a “rule” can be understood in two ways. On one way, it is a function; on the other, it is a formula,  $\gamma(x,y)$ , with  $x$  and  $y$  free, with the property that for each  $x$  there is a unique  $y$  such that  $\gamma(x,y)$ . Such a  $\gamma$  is “functional”, and can be used in place of a genuine function, in cases where our rule is “too big to be a set”, and so cannot be a function.

Function route. Suppose we’re defining a function  $F$  from  $A$  into  $B$ , and we have a well ordering  $<$  on  $A$ . We’re doing to define the function by recursion, so we want a function,  $G$ , that gives us the next value of  $F$  as a function of all earlier values of  $F$ .  $G$  must be defined on every  $F \upharpoonright \text{seg } t$ . So  $G$ ’s domain will need to be the set of all initial segments of functions from  $A$  into  $B$ . We’ll need a name for this set:

$${}^<A B = \{f \mid \text{for some } t \in A, f \text{ is a function from } \text{seg } t \text{ into } B\}$$

OK, here’s the theorem:

**Transfinite Recursion Theorem, Preliminary Form** Assume that  $<$  is a well ordering on  $A$ ,

and that  $G: {}^{<A}B \rightarrow B$ . Then there is a unique function  $F: A \rightarrow B$  such that for any  $t \in A$ ,

$$F(t) = G(F \upharpoonright \text{seg } t)$$

Here's the formula route:

**Transfinite Recursion Theorem Schema** For any formula  $\gamma(x,y)$ , the following is a theorem:

Assume that  $<$  is a well ordering on a set  $A$ . Assume that for any  $f$  there is a unique  $y$  such that  $\gamma(f,y)$ . Then there exists a unique function  $F$  with domain  $A$  such that for all  $t \in A$ :

$$\gamma(F \upharpoonright \text{seg } t, F(t))$$

Intuitive idea:  $\gamma(x,y)$  can be thought of as meaning “ $y$  is the thing that rule  $\gamma$  associates with  $x$ ”. Thus, “ $\gamma(F \upharpoonright \text{seg } t, F(t))$ ” says that  $F(t)$  is the thing that rule  $\gamma$  associates with  $F \upharpoonright \text{seg } t$ .” (Note that  $\gamma(x,y)$  is, grammatically, a sentence, not a singular term. You can't refer to  $\gamma(x,y)$ . What you can do is refer to the unique  $y$ , such that  $\gamma(x,y)$ .)

To get the idea of the schema approach, let's consider one example. Suppose we want to define a function with domain  $A$  so that each  $F(t)$  is just the set of all the earlier values of  $F$ . Then we can choose our  $\gamma(x,y)$  to be  $y = \text{ran } x$ . The recursion schema then guarantees the existence of a function with domain  $A$  such that for each  $t \in A$ ,  $F(t) = \text{ran } F \upharpoonright \text{seg } t$ . I.e.,  $F(t) = F \upharpoonright \text{seg } t$ . Note that where  $a_0 =$  the least member of  $A$ ,  $F(a_0) = \emptyset$ , since  $\text{seg } a_0 = \emptyset$ . Then, where  $a_1$  is the next member of  $A$ ,  $F(a_1) = \{\emptyset\}$ ; and  $F(a_2) = \{\emptyset, \{\emptyset\}\}$ , etc. I.e., the first  $\omega$  many values of  $F$  will simply be the members of  $\omega$ . But we haven't proved this.

The preliminary version of the recursion theorem follows from the schematic version:

Consider any function  $G$  as described, and choose  $\gamma$  to be:

$$\gamma(x,y) = \text{“either } x \in {}^{<A}B \text{ and } y = G(x), \text{ or } x \notin {}^{<A}B \text{ and } y = \emptyset\text{”}$$

It's clearly true that for each  $x$ , there is a unique  $y$  such that  $\gamma(x,y)$ . So the schematic version of the theorem guarantees the existence of a function,  $F$ , with domain  $A$ , such that

(\*) for each  $t \in A$ , either  $F \upharpoonright \text{seg } t \in {}^{<A}B$  and  $F(t) = G(F \upharpoonright \text{seg } t)$ , or  $F \upharpoonright \text{seg } t \notin {}^{<A}B$  and  $F(t) = \emptyset$

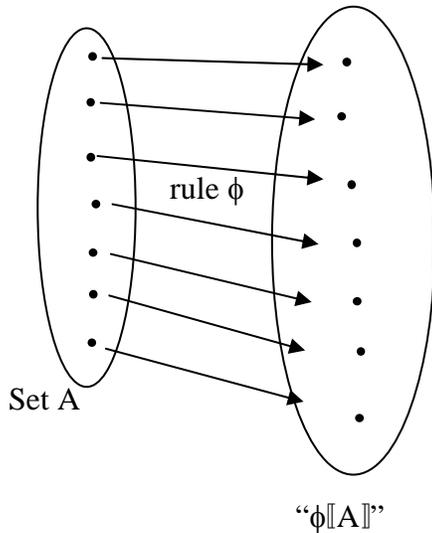
But now we can prove the following (see below): for each  $t \in A$ ,  $F \upharpoonright \text{seg } t \in {}^{<A}B$ . That means we're always on the first disjunct. So, for each  $t \in A$ ,  $F(t) = G(F \upharpoonright \text{seg } t)$ , which is what the preliminary form of the transfinite recursion theorem says.

Proof by transfinite induction that for each  $t \in A$ ,  $F \upharpoonright \text{seg } t \in {}^{<A}B$ : Let  $C = \{t \in A \mid F \upharpoonright \text{seg } t \in {}^{<A}B\}$ . Suppose that  $\text{seg } t \subseteq C$ . We must show that  $t \in C$ , i.e., that  $F \upharpoonright \text{seg } t \in {}^{<A}B$ . Clearly,  $\text{dom } F \upharpoonright \text{seg } t$  is  $\text{seg } t$  (since  $\text{dom } F = A$ ). So we just need to show that  $F \upharpoonright \text{seg } t$  is into  $B$ . Suppose otherwise. Then for some  $a \in \text{seg } t$ ,  $F(a) \notin B$ . By the inductive hypothesis,  $a \in C$ , so  $F \upharpoonright \text{seg } a \in {}^{<A}B$ . By (\*),  $F \upharpoonright \text{seg } a \in {}^{<A}B$  and  $F(a) = G(F \upharpoonright \text{seg } a)$ , or  $F \upharpoonright \text{seg } a \notin {}^{<A}B$  and  $F(a) = \emptyset$ . Thus,  $F(a) = G(F \upharpoonright \text{seg } a)$ . But  $G$  is into  $B$ ; hence  $F(a) \in B$ . Contradiction.

We still need to prove the transfinite recursion schema, but we need new axioms first.

### E. Replacement axioms

Suppose we have the following situation:



That is, we have a known set  $A$ , and we have a “rule”  $\phi$  that associates with each  $a \in A$  a unique object  $\phi(a)$ . Then, intuitively, there should exist a set of all the  $\phi(a)$ s; think of this set as  $\phi[A]$ . For after all,  $\phi[A]$  isn’t too big to be a set; it’s just the same size as  $A$ , which is a set. The axioms of replacement guarantee that  $\phi[A]$  exists. The basic idea is that  $\phi[A]$  is the result of starting with  $A$  and then “replacing” each  $a \in A$  with  $\phi(a)$ .

If the “rule”  $\phi$  had to be a function then the replacement axioms would be useless. For we’d already need to know that  $\phi[A]$  exists in order to know that the function  $\phi$  exists. The replacement axiom schema treats  $\phi$  as a formula, not as a function:

**Replacement axioms** For any formula  $\phi(x,y)$  not containing the letter  $B$ , the following is an axiom:

$$\forall A[(\forall x \in A) (\forall y_1 \forall y_2 ((\phi(x,y_1) \ \& \ \phi(x,y_2)) \rightarrow y_1=y_2) \rightarrow \exists B \forall y (y \in B \leftrightarrow (\exists x \in A) \phi(x,y)))]$$

Actually, the “rule”  $\phi$  doesn’t have to associate something with every member of  $A$ ; it just can never associate two things with a single member of  $A$ .

