14. Naive set theory

1. *Sets*

Naive definition: A **set** is an *unordered collection* of things (*not* counting multiplicities), its **elements**. Write $x \in S$ or $S \ni x$ for an element x of S. Sets are described either as comma-separated *lists* (whose order is not supposed to be significant)

$$S = \{x_1, x_2, \ldots\}$$

or by a *rule*

$$S = \{x : \textit{some condition on } x \textit{ is met}\}$$

The **empty set** is

$$\phi = \{\}$$

[1.0.1] **Theorem:** There is no set S such that $x \in S$ if and only if $x \notin x$.

Proof: Suppose there were such S. Then $S \in S$ if and only if $S \notin S$, contradiction. ///

Extension Principle *(Leibniz)* Two sets are equal if and only if they have the same elements.

[1.0.2] **Corollary:** There is only one empty set ϕ. ///

Idea: *Everything is a set.*

A **subset** T of S is a set such that for all elements x of T also x is an element of S. Write $T \subset S$ or $S \supset T$.

A subset of S is *proper* if it is neither S itself nor ϕ. The **union** of a set F of sets is

$$\bigcup_{S\in F} S = \{x : x \in S \text{ for some } S \in F\}$$

The **intersection** is

$$\bigcap_{S\in F} S = \{x : x \in S \text{ for all } S \in F\}$$

We make an exception in the case of intersections over F for $F = \phi$, since the defining condition would be vacuous, and (supposedly) *every* set would be an element of that intersection, which is not viable. The union and intersection of a finite number of sets can also be written, respectively, as

$$S_1 \cup \ldots \cup S_n$$

$$S_1 \cap \ldots \cap S_n$$

Proto-definition: The *ordered pair* construct (x, y) with *first* component x and *second* component y should have the property that

$$(x, y) = (z, w) \Longleftrightarrow x = z \text{ and } y = w$$

[1.0.3] **Remark:** As sets, taking $(x, y) = \{x, y\}$ fails, since the elements of a set are not ordered. Taking $(x, y) = \{x, \{y\}\}$ fails, since it may be that $x = \{y\}$.

[1.0.4] **Proposition:** We can construct ordered pairs as sets by defining

$$(x, y) = \{\{x\}, \{x, y\}\}$$

Proof: We must prove that $(x, y) = (z, w)$ if and only if the respective components are equal. One direction of the implication is clear. For the other implication, from

$$\{\{x\}, \{x, y\}\} = \{\{z\}, \{z, w\}\}$$

$\{x\}$ is either $\{z\}$ or $\{z, w\}$, and $\{x, y\}$ is either $\{z\}$ or $\{z, w\}$. Treat cases, using the Extension Principle. ///

For finite n, define recursively **ordered n-tuples** by

$$(x_1, \ldots, x_{n-1}, x_n) = ((x_1, \ldots, x_{n-1}), x_n)$$

[1.0.5] **Remark:** Subsequently we *ignore* the internal details of the construction of ordered pair, and only use its properties. This is a typical ruse.

The **Cartesian product** $X \times Y$ of two sets X and Y is the set of ordered pairs

$$X \times Y = \{(x, y) : x \in X, y \in Y\}$$

A **function** or **map** $f : X \longrightarrow Y$ from X to Y is a subset of $X \times Y$ such that for all $x \in X$ there is a unique y in Y such that $(x, y) \in f$. As usual, this is written $f(x) = y$ or $fx = y$. The **image** $f(X)$ of f is

$$f(X) = \text{ image of } f = \{f(x) : x \in X\}$$

[1.0.6] **Remark:** This definition identifies a function with its graph, rather than by a formula or algorithm by which to *compute* the function.

[1.0.7] **Definition:** A function $f : X \longrightarrow Y$ is **surjective** if for every $y \in Y$ there is $x \in X$ such that $f(x) = y$. It is **injective** if $f(x) = f(x')$ implies $x = x'$. If f is both surjective and injective is it **bijective**.