#### F. Proof of the recursion theorem

Now we use the replacement axioms to prove the transfinite recursion theorem. So, we are given

a well-ordering  $<$  on set  $A$ , and a formula  $\gamma(x,y)$  such that for any  $f$  there is a unique  $y$  such that  $\gamma(f,y)$ . We must now construct a unique function  $F$  with domain  $A$  such that for all  $t \in A$ :

$$\gamma(F \upharpoonright \text{seg } t, F(t))$$

The strategy will be to construct  $F$  as the union of a number of smaller approximating functions.

Recall the proof of the recursion theorem on  $\omega$ . There, we took the union of all “acceptable” functions, i.e., all functions  $v$  such that  $\text{dom } v \subseteq \omega$ ,  $\text{ran } v \subseteq A$ , and

- i) if  $0 \in \text{dom } v$  then  $v(0)=a$
- ii) if  $n^+ \in \text{dom } v$  (where  $n \in \omega$ ) then also  $n \in \text{dom } v$  and  $v(n^+) = F(v(n))$

In essence, these were functions that approximated the desired function up to a certain point. (for if an acceptable function is defined at a number,  $n$ , then it’s defined at the previous number, and then at the previous number, and then so on, all the way down to 0.)

We will here follow the same strategy. Only, it’s a bit easier to implement it now, since we have the well-ordering. (Back in the case of  $\omega$ , we hadn’t yet introduced an ordering of  $\omega$ .) Where  $t \in A$ , call  $v$  “ $\gamma$ -constructed up to  $t$ ” iff  $v$  is a function such that:

$$\text{Dom } v = \{x \in A \mid x \leq t\} \quad \text{and} \quad \forall x \in \text{dom } v \quad \gamma(v \upharpoonright \text{seg } x, v(x))$$

i.e.,  $v$ ’s values for all  $x$  up to and including  $t$  are what we want them to be – the results of applying the “ $\gamma$ -function” to all previous values of  $v$ .

(Just on an intuitive level: why don’t we need anything corresponding to the clause for 0 in the definition of an acceptable function? It’s because the “function”  $\gamma$  yields a value for all segments  $v$ , including  $\text{seg } t_0$ , where  $t_0$  is the minimal element of  $A$ .  $\text{Seg } t_0$  is  $\emptyset$ , and so in effect the same function  $\gamma$  that yields values for all other  $t \in A$  also yields a value for the initial element  $t_0$ ; whereas in the case of  $\omega$ , our function  $F$  that yielded values for the numbers that were successors didn’t yield a value for 0. *Why* did we need to do things that way? Because: we lacked an ordering at that stage on  $\omega$ , and so we lacked a neat way of stating a single rule that would yield as a special case the value of the desired function when its argument was 0. If we had had an ordering, we could have replaced the value  $a$  and the function  $F$  with a single function  $G$  which would take  $\text{seg } n$  as input and return the value of the desired function for  $n$  as output.)

1. OK, so let:

$$K = \{f \mid \exists t \in A \text{ } f \text{ is } \gamma\text{-constructed up to } t\}$$

We are going to show that  $\bigcup K$  is our desired function. But first we need to establish that there is such a set as  $K$ . For that, we’re going to use a replacement axiom. Basically, we’re going to

begin with  $A$ , and then replace each  $t \in A$  with the one and only function that is  $\gamma$  constructed up to  $t$ . To be able to do that, we need to prove that you can't get two different functions  $\gamma$  constructed up to  $t$ . So, we'll prove:

(+) If  $t_1 \leq t_2$ ,  $v_1$  is  $\gamma$  constructed up to  $t_1$ , and  $v_2$  is  $\gamma$  constructed up to  $t_2$ , then  $v_1(x) = v_2(x)$  for all  $x \leq t_1$ .

Suppose otherwise; then there's a least  $x (\leq t_1)$  for which  $v_1(x) \neq v_2(x)$ . Since this is the least such  $x$ , for all  $y < x$ ,  $v_1(y) = v_2(y)$ , and hence  $v_1 \upharpoonright \text{seg } x = v_2 \upharpoonright \text{seg } x$ . But then, since  $v_1$  and  $v_2$  are both  $\gamma$  constructed up to  $t_1$ , we have:

$\gamma(v_1 \upharpoonright \text{seg } x, v_1(x))$   
 $\gamma(v_2 \upharpoonright \text{seg } x, v_2(x))$ , and so  $\gamma(v_1 \upharpoonright \text{seg } x, v_2(x))$

And then, given what we're told about  $\gamma$ , it follows that  $v_1(x) = v_2(x)$  – contradiction. So (+) is true.

It follows from (+) that

For any  $t \in A$ ,  $v_1$  and  $v_2$ , if  $v_1$  is  $\gamma$ -constructed up to  $t$  and  $v_2$  is  $\gamma$ -constructed up to  $t$ , then  $v_1 = v_2$

It then follows by replacement that

There exists a set,  $K$ , such that  $\forall v (v \in K \text{ iff } (\exists t \in A) v \text{ is } \gamma\text{-constructed up to } t)$

(we have chosen our  $\phi$  in the replacement schema to be  $\phi(t, v) = \text{"}v \text{ is } \gamma\text{-constructed up to } t\text{"}$ ). So our set  $K$  exists.

Now let  $F = \bigcup K$ . Thus:

(\*)  $\langle x, y \rangle \in F$  iff  $y = v(x)$  for some  $v \in K$

$F$  is a function: suppose  $\langle x, y \rangle, \langle x, z \rangle \in F$ . So for some  $v_1$  and  $v_2 \in K$ ,  $v_1(x) = y$  and  $v_2(x) = z$ . Since  $v_1, v_2 \in K$ , for some  $t_1, t_2 \in A$ ,  $v_1$  is  $\gamma$ -constructed to  $t_1$  and  $v_2$  is  $\gamma$ -constructed to  $t_2$ . Suppose without loss of generality that  $t_1 \leq t_2$ . Then, since  $x \leq t_1$  ( $x \in \text{dom } v_1$ ), by (+)  $v_1(x) = v_2(x)$ , so  $y = z$ .

2. Next we'll show:

For any  $x \in \text{dom } F$ ,  $\gamma(F \upharpoonright \text{seg } x, F(x))$

Consider any such  $x$ . Then  $\langle x, F(x) \rangle \in F$ , so by (\*),  $F(x) = v(x)$  for some  $v \in K$ . Since the  $v$ 's are

all  $\gamma$ -constructed up to some  $t$  (for  $x \leq t$ ), we have:  $\gamma(v \upharpoonright \text{seg } x, v(x))$ . But since  $v \subseteq F$  and  $F$  is a function,  $v \upharpoonright \text{seg } x = F \upharpoonright \text{seg } x$ , and  $v(x) = F(x)$ .

3. Now we want to show that  $\text{dom } F = A$ . Obviously  $\text{dom } F \subseteq A$ ; suppose for reductio that  $\text{dom } F \subsetneq A$ . Let  $a$  be the least member of  $A - \text{dom } F$ . Consider  $F \upharpoonright \text{seg } a$ . By our hypothesis about  $\gamma$ , there exists a unique  $y$  such that  $\gamma(F \upharpoonright \text{seg } a, y)$ . Consider now

$$v = F \upharpoonright \text{seg } a \cup \{ \langle a, y \rangle \}.$$

We will argue that  $v$  is  $\gamma$ -constructed up to  $a$ , from which it follows that  $v \in K$ , and hence that  $\langle a, y \rangle \in F$ , and hence that  $a \in \text{dom } F$  after all.

First,  **$v$  is a function**; for  $F \upharpoonright \text{seg } a$  is a function and  $a \notin \text{dom } F \upharpoonright \text{seg } a$ .

Second, since  $a$  is the least member of  $A - \text{dom } F$ , for all  $x < a$ ,  $x \in \text{dom } F$ . Hence  **$\text{dom } v = \{x \in A \mid x \leq a\}$** . (Note that for no  $b > a$  is  $b \in \text{dom } F$  (since  $F$  is  $\bigcup K$ , and when  $b \in \text{dom } v' \in K$ , then all  $y < b$  are also in  $\text{dom } v'$ ). Hence  $\text{dom } F = \text{seg } a$ , and  $F \upharpoonright \text{seg } a = F$ .)