The **composition** $f \circ g$ of two functions $f : Y \longrightarrow Z$ and $g : X \longrightarrow Y$ is defined by

$$(f \circ g)(x) = f(g(x))$$

A **left inverse** g (if it exists) to a function $f : X \longrightarrow Y$ is a function $g : Y \longrightarrow X$ such that $g \circ f = 1_X$, where 1_X is the **identity function** on X, defined by $1_X(x) = x$ for all $x \in X$. A **right inverse** g (if it exists) to a function $f : X \longrightarrow Y$ is a function $g : Y \longrightarrow X$ such that $f \circ g = 1_Y$

Let F be a set of sets. A **choice function** f on F (if it exists) is any function

$$f : F \longrightarrow \bigcup_{S \in F} S$$

such that

$$f(S) \in S$$

for all S in F. To postulate that at least one choice function exists for any set F of sets is a non-trivial thing, and, roughly, is the **Axiom of Choice**. The collection of all choice functions on F is the **direct product** of the sets, denoted

$$\prod_{S \in F} S$$

Again, to know that this is non-empty (for F infinite) requires something!

K. Godel and P. Cohen proved that the Axiom of Choice is not only not provable from other more mundane axioms for sets, but is *independent* of them, in the sense that it is equally consistent to assume the negation of the Axiom of Choice.

A **relation** R between sets X and Y is a subset of $X \times Y$. A (binary) relation on a set X is a subset of $X \times X$. A relation R on X is
• **reflexive** if $(x, x) \in R$ for all $x \in X$
• **symmetric** if $(x, y) \in R$ implies $(y, x) \in R$ for all $x, y \in X$
• **transitive** if $(x, y) \in R$ and $(y, z) \in R$ implies $(x, z) \in R$ An **equivalence relation** is a relation that enjoys all three of these properties. For an equivalence relation R, the **equivalence class** of x is

$$\text{equivalence class of } x = \{y \in X : (x, y) \in R\}$$

2. *Posets, ordinals*

A **partial order** $\leq$ on a set X is a relation R on X, written $x \leq y$ if $(x, y) \in R$, such that
• *(Reflexivity)* $x \leq x$ for all $x \in X$
• If $x \leq y$ and $y \leq x$ then $x = y$
• *(Transitivity)* If $x \leq y$ and $y \leq z$ then $x \leq z$ Then X is a **partially ordered set** or **poset**. We may write $x < y$ if $x \leq y$ and $x \neq y$.

A partial ordering on X is a **total ordering** if for all $x, y \in X$ either $x \leq y$ or $y \leq x$.

A **well ordering** [sic] on a set X is a total ordering on X such that any non-empty subset Y of X has a **minimal element** (also called **least element**). That is, there is an element $y \in Y$ such that for all $z \in Y$ we have $y \leq z$.

[2.0.1] **Proposition:** Let X be a well-ordered set. Let $f : X \longrightarrow X$ be an order-preserving injective map (so $x \leq x'$ implies $f(x) \leq f(x')$). Then

$$f(x) \geq x$$

for all $x \in X$.

Proof: Let Z be the subset of X consisting of elements x such that $f(x) < x$. If Z is non-empty, then it has a least element x. Thus, on one hand, $f(x) < x$. On the other hand, $f(x) \notin Z$, so $f(f(x)) > f(x)$. But, since f preserves order and is injective, $f(x) < x$ implies $f(f(x)) < f(x)$, contradiction. ///

[2.0.2] **Corollary:** The only order-preserving bijection of a well-ordered set X to itself is the identity map. ///

[2.0.3] **Corollary:** There is no order-preserving bijection of a well-ordered set X to a proper **initial segment**

$$X^{<x} = \{y \in X : y < x\}$$

of it for any $x \in X$. ///

[2.0.4] **Example:** The set

$$Z = \{X^{<x} = \{y \in X : y < x\} : x \in X\}$$

of initial segments $X^{<x}$ of a well-ordered set X, with ordering

$$z \leq w \Longleftrightarrow z \subset w$$

has an order-preserving bijection to X by

$$X^{<x} \longleftrightarrow x$$

An **ordinal** is a well-ordered set X such for every element $x \in X$

$$x = X^{<x}$$

That is, x is the set $X^{<x} = \{y \in X : y < x\}$ of its predecessors in X.

[2.0.5] **Example:** The empty set is an ordinal, since the defining condition is met vacuously. Let X be an ordinal that is not the empty set. Then X (being non-empty) has a least element x. Since x is the union of its predecessors, of which there are none, $x = \phi$. So ϕ is the least element of every ordinal.

[2.0.6] **Example:** If X is an ordinal, and $x \in X$, then the **initial segment** below x

$$X^{<x} = \{y \in X : y < x\}$$

is also an ordinal. Indeed, the well-ordering is preserved, and by transitivity the predecessors of y in $X^{<x}$ are exactly the predecessors of y in X, so the defining property of ordinals holds.

[2.0.7] **Example:** If X is an ordinal, then $Y = X \cup \{X\}$, with ordering

$$a \leq b \Longleftrightarrow a \subset b$$

is an ordinal, the **successor** of X. To see this, first note that, for all $y \in Y$ we have $y \subset X$, that is (by definition of the ordering) $y \leq X$. Thus, for $y \in Y$, if $y \neq X$, then $y \subset X$ and (since X is an ordinal) is the set of its predecessors *in* X. And since $y < X$ in Y, X is not among y's predecessors in Y, so y really is the set of its predecessors in Y. And X is the set of its predecessors in Y. ///

Since everything is to be a set, following J. von Neumann, define the initial (finite) ordinals by

$$0 = \phi = \{\}$$

$$1 = \{0\} = \{\phi, \{\phi\}\} = \{\{\}, \{\{\}\}\}$$

$$2 = \{0, 1\} = \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\} = \{\{\}, \{\{\}\}, \{\{\}, \{\{\}\}\}\}$$

$$3 = \{0, 1, 2\} = \{\phi, \{\phi\}, \{\phi, \{\phi\}\}, \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\}\}$$

$$= \{\{\}, \{\{\}\}, \{\{\}, \{\{\}\}\}, \{\{\}, \{\{\}\}, \{\{\}, \{\{\}\}\}\}\}$$

and so on

The set ω of **natural numbers** is[1]

$$\omega = \{0, 1, 2, \ldots\}$$

Define an order $\leq$ on ω by

$$x \leq y \Longleftrightarrow x \subset y$$

It is not at all immediate (with the present definition of the symbols) that ω is an ordinal.

[2.0.8] **Proposition:** If X and Y are ordinals and $Y \subset X$ then there is $x \in X$ such that Y is the initial segment

$$Y = \{y \in X : y < x\} = x$$

Proof: Let Z be the set of elements of X that are not in Y but are below some element of Y. The claim is that Z is empty. If not, let z be the least element in Z. Let $y \in Y$ be such that $z < y$. Since y is the set of its predecessors in X, $x \in y$. But also y is the set of its predecessors in Y, so $x \in y$, contradiction. ///

[2.0.9] **Theorem:** Any two ordinals X, Y are *comparable*, in the sense that either $X = Y$, or X is an initial segment of Y, or Y is an initial segment of X.