Third, consider any  $x \leq a$ .

If  $x < a$  then, since  $x \in \text{dom } F$ , for some  $v' \in K$ ,  $F(x) = v'(x)$ , and so  $\gamma(v' \upharpoonright \text{seg } x, F(x))$ ; but  $v' \upharpoonright \text{seg } x = F \upharpoonright \text{seg } x = v \upharpoonright \text{seg } x$  (since  $v' \subseteq F$ ,  $F$  is a function), so  $\gamma(v \upharpoonright \text{seg } x, v(x))$ .

If  $x = a$  then by construction of  $v$ ,  $\gamma(v \upharpoonright \text{seg } a, v(a))$

Either way, **for any  $x \in \text{dom } v$ ,  $\gamma(v \upharpoonright \text{seg } x, v(x))$** .

So  $v$  is  $\gamma$ -constructed up to  $a$ .

4. Now we want to show that  $F$  is unique (it's the only function with domain  $A$  such that for all  $t \in A$ ,  $\gamma(F \upharpoonright \text{seg } t, F(t))$ .) Let  $F_1$  and  $F_2$  be any two such functions. Let  $B$  be  $\{t \in A \mid F_1(t) = F_2(t)\}$ . We'll show by transfinite induction that  $B = A$ . Suppose that all  $x < t$  are in  $B$ . Then  $F_1 \upharpoonright \text{seg } t = F_2 \upharpoonright \text{seg } t$ . But  $\gamma(F_1 \upharpoonright \text{seg } t, F_1(t))$  and  $\gamma(F_2 \upharpoonright \text{seg } t, F_2(t))$ , and  $\gamma$  is "functional"; hence  $F_1(t) = F_2(t)$ . So  $t \in B$ . By transfinite induction,  $B = A$ .

### G. Idea of Ordinal numbers

The cardinal numbers measured size in one sense: two sets got assigned the same cardinal

number iff they were equinumerous.

We're going to introduce another measure of size. Consider a well-ordered structure  $\langle A, R \rangle$ . This isn't just the set  $A$ ; it's  $A$  together with a certain ordering; hence  $R$  strings  $A$  out on a line, and it's a well-ordering. We're going to introduce a measure of the "length" of  $A$  relative to  $R$ . The length assigned to  $\langle A, R \rangle$  will be the same as that assigned to  $\langle A', R' \rangle$  iff they "look" exactly alike -- iff  $R'$  strings  $A'$  out into a line in exactly the same way that  $R$  strings  $A$  out into a line. This "length" will be an ordinal number -- a measure of the length of well-ordered structures.

For bare sets, looking alike is standing in one-to-one correspondence. For well ordered sets we have a richer notion of looking alike; the well-ordered sets must be isomorphic. Basically, not only must the sets stand in one-to-one correspondence; but the correspondence must preserve the holding of the well orderings. We'll make this more precise soon.

Our method will be to use the membership relation to construct "standardized" well-orderings.

#### H. Epsilon images

Take any well ordering  $<$  on  $A$ . Our goal is to construct a "copy" of this well-ordered set where the ordering is not  $<$ , but is rather the relation of set membership.

We described earlier the way to construct a function,  $E$ , where for all  $t \in A$ ,  $E(t)$  is the set of all earlier values  $E(x)$  for all  $x < t$ . Thus,  $E(a_0) = \emptyset$ ,  $E(a_1) = \{\emptyset\}$ , and so on. The method is to choose  $\gamma(x, y)$  in the recursion schema to be " $y = \text{ran } x$ ". We then are guaranteed that a function,  $E$ , exists such that for each  $t \in A$ ,  $\gamma(E \upharpoonright \text{seg } t, E(t))$ ; i.e.,  $E(t) = \text{ran } E \upharpoonright \text{seg } t = E \upharpoonright \text{seg } t$  = the set of all earlier values of  $E$ .

Call the range of  $E$  " $\alpha$ ".  $\alpha$  is called the "epsilon-image" of the well-ordered structure  $\langle A, < \rangle$ .  $\alpha$  is also what we're going to call an ordinal number. Our immediate goal is to show that the relation  $\in_\alpha$ , (the relation  $\{<x, y> \mid x, y \in \alpha \text{ and } x \in y\}$ ) holds in the same pattern over  $\alpha$  as  $<$  does over  $A$ :

#### **Theorem 7D**

Let  $A$ ,  $<$ ,  $E$ , and  $\alpha$  be as described above. Then:

- (a)  $E(t) \notin E(t)$  for all  $t \in A$
- (b)  $E$  maps  $A$  one-to-one onto  $\alpha$
- (c) For any  $s$  and  $t \in A$ ,  $s < t$  iff  $E(s) \in E(t)$
- (d)  $\alpha$  is a transitive set

Proof:

(a): let  $t$  be the least (relative to  $<$ ) counterexample to this claim. By definition of  $E$ ,  $E(t) = E[\text{seg } t]$ , so  $E[\text{seg } t] \in E[\text{seg } t]$ . So, for some  $x < t$ ,  $E(x) = E[\text{seg } t]$ . But  $E(x) \in E[\text{seg } t]$  (since  $x < t$ ), so  $E(x) \in E(x)$ , contradicting  $t$ 's leastness.

(b):  $E$  is obviously onto  $\alpha$  since  $\alpha$  was defined as  $\text{ran } E$ . Now suppose that  $t_1, t_2 \in A$  and  $t_1 \neq t_2$ . Suppose without loss of generality that  $t_1 < t_2$ . Then  $t_1 \in \text{seg } t_2$ . Hence  $E(t_1) \in E[\text{seg } t_2] = E(t_2)$ . So by part (a),  $E(t_1) \neq E(t_2)$ .

(c): Suppose  $s < t$ . Then  $s \in \text{seg } t$ ; so  $E(s) \in E[\text{seg } t] = E(t)$ . Suppose  $E(s) \in E(t)$ . Then since  $E(t) = E[\text{seg } t]$ ,  $E(s) = E(s')$ , for some  $s' < t$ . Since  $E$  is one-to-one (part b),  $s = s'$ ; so  $s < t$ .

(d): Suppose  $a \in b \in \alpha = \text{ran } E$ . Then since  $b \in \text{ran } E$ , for some  $t \in A$ ,  $b = E(t) = E[\text{seg } t]$ . So, since  $a \in E[\text{seg } t]$ , for some  $x < t$ ,  $a = E(x)$ , so  $a \in \text{ran } E = \alpha$ .

## I. Isomorphisms

OK, now we'll formalize the idea of ordered structures "looking the same". The ultimate goal of this will be a proof that, in light of the theorem just proved (7D),  $\in_\alpha$  well-orders  $\alpha$ .

**Definition.** Let  $\langle A, R \rangle$  and  $\langle B, S \rangle$  be structures (i.e., pairs of sets with orderings that are at least partial.) Then  $f$  is an *isomorphism* between  $\langle A, R \rangle$  and  $\langle B, S \rangle$  iff  $f$  is a one-one function,  $E$ , from  $A$  onto  $B$ , such that  $xRy$  iff  $f(x)Sf(y)$ . If such an  $f$  exists then  $\langle A, R \rangle$  and  $\langle B, S \rangle$  are said to be *isomorphic* – " $\langle A, R \rangle \cong \langle B, S \rangle$ "

Thus, given theorem 7D,  $\langle A, < \rangle \cong \langle \alpha, \in_\alpha \rangle$ .

The following theorems are straightforward (I won't prove them.)