Proof: The intersection $X \cap Y$ is an ordinal, since for $z \in X \cap Y$

$$\{w \in X \cap Y : w < z\} = \{x \in X : x < z\} \cap \{y \in Y : y < z\} = z \cap z = z$$

Suppose that X is not contained in Y, and Y is not contained in X. From above, $X \cap Y \subset$ is an initial segment

$$X \cap Y = \{z \in X : z < x\} = x$$

[1] Shuddering at the casual formation of this ostensibly infinite set is reasonable, since its existence as a set is not formally assured by the existence of the separate finite ordinals.

in X for some $x \in X$, and also an initial segment

$$X \cap Y = \{w \in Y : z < y\} = y$$

in Y for some $y \in Y$. But then $x = y$, contradiction. ///

[2.0.10] **Corollary:** Two ordinals admit an order-preserving bijection between them if and only if they are identical, and in that case the only order-preserving bijection is the identity map.

Proof: We already saw that there is at most one order-preserving bijection between two well-ordered sets. Thus, let X and Y be ordinals, and $X \neq Y$. By the theorem, one is an initial segment of the other, so assume without loss of generality that Y is an initial segment

$$Y = \{y \in X : y < x\}$$

for some x in X. Let $f : X \longrightarrow Y$ be an order-preserving bijection. We saw earlier that $f(z) \geq z$ for any well-ordered sets in this situation. But then $f(x) \geq x$, which is impossible. ///

[2.0.11] **Corollary:** The relation on ordinals defined by $x < y$ if and only if x is an initial segment of y is a total ordering. ///

[2.0.12] **Corollary:** Given an ordinal x, its successor ordinal $y = x \cup \{x\}$ has the property that $x < y$. ///

[2.0.13] **Corollary:** There is no largest ordinal. ///

[2.0.14] **Theorem:** The union of any set of ordinals is an ordinal.

Proof: Let F be a set of ordinals, and

$$E = \bigcup_{X \in F} X$$

is also a set of ordinals. Define a relation $<$ on E by $x < y$ if x is an initial segment in y, that is, is an element of y. The transitivity of $<$ follows (again) from the fact that every element of an ordinal is an ordinal. The comparability of all ordinals (from above) says that this is a *total* ordering. To prove that $<$ is a well-ordering, let D be a non-empty subset of E, and let d be any element of D. If d is least in D, we are done. If d is not least in D, then nevertheless $c \in D$ with $c < d$ are elements of d, since $c < d$ only for c an initial segment of d, that is an element of d. Since d is an ordinal, it is well-ordered, so

$$\{c \in D : c < d\} = D \cap d$$

is well-ordered. Thus, D contains a least element. Finally, we must prove that any element e of E is the set of its predecessors in E. Let X be an element of F such that $e \in X$. Since X is an ordinal, e is the set of its predecessors d in X. Thus, all such predecessors d are elements of X, so are elements of the union E. Thus,

$$e = \{d \in X : d < e\} \subset \{d \in E : d < e\}$$

On the other hand, for any $d \in E$, the definition of $d < e$ is that d is an initial segment of e, that is, that $d \in e$. In that case, $d \in X$ for *every* ordinal containing e. That is, we have the opposite inclusion

$$e = \{d \in X : d < e\} \supset \{d \in E : d < e\}$$

and e is exactly the set of its predecessors in the union E. ///

[2.0.15] **Theorem:** Every well-ordered set has an order-preserving bijection to exactly one ordinal.

Proof: First, let X be a well-ordered set with each initial segment

$$X^{<x} = \{y \in X : y < x\}$$

for $x \in X$ isomorphic [2] to an ordinal ω_x. We claim that X is isomorphic to an ordinal. From above, since no two distinct ordinals are isomorphic, and since an ordinal admits no non-trivial maps to itself, for each $x \in X$ the ordinal ω_x is uniquely determined and the order-preserving map $f_x : x \longrightarrow \omega_x$ is unique. We view $F : x \longrightarrow \omega_x$ as an ordinal-valued function F on X.