**Theorem 7E**  $\cong$  is an "equivalence concept"

**Lemma 7F** Assume  $f$  is a one-to-one function from  $A$  into  $B$  and  $<_B$  is a partial ordering on  $B$ . Define the following binary relation on  $A$ :

$x <_A y$  iff  $f(x) <_B f(y)$  for any  $x, y \in A$

then:

- (a)  $<_A$  is a partial ordering on  $A$
- (b) If  $<_B$  is a linear ordering on  $B$  then  $<_A$  is a linear ordering on  $A$
- (c) If  $<_B$  is a well ordering on  $B$  then  $<_A$  is a well ordering on  $A$

This lemma is a special case of the following intuitive idea. If a relation  $R$  holds on set  $B$ , and we have a one-one function from  $A$  to  $B$  that preserves the holding of  $R$ , then for any facts about the holding of  $R$  on  $B$ , there are corresponding facts about the image of this relation  $R$  in  $A$ .

**Theorem 7G** If  $\langle A, R \rangle \cong \langle B, S \rangle$  then if one is partially (or linearly, or well) ordered, then so is the other.

**Corollary 7H** If  $<$  is a well ordering on  $A$  and  $\alpha$  is the  $\in$ -image of  $\langle A, < \rangle$ , then  $\alpha$  is a transitive set and  $\in_A$  is a well ordering on  $\alpha$ .

## J. Ordinal numbers defined

As I said, we want ordinal numbers to measure length. Same length = being isomorphic. So we want  $\text{ord } \langle A, R \rangle = \text{ord } \langle B, S \rangle$  iff  $\langle A, R \rangle \cong \langle B, S \rangle$ .

We will define  $\text{ord } \langle A, R \rangle$  as the  $\in$ -image of  $\langle A, R \rangle$ . For this to be an acceptable definition, we need to establish the following theorem:

**Theorem 7I** Well-ordered structures are isomorphic iff they have the same  $\in$ -image

Right to left is easy. Suppose  $\langle A, < \rangle$  and  $\langle B, <' \rangle$  have the same  $\in$ -image. By 7D, each is isomorphic to this common  $\in$ -image, and so they are isomorphic to each other.

Left to right: Suppose  $\langle A, < \rangle \cong \langle B, <' \rangle$ . Let  $f$  be an isomorphism from  $\langle A, < \rangle$  onto  $\langle B, <' \rangle$ . And let  $E_A$  and  $E_B$  be the isomorphisms described earlier between  $\langle A, < \rangle$  and  $\langle B, <' \rangle$  and their  $\in$ -images.

Let's now show that **for all  $a \in A$ ,  $E_A(a) = E_B(f(a))$**  (and thus that the  $\in$ -image of  $A$  is a subset of the  $\in$ -image of  $B$ ). We'll do this by transfinite induction. Let  $T = \{a \in A \mid$

$E_A(a)=E_B(f(a))$ . Suppose for the induction that  $\forall x < a \in T$ . Now,  $E_A(a)=E_A[\text{seg } a]$  and  $E_B(f(a))=E_B[\text{seg } f(a)]$ . Since  $f$  is an isomorphism,  $b < f(a)$  iff  $f^{-1}(b) < a$ . So  $b < f(a)$  iff  $f^{-1}(b) \in \text{seg } a$ . So,  $\text{seg } f(a) = \{b \mid f^{-1}(b) \in \text{seg } a\} = \{f(a') \mid a' \in \text{seg } a\}$ . Thus,  $E_B[\text{seg } f(a)] = E_B[\{f(a') \mid a' \in \text{seg } a\}] = \{E_B(f(a')) \mid a' \in \text{seg } a\}$ , which, by the ih,  $= \{E_A(a') \mid a' \in \text{seg } a\} = E_A[\text{seg } a]$ . Thus,  $E_B(f(a))=E_A(a)$ , so  $a \in T$ .

We must now use this fact to show that  $\alpha_A = \text{ran } E_A = \alpha_B = \text{ran } E_B$ . Let  $x \in \text{ran } E_A$ . Then  $E_A(a)=x$  for some  $a \in A$ ; so  $E_B(f(a))=x$  by what was just shown, so  $x \in \text{ran } E_B$ . Next let  $x \in \text{ran } E_B$ . Then  $E_B(b)=x$  for some  $b \in B$ . Since  $f$  is onto  $B$ ,  $b=f(a)$  for some  $a \in A$ . By what was just shown,  $E_A(a)=E_B(b)$ , so  $x \in \text{ran } E_A$ .

This then justifies the following definition:

Where  $<$  is a well ordering on  $A$ , the *ordinal number* of  $\langle A, < \rangle$  is its  $\in$ -image.

For given the definition, we can infer this:  $\text{ord } \langle A, R \rangle = \text{ord } \langle B, S \rangle$  iff  $\langle A, R \rangle \cong \langle B, S \rangle$

An ordinal number is defined as the ordinal number of some well-ordered structure.

### K. Isomorphisms between well-ordered structures

The first thing we'll show is that if you have two well-ordered structures, either they are isomorphic, or one is isomorphic to an "initial segment" of the other. This is a "trichotomy law" for well-ordered structures.

We need a definition first. Suppose you take an ordering  $<$  on  $A$ . And suppose you consider  $C \subseteq A$ . You can "chop off" the bit of  $<$  that isn't in  $C$  thus:

$$\langle C, <^\circ \rangle = \langle C, < \cap C \times C \rangle$$

And we'll need an intermediate theorem:

**Theorem 7J** Assume that  $<$  is a partial (linear, well) ordering on  $A$  and that  $C \subseteq A$ . Then  $<^\circ$  is a partial (linear, well) ordering on  $C$ .

This makes sense (won't prove it; it's easy.)

Now the trichotomy law:

**Theorem 7K** Let  $\langle_A$  and  $\langle_B$  be well orderings on A and B. Then one of the following alternatives holds:

$$\begin{aligned} \langle_A, \langle_A \rangle &\cong \langle_B, \langle_B \rangle \\ \langle_A, \langle_A \rangle &\cong \langle \text{seg } b, \langle_B^\circ \rangle \quad \text{for some } b \in B \\ \langle \text{seg } a, \langle_A^\circ \rangle &\cong \langle_B, \langle_B \rangle \quad \text{for some } a \in A \end{aligned}$$

The proof works by starting to pair objects of A with objects of B, in order of  $\langle_A$  and  $\langle_B$ , respectively; eventually we'll run out of one set or the other, or they will be isomorphic. We'll define this pairing by transfinite recursion. The function we'll define will map A into B, so long as B hasn't run out; after that, we need to map the members of A to something. To that end, we'll choose some "extraneous" object, e, to map members of A to. e must not be in either A or B. Here's the function:

$$F(t) = \begin{cases} \text{The least element of } B - F[\text{seg } t] & \text{if } B - F[\text{seg } t] \text{ is nonempty} \\ e & \text{if } B - F[\text{seg } t] \text{ is empty} \end{cases}$$

The transfinite recursion theorem guarantees that F exists. It's a function from A into  $B \cup \{e\}$ . Now we consider three cases, corresponding to whether and when F "runs out". They correspond to what ran F is like.

*Case I*  $e \in \text{ran } F$ . This means that F "ran out" of members of A. So we ought to be able to prove, then,  $\langle \text{seg } a, \langle_A^\circ \rangle \cong \langle_B, \langle_B \rangle$  for some  $a \in A$ . Which a? Let a be the least member of A such that  $f(a)=e$ . We'll show that  $F^\circ = F \upharpoonright \text{seg } a$  is an isomorphism between  $\langle \text{seg } a, \langle_A^\circ \rangle$  and  $\langle_B, \langle_B \rangle$ .

Well,  $F^\circ$  is clearly defined on  $\text{seg } a$  (since F was defined on all of A).

And its range is all of B: since  $F(a)=e$  (and  $e \notin B$ ), we're on the second case of the definition of F, and so  $B - F[\text{seg } a]$  is empty. But  $F[\text{seg } a]$  is  $\text{ran } F^\circ$

Next, note that, for  $x, y \in \text{seg } a$ :

$$(*) \text{ If } x \leq_A y \text{ then } F(x) \leq_B F(y)$$

For if  $x \leq_A y$  then  $F[\text{seg } x] \subseteq F[\text{seg } y]$ . But then, given the definition of complement, we have:  $B - F[\text{seg } y] \subseteq B - F[\text{seg } x]$ . But now, remember that  $F(t)$ , for all  $t < a$ , is the least element of  $B - F[\text{seg } t]$ . But for any two sets, X and Y, if  $X \subseteq Y$ , then the least member of Y is  $\leq$  the least member of X (for  $x_0$  is in Y, so  $y_0 \leq x_0$ .) Hence,  $F(x) \leq_B F(y)$ .