Consider $x < y$ in X. Since x and y are distinct initial segments of X, they are not isomorphic as ordered sets (indeed, there is no order-preserving injection of y to x). Thus, $F(x) = \omega_x$ is not isomorphic to $F(y) = \omega_y$. Thus, since any two ordinals can be compared, either $F(x) = \omega_x$ is an initial segment of $F(y) = \omega_y$ or *vice versa*. Unsurprisingly, if $\omega_y < \omega_x$ then

$$y \approx \omega_y \subset \omega_x \approx x$$

would give an isomorphism of y to a *proper* initial segment x, but (again) this is impossible. Thus, F is an order-preserving bijection of X to a set $\Omega = \{\omega_x = F(x) : x \in X\}$ of ordinals. Since $\Omega = F(X)$ is the image of the well-ordered set X, Ω is well-ordered. To show that Ω is an ordinal, by definition, we must show that for $\omega \in \Omega$ the initial segment

$$\Omega^{<\omega} = \{\omega' \in \Omega : \omega' < \omega\}$$

is equal to ω. Indeed, the hypothesis is exactly this, so Ω is an ordinal, and X is an ordinal (being isomorphic to Ω).

Now we prove the theorem. First we prove that every element of a (non-empty) well-ordered set X is isomorphic to an ordinal. The least element of X is isomorphic to the ordinal ϕ. Given α in X with β isomorphic to an ordinal for all $\beta < \alpha$, then apply the claim to α (in place of X) to conclude that α is isomorphic to an ordinal. And then the claim implies that X is isomorphic to an ordinal. Since two distinct ordinals are not isomorphic, there is exactly one ordinal to which X is isomorphic. ///

The following corollary is sometimes recast as a paradox:

[2.0.16] **Corollary:** *(Burali-Forti)* The collection of all ordinals is not a set.

Proof: Suppose the collection F of all ordinals were a set. Then (by the theorem) the union

$$E = \bigcup_{S \in F} S$$

would be an ordinal. Thus, E would be an element of itself, contradiction. ///

---

3. *Transfinite induction*

[3.0.1] **Theorem:** Let $P(\alpha)$ be a property that may or may not hold of ordinals α. Suppose that for any ordinal α if $P(\beta)$ for all ordinals $\beta < \alpha$ then $P(\alpha)$ holds. The $P(\alpha)$ holds for *all* ordinals α.

Proof: Let $\omega = \alpha \cup \{\alpha\}$, so ω is an ordinal containing α. Then we can do induction on the *set* ω: prove that $P(\beta)$ holds for all $\beta \in \omega$ (including α). If $P(\gamma)$ *failed* for some γ in ω, then there would be a *least* γ in ω for which it failed. But $P(\delta)$ holds for all $\delta < \gamma$, and the hypothesis assures that $P(\gamma)$ *does* hold, after all. This contradiction shows that $P(\gamma)$ holds for all $\gamma \in \omega$, in particular, for α. ///

In some situations the induction step, namely, proving that $P(\alpha)$ holds if $P(\beta)$ holds for all $\beta < \alpha$, must be broken into cases, depending on the nature of α.
• The **initial ordinal**, ϕ.

[2] As ordered set, of course.

• **Successor ordinals** $\alpha = \beta \cup \{\beta\}$ for some β.
• **Limit ordinals** $\alpha = \bigcup_{\beta<\alpha} \beta$.

[3.0.2] **Remark:** First, contrast the definition of *limit ordinal* with the property enjoyed by *every* ordinal, namely

$$\alpha = \{\beta : \beta \in \alpha\} = \{\beta : \beta < \alpha\}$$

A successor ordinal α is not a limit ordinal, since if $\alpha = \beta \cup \{\beta\}$ then all predecessors of α are subsets of β, and likewise their union, which cannot contain β *as an element*.

[3.0.3] **Proposition:** Every ordinal is either the initial ordinal ϕ, a successor ordinal, or a limit ordinal.

Proof: Suppose α is not ϕ and is not a successor. Let β be the union of the predecessors of α. Since a union of ordinals is an ordinal, β is an ordinal, and $\beta \leq \alpha$. If $\beta < \alpha$ then β is among α's predecessors, so is in the union of predecessors, so is the largest among the predecessors of α. The assumption $\beta < \alpha$ gives $\beta \cup \{\beta\} \leq \alpha$. It cannot be that $\beta \cup \{\beta\} \leq \alpha$ since otherwise $\beta \cup \{\beta\}$ would be a predecessor of α, and thus $\beta \geq \beta \cup \{\beta\}$, which is false. So, then, the successor $\beta \cup \{\beta\}$ of β is α, contradiction to the hypothesis that α is not a successor. Thus, $\beta = \alpha$. ///

Thus, we can rewrite the first theorem in a manner that refers explicitly to the types of ordinals: to prove a property $P(\alpha)$ holds for all ordinals α:
• Prove $P(\phi)$ holds.
• Prove (for all α) that if $P(\alpha)$ holds then $P(\alpha \cup \{\alpha\})$ holds.
• Prove for every limit ordinal λ that if $P(\alpha)$ holds for all $\alpha < \lambda$ then $P(\lambda)$ holds.

4. Finiteness, infiniteness

A set S is **Peano finite** if there is some $n \in \omega$ such that there is a bijection of S to

$$n = \{0, 1, 2, \ldots, n-1\}$$

The set is **Peano infinite** if it is *not* Peano finite.

A set S is **Dedekind infinite** if there is an injection from S to a *proper* subset of S. It is Dedekind *finite* if it is *not* Dedekind infinite.

[4.0.1] **Theorem:** (Granting the Axiom of Choice) The two notions of *infinite* are the same.

[4.0.2] **Remark:** To avoid circularity, we should not presume *arithmetic* at this point.

Proof: Let $f : S \longrightarrow S$ be an injection of S to a proper subset of itself. Choose $s_1 \in S$ but not lying in the image $f(S)$. Claim $f(f(S))$ is a proper subset of $f(S)$. Indeed, $f(s_1)$ cannot be in $f(f(S))$, or there would be $t \in f(S)$ such that $f(t) = f(s_1)$, and then by injectivity of f we would have $t = s_1$, contradicting the fact that $s_1 \notin f(S)$. Certainly f restricted to $f(S)$ is still injective.

Thus, $f(f(f(S)))$ is strictly smaller than $f(f(S))$ By induction, we can find $s_1, s_2, \ldots$ such that $s_1 \notin f(S)$, $s_2 \in f(S)$ but $s_2 \notin f(f(S))$, $s_3 \in f(f(S))$ but $s_3 \notin f(f(f(S)))$, etc. In particular, all these s_i are distinct, so we have an injection

$$\{1, 2, 3, \ldots\} \longrightarrow S$$

Thus, Dedekind infinite implies Peano infinite. ///

5. Comparison of infinities

The **Cantor-Schroeder-Bernstein Theorem** proven here is the key result that allows comparison of infinities. Perhaps it is the first serious theorem in set theory after Cantor's diagonalization argument. Apparently Cantor *conjectured* this result, and it was proven independently by F. Bernstein and E. Schröder in the 1890s. The proof given below is a *natural* proof that one might find after sufficient experimentation and reflection.

It is noteworthy that there is no invocation of the Axiom of Choice, since one can imagine that it would have been needed.

The argument below is not the most succinct possible, but is intended to lend a greater sense of inevitability to the conclusion than would the shortest possible version.

[5.0.1] **Theorem:** (*Cantor-Schroeder-Bernstein*) Let A and B be sets, with injections $f : A \longrightarrow B$ and $g : B \longrightarrow A$. Then there exists a canonical *bijection* $F : A \longrightarrow B$.