Further note that

if  $x <_A y$  then  $F(x) \neq F(y)$

For if  $x <_A y$  then  $F(x) \in F[\text{seg } y]$ , but  $F(y)$  can never be a member of  $F[\text{seg } y]$  (because it's defined as a member of  $B - F[\text{seg } y]$ .)

This tells us two things, first that

$F$  is one-to-one

And second that:

If  $x <_A$  then  $F(x) <_B F(y)$

But finally notice that from (\*), plus facts about well-orders, it follows that:

if  $F(x) <_B F(y)$  then  $x <_A y$

For if  $F(x) <_B F(y)$  then  $F(y)$  is not  $\leq_B F(x)$  (trichotomy for  $\leq_B$ ); given (\*) we have not:  $y \leq_A x$ , so by trichotomy  $x <_A y$ . So,  $F$  is an isomorphism as claimed.

$X$  can't be less than the least member of  $Y$ , for the least member of  $X$  is in  $Y$ , and so

then, since  $x, y <_A$ , given the definition of  $F$  we have:

*Case 2*  $\text{ran } F = B$  Here  $F$  exactly works out. In this case, the proof from the last case goes through to show that  $F$  is one-to-one and preserves order. But now,  $\text{ran } F$  is  $B$ . so  $F$  is an isomorphism between the entire structures.

*Case 3*  $\text{ran } F \subset B$ . This is the only remaining case. We now want to show that  $F$  is an isomorphism from all of  $A$  to an initial segment of  $B$ . Which one? Let  $b$  = the least member of  $B - \text{ran } F$ .

First we'll show that  $\text{ran } F = \text{seg } b$ . Since  $b$  is the least member of  $B$  not in  $\text{ran } F$ , we know that  $\text{seg } b \subseteq \text{ran } F$ . But  $\text{ran } F \subseteq \text{seg } b$ , because if  $x \in \text{ran } F$  were not in  $\text{seg } b$ , since  $b$  is not in  $\text{ran } F$ , so  $b < x$ , contradicting the fact that  $b$  is the *least* member of  $B$  not in  $\text{ran } F$ .

As before,  $F$  is one-to-one and preserves order. So  $F$  is the isomorphism we're looking for.

So: for any two well-ordered structures, one is isomorphic to an initial segment of the other, or they're isomorphic.

What this means is that well-orderings are basically all alike except for the “length”. So ordinals are arbitrary ones we choose to measure length. They’re nice because the same relation works in each case ( $\in$ ); we don’t have to pair the set with the relation; we can let the relation be understood. And it’s convenient to have the relation be  $\in$ .

L. More about ordinals

**Definition** A set  $A$  is *well-ordered by epsilon* iff  $\in_A (\{ \langle x, y \rangle \in A \times A \mid x \in y \})$  is a well-ordering on  $A$

7H earlier showed that each ordinal is a transitive set well-ordered by epsilon. We’ll next show that the converse holds. Thus, being a transitive set well-ordered by epsilon is necessary and sufficient for being an ordinal number.

**Theorem 7L.** Let  $\alpha$  be any transitive set well-ordered by epsilon. Then  $\alpha$  is an ordinal number. In fact,  $\alpha$  is the  $\in$ -image of  $\langle \alpha, \in_A \rangle$

Let  $E$  be our usual function from  $\alpha$  to its  $\in$ -image. We’re going to use transfinite induction to show that  $E$  is the identity function on  $\alpha$ . Now, the  $\in$ -image of  $\alpha$  is defined as  $\text{ran } E$ ; since  $E$  is the identity function, that means that the  $\in$ -image of  $\alpha$  is just  $\alpha$ . And hence, since  $\alpha$  is the epsilon image of something, it will follow that  $\alpha$  is an ordinal.

First note that since  $\alpha$  is a transitive set, then if  $t \in \alpha$ , then  $x \in t$  iff  $x \in_{\alpha} t$ . Now, the well-ordering here on  $\alpha$  is  $\in_{\alpha}$ . So  $\text{seg } t = \{ x \in \alpha \mid x \in_{\alpha} t \} = \{ x \in \alpha \mid x \in t \} = t$ .

Suppose, now, for transfinite induction (on  $\alpha$ , which we can do since  $\alpha$  is well-ordered by epsilon) that for all  $x \in \text{seg } t$ ,  $E(x) = x$ . We must show that  $E(t) = t$ .

$$E(t) = \{ E(x) \mid x \in_{\alpha} t \}$$

But  $t = \text{seg } t$ , so by the i.h., this is

$$= \{ x \mid x \in_{\alpha} t \}$$

$$= \text{seg } t$$

$$= t$$

So now we have a necessary and sufficient condition for being an ordinal number: that it be a transitive set that is well-ordered by epsilon. (One could have taken that as a definition, and then shown it to follow that ordinals are  $\in$ -images of sets.)

Next:

**Theorem 7M** The following are valid for any ordinal numbers  $\alpha$ ,  $\beta$ , and  $\gamma$ :

- (a) (transitive class) Any member of  $\alpha$  is itself an ordinal number
- (b) (transitivity)  $\alpha \in \beta \in \gamma \Rightarrow \alpha \in \gamma$
- © (irreflexivity)  $\alpha \neq \alpha$
- (d) (trichotomy) Exactly one of the alternatives holds:  
 $\alpha \in \beta \quad \alpha = \beta \quad \beta \in \alpha$
- (e) (Well) Any nonempty set  $S$  of ordinal numbers has a least element  $\mu$  (i.e.,  $\mu \in$  or  $=$  to  $\alpha$  for each  $\alpha \in S$ )

What this theorem is saying is that the “class” of ordinals itself is kind of like an ordinal number. Remember that we just showed a bit ago that ordinal numbers are transitive sets that are well-ordered by  $\in$ ; this theorem says the same thing about the “class” of ordinals.

I want to prove just the first part of this theorem: that any member of an ordinal is an ordinal. Suppose  $x \in \alpha$ . By definition of an ordinal,  $\alpha$  is the  $\in$ -image of some well-ordered structure  $\langle A, < \rangle$ . Let  $E$  be the usual function defined on  $A$ .  $E$ 's range is  $\alpha$ , so for some  $a \in A$ ,  $x = E(a)$ . Let's show that  $x$  is the  $\in$ -image of  $\langle \text{seg } a, <^\circ \rangle$ . (This will show that  $x$  is an ordinal, since  $<^\circ$  well orders  $\text{seg } a$  by 7J.) To do this, we just need to show that when you define the usual function  $E'$  on  $\text{seg } a$ , its range is  $x$ .

Let's show by transfinite induction that  $E$  and  $E'$  agree over  $\text{seg } a$ . Let  $T = \{y \in \text{seg } a \mid E(y) = E'(y)\}$ . Suppose that all  $z \in \text{seg } y$  are in  $T$ . Then  $E[\text{seg } y] = E'[\text{seg } y]$ . So, by definition of  $E$  and  $E'$ ,  $E(y) = E'(y)$ . So  $y \in T$ . So, by transfinite induction,  $T =$  all of  $\text{seg } a$ .

Thus,  $\text{ran } E'$  is  $E[\text{seg } a]$ . But  $E[\text{seg } a] = E(a) = x$ .

This then has some important consequences:

*Corollary 7N*

- (a) Any transitive set of ordinal numbers is itself an ordinal number
- (b) 0 is an ordinal number
- (c) if  $\alpha$  is an ordinal number, so is  $\alpha^+$
- (d) if  $A$  is a set of ordinal numbers, so is  $\bigcup A$

Let's work through this proof. First (a). Suppose  $A$  is a transitive set of ordinal numbers, and consider  $\in_A$ . 7M tells us that  $\in_A$  is transitive, irreflexive, satisfies trichotomy, and obeys the leastness principle. So  $A$  is well ordered by  $\in$ , and so by 7L is an ordinal.

Second, (b). follows from (a), since  $0$  is a transitive set of ordinals (vacuously).

Third, ©. Suppose  $\alpha$  is an ordinal. Let's get ourselves in a position to use part (a). We know from 7Ma that each member of  $\alpha$  is itself an ordinal. That means that each member of  $\alpha \cup \{\alpha\}$  (i.e.,  $\alpha^+$ ) is an ordinal. All that remains is to show that  $\alpha^+$  is a transitive set. Now, 4E says that if  $a$  is a transitive set, then  $\bigcup a^+ = a$ . Applied to the present case, this means that  $\bigcup \alpha^+ = \alpha$ . So now, suppose  $x \in y \in \alpha^+$ . Thus,  $x \in \bigcup \alpha^+$ . So  $x \in \alpha$ . So, given the definition of  $^+$ ,  $x \in \alpha^+$ .

Fourth, (d). Suppose  $A$  is a set of ordinal numbers, and consider  $\bigcup A$ . Any member of  $\bigcup A$  is an ordinal, by (a). And  $\bigcup A$  is a transitive set, because it is a union of transitive sets (exercise 5, p. 73). So by (a),  $\bigcup A$  is an ordinal.

Let's make a couple other observations about the ordinals that this corollary tells us about. First, where  $A$  is a set of ordinals, consider the ordinal  $\bigcup A$ . We can show that this is a least upper bound of  $A$ .

First, let's show that for any ordinals  $\alpha, \beta$ ,  $\alpha \in \beta$  iff  $\alpha \subset \beta$ . If  $\alpha \in \beta$  then since ordinals are transitive sets,  $\alpha \subseteq \beta$ . But  $\alpha \neq \beta$  by irreflexivity. Next, suppose  $\alpha \subset \beta$ . Then  $\alpha \neq \beta$ , so either  $\alpha \in \beta$  or  $\beta \in \alpha$  by trichotomy. But if  $\beta \in \alpha$  then  $\beta \in \beta$  (since  $\alpha \subset \beta$ ), violating irreflexivity. So  $\alpha \in \beta$ .

That means that the ordinals are also ordered by  $\subset$ . But now, take any  $\alpha \in A$ . obviously,  $\alpha \subseteq \bigcup A$ , and so by the previous step,  $\alpha \in$  or  $=$  to  $\bigcup A$ . Thus  $\bigcup A$  is an upper bound of  $A$ .

Next let  $b$  be any other ordinal that is an upper bound of  $A$  in the  $\in$  ordering – i.e., each  $\alpha \in A$  is  $\in b$ . So by what we just showed, for each  $a \in A$ ,  $a \subseteq b$ . So  $\bigcup A \subseteq b$ . So by what we just showed,  $\bigcup A \in$  or  $=$  to  $b$ . So  $\bigcup A$  is a *least* upper bound of  $A$ .

Next, where  $\alpha$  is an ordinal, we can show that  $\alpha^+$  is the least ordinal greater than  $\alpha$ . Obviously,  $\alpha \in \alpha^+$ . Moreover, consider any other  $\beta$  such that  $\alpha \in \beta$ . It cannot be that  $\beta \in \alpha^+ = \alpha \cup \{\alpha\}$ , for then  $\beta \in \alpha$  (violating trichotomy) or  $\beta = \alpha$ , violating irreflexivity. So by trichotomy,  $\alpha^+ \in$  or equal to  $\beta$ .

Finally, we can also show that each ordinal is the set of all earlier ordinals. Trivially, each ordinal  $\alpha = \{x \mid x \in \alpha\}$ ; but by 7Ma this becomes  $\alpha = \{x \mid x \text{ is an ordinal and } x \in \alpha\}$ .

Putting all this together, we get a clearer picture of what the ordinals are like. The first one is 0. Each ordinal is the set of earlier ones. For each ordinal, there is an immediate next ordinal – its successor. (by the old successor relation we introduced in discussion of  $\omega$ .) So, the first ordinals are 0,  $0^+$ ,  $0^{++}$ , etc. (It follows immediately by induction from 7Nb and c that each member of  $\omega$  is an ordinal.) Now take  $\omega$ . It is a set of ordinals, so its union is the least ordinal after all the members of  $\omega$ . But  $\bigcup \omega$  is just  $\omega$ . So  $\omega$  is the first ordinal after all the members of  $\omega$ . Then the next ordinals are  $\omega^+$ ,  $\omega^{++}$ , etc.

Note that  $\omega^+$ ,  $\omega^{++}$ , etc., are countable sets (use the usual Hilbert hotel strategy.) (The ordering on ordinals doesn't indicate size, in the sense of cardinals; one ordinal can be a member of another, and thus precede it in the ordinal ordering, even though they are equinumerous.) We'll show soon that there are some uncountable ordinals. Take the set of all the countable ordinals, C. Its union is the first ordinal that is greater than all of them; and it can't be a countable ordinal.

As the first step to showing that there are uncountable ordinals, let's show that there is no set of all the ordinals:

**Burali-Forti theorem:** there is no set of all the ordinals.

Suppose there were; call it  $\alpha$ . Then by 7M it would be a transitive set well-ordered by  $\in_\alpha$ , and so by 7L would be an ordinal. But then  $\alpha \in \alpha$ , violating irreflexivity.

Apparently this was the first set-theoretic paradox discovered. It is a paradox in the sense that it shows that the assumption that for any well-defined condition, for instance "is an ordinal", there is a corresponding set, leads to a contradiction.

### M. Definition of cardinal numbers

We've now finally got the resources available to do a bunch of things we've needed to do for awhile. We can finish showing all the "versions" of the axiom of choice to be equivalent, we can define what cardinal numbers are, and we can be more precise about the set theoretic hierarchy described way back at the beginning of the course.

**First, let's prove**

**Hartog's theorem** For every set, A, there is an ordinal not dominated by A

This is our first connection between the ordinal conception of size (= length of well-orderings) to the cardinal conception of size (which proceeds in terms of one-to-one maps irrespective of order-preservation). It says, in essence, that there are arbitrarily large (in the cardinal sense) ordinals. (We already know that there are arbitrarily large ordinals, since for each ordinal,  $\alpha \in$  the ordinal  $\alpha^+$ .)

Proof:

Consider any set  $A$ ; we will try to show that there is a set,  $\alpha$ , such that

$$\alpha = \{\beta \mid \beta \text{ is an ordinal and } \beta \preceq A\}$$

If we succeed, then the theorem will have been proved. For suppose for reductio that the theorem is false. Then some  $A$  dominates every ordinal, and so the corresponding  $\alpha$  would be a set of all the ordinals, violating Burali-Forti.

First, we can define the following set:

$$W = \{\langle B, < \rangle \mid B \subseteq A \text{ and } < \text{ is a well-ordering on } B\}$$

$W$  is a set because it's a subset of  $\wp A \times \wp(A \times A)$ .

Next, we can use a replacement axiom to construct another set: the set  $E$  of  $\in$ -images of members of  $W$  (since every well ordered structure has a unique  $\in$ -image).

Next, we will show that we can define  $\alpha$  as a subset of  $E$ . That is, we must show that if  $\beta$  is an ordinal and  $\beta \preceq A$ , then  $\beta \in E$ . (Then we can just use a subset axiom to pick out  $\alpha = \{\beta \in E \mid \beta \text{ is an ordinal and } \beta \preceq A\}$ .)

Suppose  $\beta$  is an ordinal and  $\beta \preceq A$ . Then some function  $f$  maps  $\beta$  one-one into a subset of  $A$ . Let  $B = \text{ran } f$ . Define  $< = \{\langle b, c \rangle \in B \times B \mid f^{-1}(b) \in f^{-1}(c)\}$ . (i.e.,  $<$  is the image of  $\in_\beta$  via  $f$ .) Then clearly  $f$  is an isomorphism between  $\langle \beta, \in_\beta \rangle$  and  $\langle B, < \rangle$ , and so by 7G,  $\langle B, < \rangle$  is a well-ordered structure, and so is a member of  $W$ . But by 7I, these two well ordered structures have the same  $\in$ -image, which is  $\beta$  by 7L. So  $\beta \in E$ .