Proof: Let

$$A_o = \{a \in A : a \notin g(B)\} \qquad B_o = \{b \in B : b \notin f(A)\}$$

The sets

$$A_{2n} = (g \circ f)^n(A_o) \qquad A_{2n+1} = (g \circ f)^n g(B_o)$$

are disjoint. Let A_∞ be the complement in A to the union $\bigcup_n A_n$. Define F by

$$F(a) = \begin{cases} f(a) & (\text{for } a \in A_n,\ n \in 2\mathbb{Z}) \\ g^{-1}(a) & (\text{for } a \in A_n,\ n \in 1 + 2\mathbb{Z}) \\ f(a) & (\text{for } a \in A_\infty) \end{cases}$$

We must verify that this moderately clever apparent definition really gives a well-defined F, and that F is a bijection. For $n \geq 1$, let

$$B_n = f(A_{n-1})$$

and also let $B_\infty = f(A_\infty)$.

The underlying fact is that $A \cup B$ (disjoint union) is *partitioned* into one-sided or two-sided maximal sequences of elements that map to each other under f and g: we have three patterns. First, one may have

$$a_o \stackrel{f}{\longrightarrow} b_1 \stackrel{g}{\longrightarrow} a_1 \stackrel{f}{\longrightarrow} b_2 \stackrel{g}{\longrightarrow} a_2 \longrightarrow \ldots \stackrel{f}{\longrightarrow} b_n \stackrel{g}{\longrightarrow} a_n \longrightarrow \ldots$$

beginning with $a_o \in A_o$, all $a_i \in A$ and $b_i \in B$. Second, one may have

$$b_o \stackrel{g}{\longrightarrow} a_1 \stackrel{f}{\longrightarrow} b_1 \stackrel{g}{\longrightarrow} a_2 \stackrel{f}{\longrightarrow} b_2 \longrightarrow \ldots \stackrel{g}{\longrightarrow} a_n \stackrel{f}{\longrightarrow} b_n \longrightarrow \ldots$$

with $b_o \in B_o$, and $a_i \in A$ and $b_i \in B$. The third and last possibility is that none of the elements involved is an image of A_o or B_o under any number of iterations of $f \circ g$ or $g \circ f$. Such elements fit into pictures of the form

$$\ldots \stackrel{g}{\longrightarrow} a_{-2} \stackrel{f}{\longrightarrow} b_{-1} \stackrel{g}{\longrightarrow} a_{-1} \stackrel{f}{\longrightarrow} b_o \stackrel{g}{\longrightarrow} a_o \stackrel{f}{\longrightarrow} b_1 \stackrel{g}{\longrightarrow} \ldots$$

where $a_i \in A$ and $b_i \in B$. The fundamental point is that any two distinct such sequences of elements are disjoint. And any element certainly lies in such a sequence.

The one-sided sequences of the form

$$a_o \stackrel{f}{\longrightarrow} b_1 \stackrel{g}{\longrightarrow} a_1 \stackrel{f}{\longrightarrow} b_2 \stackrel{g}{\longrightarrow} a_2 \longrightarrow \ldots \stackrel{f}{\longrightarrow} b_n \stackrel{g}{\longrightarrow} a_n \longrightarrow \ldots$$

beginning with $a_o \in A_o$, can be broken up to give part of the definition of F by

$$F : a_o \stackrel{f}{\longrightarrow} b_1 \quad F : a_1 \stackrel{f}{\longrightarrow} b_2 \ldots$$

The one-sided sequences of the form

$$b_o \stackrel{g}{\longrightarrow} a_1 \stackrel{f}{\longrightarrow} b_1 \stackrel{g}{\longrightarrow} a_2 \stackrel{f}{\longrightarrow} b_2 \longrightarrow \ldots \stackrel{g}{\longrightarrow} a_n \stackrel{f}{\longrightarrow} b_n \longrightarrow \ldots$$

with $b_o \in B_o$, beginning with $b_o \in B_o$, can be broken up to give another part of the definition of F

$$b_o \stackrel{g}{\longrightarrow} a_1 \quad b_1 \stackrel{g}{\longrightarrow} a_2 \ldots$$

which is to say

$$F : a_1 \stackrel{g^{-1}}{\longrightarrow} b_o \quad F : a_2 \stackrel{g^{-1}}{\longrightarrow} b_1 \ldots$$