(Enderton also goes on and proves something further, that  $\alpha$  itself is an ordinal, and is the least ordinal that is not dominated by  $A$ .)

Given the axiom of choice and other things we have now proved, one can prove:

### Well-ordering theorem **For any set $A$ , there is a well-ordering on $A$**

I won't prove this in full; I'll only sketch it. Consider any choice function,  $G$ , for  $A$ . What we want to do is go through and arbitrarily start picking things in  $A$ . We'll pick something in  $A$  to be the first thing in the well-ordering, and then pick something else to be the next thing, and so on. To do this rigorously, we'll want to use transfinite recursion. (To make sense of the "process" of constructing the well-ordering.) But to do this, we need a well-ordered set "big enough" to be the domain of the function.

So, we use Hartog's theorem to get an ordinal,  $\alpha$ , that's not dominated by  $A$ . We now define by transfinite induction, a function,  $F$ , like this:

$$\text{For any } \gamma \in \alpha, F(\gamma) = G(A - F \upharpoonright \text{seg } \gamma)$$

We can then define our well-ordering:  $a < a'$  if  $a$  shows up earlier in  $F$ 's range.

Numeration theorem **Any set is equinumerous to some ordinal**

This follows immediately from the well-ordering theorem. Simply calculate the  $\in$ -image  $\alpha$  of  $\langle A, < \rangle$ , where  $<$  is a well ordering on  $A$ ; the resultant function  $E$  is a one-one function from  $A$  onto  $\alpha$  (since  $\alpha$  is defined as the range of  $E$ .)

Now we can define cardinal numbers:

**Definition**  $\text{card } A$  is defined as the least (in the ordinal ordering) ordinal that is equinumerous to  $A$

This is an acceptable definition (Enderton skips this): First, we know from the numeration theorem that *some* ordinal is equinumerous to  $A$ . Moreover, we know from the proof of Hartog's theorem that there exists a set of all ordinals  $\beta$  that are dominated by  $A$ , so we can pick out a set of all ordinals that are equinumerous to  $A$  as a subset of this set. Then by 7Me, we know that this latter set has a least element. So there is always a least ordinal that is equinumerous to  $A$ .

We'll now prove that the definition works the way we want it to:

**Theorem 7P** (a) For any sets  $A$  and  $B$ ,  $\text{card } A = \text{card } B$  iff  $A \approx B$   
(b) if  $A$  is finite,  $\text{card } A$  is the natural number  $\approx$  to  $A$

Pf: (a): given the way we defined  $\text{card } A$ ,  $A \approx \text{card } A$ . Thus  $A \approx \text{card } A$  and  $B \approx \text{card } B$ . So, suppose  $\text{card } A = \text{card } B$ . Then  $A \approx B$ . Conversely, suppose  $A \approx B$ . Then  $A$  and  $B$  are equinumerous to the same ordinals, so the least equinumerous ordinal is the same in each case.

I'll skip the proof of (b).

A couple further consequences of our definition of  $\text{card } A$ .

**Definition** an *initial* ordinal is an ordinal that is not equinumerous to an earlier ordinal

So then, initial ordinals are their own cardinal numbers; and the cardinal numbers, in fact, are just the initial ordinals.

We can show that the ordering relation on the cardinals is the same as that on the ordinals.

Suppose  $\kappa \in \lambda$ . Then  $\kappa \subset \lambda$  (we proved earlier that for ordinals,  $\in$  and  $\subset$  are equivalent). So  $\kappa \leq \lambda$ . But  $\text{card } \kappa$  is  $\kappa$  and  $\text{card } \lambda$  is  $\lambda$ . So  $\kappa \leq \lambda$ . And since  $\kappa \subset \lambda$ ,  $\kappa \neq \lambda$ .  $\kappa < \lambda$ .

Next suppose  $\kappa \neq \lambda$ . Then by trichotomy,  $\lambda \in \kappa$  or  $\lambda = \kappa$ . But if  $\lambda \in \kappa$ , by what we just showed,  $\lambda < \kappa$ . Hence, either  $\lambda < \kappa$  or  $\lambda = \kappa$ , and so by trichotomy,  $\kappa$  is not less than  $\lambda$ .

Thus, we can infer things like this: every nonempty set of cardinal numbers has a least element. Here's another thing we can say. Consider the cardinal  $\aleph_0$ . We call the next highest cardinal number  $\aleph_1$  --- it's the least cardinal greater than  $\aleph_0$ .

How do we know that there is such a thing as the least cardinal greater than  $\aleph_0$ ? Well, let  $\kappa$  be any cardinal greater than  $\aleph_0$ , and let  $\gamma$  be any cardinal greater than  $\kappa$ .  $\gamma$  is an ordinal too, so has all the smaller ordinals as members, and so has all the smaller cardinals as members. So define  $\aleph_1$  as the least member of  $\gamma$ . (Can't use  $\kappa$ , because  $\kappa$  might be  $\aleph_1$  itself.)

Notice that  $\aleph_1$  is the least uncountable ordinal. Then we can call the next highest cardinal  $\aleph_2$ .

We won't go into it here (it's in chapter 8), but one can define  $\aleph_\alpha$ , for all ordinals  $\alpha$ . The basic idea is this:

$\aleph_\alpha$  is the least infinite cardinal different from  $\aleph_\beta$ , for each ordinal  $\beta < \alpha$

(chapter 8 shows how to make this into a legitimate definition.) One can then go on to prove that this construction enumerates *all* the cardinals in the right order – i.e., that each cardinal is  $\aleph_\alpha$  for some ordinal  $\alpha$ , and if  $\alpha < \beta$  then  $\aleph_\alpha < \aleph_\beta$

One can then go on to finish the proof that the various versions of the axiom of choice are equivalent; but I won't go through those proofs.

## N. Rank

The next thing to do is capture precisely the description of the set theoretic hierarchy we gave informally at the beginning of the class. We want to define  $V_\alpha$ , for every ordinal  $\alpha$ . Every set is to show up somewhere in one of the  $V_\alpha$ s; and the hierarchy is to be cumulative in that each  $V_\alpha$  will contain all the earlier ones as subsets, i.e., all the  $V_\beta$ s where  $\beta < \alpha$ . Moreover, for every member of the hierarchy, the next member will be the power set of the previous one. Note how the ordinals are our indices for the levels of the hierarchy, and generate the ordering of the levels. That's another thing that the ordinals are good for.

Now, we can't just define a function  $V$  from ordinals into sets, since there is no function big enough; its domain would have to be all the ordinals. So we need a different strategy.

**Lemma 7Q** For every ordinal  $\delta$ , there is a function  $F_\delta$  with domain  $\delta$  such that for all  $\alpha \in \delta$ ,

$$F_\delta(\alpha) = \bigcup \{ \wp F_\delta(\beta) \mid \beta \in \alpha \}$$

This I won't prove; but obviously one defines  $F_\delta$  by transfinite recursion.

So the idea is that the  $F_\delta$  function generates the hierarchy up until  $\delta$ . We'll construct the whole hierarchy using these  $F_\delta$  functions. First, we'll need this lemma:

**Lemma 7R** For any ordinals  $\delta$  and  $\varepsilon$  where  $\delta \leq \varepsilon$ , any functions  $F_\delta$  and  $F_\varepsilon$  from 7Q agree up to  $\delta$  -- i.e., for any  $\alpha \in \delta$ ,  $F_\delta(\alpha) = F_\varepsilon(\alpha)$ .

I won't prove this; it's done by transfinite induction.

It then follows that (setting  $\delta = \varepsilon$ ) that the functions from 7Q are unique. Now we can define  $V_\alpha$ :

**Definition** For any ordinal  $\alpha$ ,  $V_\alpha$  is defined as  $F_\delta(\alpha)$ , where  $\delta$  is any ordinal greater than  $\alpha$ .

(there always is a greater ordinal, because there's no largest cardinal.)

Let's then state some of the consequences of this definition. We can divide ordinals into three groups:

|                                |                        |
|--------------------------------|------------------------|
| 0                              |                        |
| $\alpha^+$ , for some $\alpha$ | ("successor ordinals") |
| others                         | ("limit ordinals")     |

(Note that if  $\lambda$  is a limit ordinal and  $\beta \in \lambda$ , then  $\beta^+ \in \lambda$ . For since  $\lambda$  is a limit ordinal,  $\beta^+ \neq \lambda$ . Moreover, it cannot be that  $\lambda \in \beta^+$ , for then  $\lambda = \beta$  or  $\lambda \in \beta$ , either way contradicting  $\beta \in \lambda$  via trichotomy. So, by trichotomy,  $\beta^+ \in \lambda$ .)

**Theorem 7U**

- (a) For any ordinals  $\alpha, \beta$  where  $\beta \in \alpha$ ,  $V_\beta \subseteq V_\alpha$ .
- (b)  $V_0 = \emptyset$
- (c)  $V_{\alpha^+} = \wp V_\alpha$ , for any ordinal  $\alpha$
- (d)  $V_\lambda = \bigcup_{\beta \in \lambda} V_\beta$ , for any limit ordinal  $\lambda$

So the first  $V$  is  $\emptyset$ . Then the next is the power set of the first. And so on.  $V_\omega$  is the union of all the finite  $V_i$ s.  $V_{\omega^+}$  is the power set of  $V_\omega$ . And so on. The ordinals have allowed us to make

precise the idea of “never stopping” in this construction.

**Definition** set  $A$  is grounded iff  $A \subseteq V_\alpha$ , for some ordinal  $\alpha$ . In this case, we define the *rank* of  $A$  as the least  $\alpha$  such that  $A \subseteq V_\alpha$ .

Now, it cannot be proved from the axioms so far that every set is grounded. But the iterative conception demands this. So we need a new axiom. It could be “every set is grounded”; but instead it will be something else:

**Theorem 7W** The following two statements are equivalent:

- (a) Every set is grounded
- (b) (Regularity) Every nonempty set  $A$  has a member  $m$  such that  $m \cap A = \emptyset$

Our new axiom is Regularity. Here are some of its consequences:

**Theorem 7X**

- (a) No set is a member of itself
- (b) There are no sets  $a$  and  $b$  such that  $a \in b$  and  $b \in a$
- (c) There is no function  $f$  with domain  $\omega$  such that ...  $f(2) \in f(1) \in f(0)$

Proof of (c): suppose for reductio that for some function  $f$  with domain  $\omega$ ,  $f(n^+) \in f(n)$  for all  $n \in \omega$ , and let  $A = \text{ran } f$ . Let  $m$  be any member of  $A$ . So  $f(n) = m$ , for some  $n \in \omega$ . So  $f(n^+) \in m$  and also  $f(n^+) \in A$ ; so  $A \cap m \neq \emptyset$ . This violates Regularity.

Note that (a) then follows. For suppose  $a \in a$ , and define  $f(n) = a$ , for all  $n \in \omega$ . Then  $f(n^+) = a \in a = f(n)$ , for all  $n$ , violating (c).

## VIII. Natural models

I want to skip ahead to a topic in chapter 9. We’ll only have time to sketch.

In metalogic, we use models plus the soundness theorems to show that certain statements are not consequences of other statements. We’ll do something a little different from that to show that certain of our axioms do not imply others. We’ll construct a model of some of the axioms within the set theoretic hierarchy itself. But it will be too big to be a model in the usual sense; its domain will be too big to be a set.

Let  $\sigma$  be any formula in the language of set theory, and let  $M$  be any set. We can define the *relativization of  $\sigma$  to  $M$*  to be the formula  $\sigma^M$  that results by replacing every  $\forall v$  in  $\sigma$  with  $(\forall v \in M)$  and every  $\exists v$  with  $(\exists v \in M)$ . I.e., we restrict the quantifiers in  $\sigma$  to members of  $M$ .

We can think of any set  $M$  as a model in the following sense: we can consider the truth values of the axioms of set theory, as relativized to  $M$ . The idea is to keep ‘ $\in$ ’ meaning set membership, but to restrict the domain of the quantifiers.

What can be shown is that various of the  $V_\alpha$ s are indeed models of the various axioms of set theory:

**Extensionality:** the relativization of extensionality to any transitive set (each  $V_\alpha$  can be shown to be a transitive set) is true.

Let  $M$  be any transitive set; we must show:

$$\forall x \in M \forall y \in M [\forall z \in M (z \in x \leftrightarrow z \in y) \rightarrow x = y]$$

So, let  $x, y \in M$ , and suppose that (\*)  $\forall z \in M (z \in x \leftrightarrow z \in y)$ . Now, consider any  $a \in x$ . Since  $M$  is transitive,  $a \in M$ . By (\*),  $a \in y$ . So  $x \subseteq y$ . Similarly,  $y \subseteq x$ . So, by (real) extensionality,  $x = y$ .

**Empty set:** if  $\alpha \neq \emptyset$ , then the relativization of Empty Set to  $V_\alpha$  is true

**Pairing:** if  $\lambda$  is a limit ordinal then the relativization of Pairing to  $V_\lambda$  is true

**Union:** for any ordinal  $\alpha$ , the relativization of Union to  $V_\alpha$  is true

**Power set:** for any limit ordinal  $\lambda$ , the relativization of Power Set to  $V_\lambda$  is true

**Subset:** for any ordinal  $\alpha$ , the relativization of any subset axiom to  $V_\alpha$  is true

**Infinity:** for any ordinal  $\alpha$  greater than  $\omega$ , the relativization of Infinity to  $V_\alpha$  is true

**Choice:** For any ordinal  $\alpha$ , the relativization of Choice (version I) to  $V_\alpha$  is true

**Regularity:** for any ordinal  $\alpha$ , the relativization of Regularity to  $V_\alpha$  is true

These are called the Zermelo axioms.

Now, Consider the following ordering of  $\mathbb{Z}$ :  $0 < 1 < 2 \dots < -1 < -2 < \dots$ . It can be shown that it is a well ordering. So  $\langle \mathbb{Z}, < \rangle$  is isomorphic to some ordinal  $\alpha$ . It can be shown that this is a limit

ordinal greater than  $\omega$ . So  $V_\alpha$  is a model for all the axioms of set theory other than the regularity axioms, in the sense that their relativizations to  $V_\alpha$  are true. (What is  $V_\alpha$  like? It's basically this:  $V_\omega \cup \wp V_\omega \cup \wp \wp V_\omega \cup \wp \wp \wp V_\omega \dots$ . There's a ton of stuff in here; e.g., all the reals.)

Is there a problem with the Godel result here? Godel says that for any theory strong enough to represent a certain minimal portion of arithmetic, there can be no proof of that theory's consistency within that theory. But haven't we just used the Zermelo axioms to prove their own consistency? No – the proof of this theorem required lots of stuff about ordinals, and to develop them, we needed transfinite induction and recursion, which in turn required the regularity axioms.

So what about the regularity axioms? Are these true in  $V_\alpha$ ? No -- one can show that the relativizations of the regularity axioms to  $V_\alpha$  are not all true. So it follows that not all of the regularity axioms are consequences of the other axioms. For if  $\Gamma \vdash \phi$ , then  $\Gamma^M \vdash \phi^M$ .

Sketch: Where  $F$  is any (possibly complex) predicate and  $\psi$  is any sentence, call  $\psi^F$  the result of restricting all quantifiers using  $F$ . First prove that (\*) for any model (in the usual model-theoretic sense)  $K$  in which  $F$  is interpreted, and any sentence  $\psi$ ,  $\psi^F$  is true in  $K$  iff  $\psi$  is true in the submodel of  $K$  whose domain is the extension in  $K$  of  $F$ . Now, suppose  $\phi$  is a semantic consequence of  $\Gamma$ , and consider any model  $K$  in which all the members of  $\Gamma^M$  are true. Let  $K^-$  be the submodel whose domain is the extension of the predicate ' $\in M$ '. Since every member of  $\Gamma^M$  is true in  $K$ , by (\*) every member of  $\Gamma$  is true in  $K^-$ . So  $\phi$  is true in  $K^-$ . So by (\*),  $\phi^M$  is true in  $K$ .