For a double-sided sequence,

$$\ldots \stackrel{g}{\longrightarrow} a_{-2} \stackrel{f}{\longrightarrow} b_{-1} \stackrel{g}{\longrightarrow} a_{-1} \stackrel{f}{\longrightarrow} b_o \stackrel{g}{\longrightarrow} a_o \stackrel{f}{\longrightarrow} b_1 \stackrel{g}{\longrightarrow} \ldots$$

there are two equally simple ways to break it up, and we choose

$$F : a_i \stackrel{f}{\longrightarrow} b_{i+1}$$

Since the sequences partition $A \cup B$, and since every element of B (and A) appears, F is surely a bijection from A to B. ///

6. *Example: transfinite Lagrange replacement*

Let V be a vector space over a field k. Let $E = \{e_\alpha : \alpha \in A\}$ be a set of *linearly independent* elements, and $F = \{f_\beta : \beta \in B\}$ be a *basis* for V.

[6.0.1] **Theorem:** We have an inequality of cardinalities: $|A| \le |B|$.

Proof: Well order[3] A. We prove by transfinite induction that there is an injection $j : A \longrightarrow B$ such that

$$\{e_\alpha : \alpha \in A\} \cup \{f_\beta : \beta \in B,\ \beta \notin j(A)\}$$

is a basis for V. That is, we can *exchange* (following Lagrange) every element in E for a basis element in F and still have a basis. Thus, since E injects to F we have an inequality of cardinalities.

Fix $\alpha \in A$. Let

$$A^{<\alpha} = \{\gamma \in A : \gamma < \alpha\}$$

For the induction step, suppose that we have an injection

$$j : A^{<\alpha} \longrightarrow B$$

such that

$$\{e_\gamma : \gamma < \alpha\} \cup \{f_\beta : \beta \notin j(A^{<\alpha})\}$$

is a disjoint union, and is still a basis for V. Then, since these elements *span* V, there exist elements a_γ and b_β in the field such that

$$e_\alpha = \sum_{\gamma<\alpha} a_\gamma \cdot e_\gamma + \sum_{\beta \notin j(A^{<\alpha})} b_\beta \cdot f_\beta$$

Since the e's were linearly independent, not all the b_βs can be 0. Pick $\beta \notin j(A^{<\alpha})$ such that $b_\beta \neq 0$, and extend j by defining $j(\alpha) = \beta$.

We must check that

$$\{e_\gamma : \gamma \le \alpha\} \cup \{f_\beta : \beta \notin j(A^{\le\alpha})\}$$

is still a basis (and that the union is disjoint). For linear independence, since

$$\{e_\gamma : \gamma < \alpha\} \cup \{f_\delta : \delta \notin j(A^{<\alpha})\}$$

is a basis, any linear relation must properly involve e_α, as

$$e_\alpha = \sum_{\gamma<\alpha} c_\gamma e_\gamma + \sum_{\delta \notin j(A^{\le\alpha})} d_\delta f_\delta$$

Replace e_α by its expression

$$e_\alpha = \sum_{\gamma<\alpha} a_\gamma \cdot e_\gamma + \sum_{\delta \notin j(A^{<\alpha})} b_\delta \cdot f_\delta$$

to obtain

$$\sum_{\gamma<\alpha} a_\gamma \cdot e_\gamma + \sum_{\delta \notin j(A^{\le\alpha})} b_\delta \cdot f_\delta + b_\beta f_\beta = \sum_{\gamma<\alpha} c_\gamma e_\gamma + \sum_{\delta \notin j(A^{\le\alpha})} d_\delta f_\delta$$

But $b_\beta \neq 0$, f_β occurs only on the left-hand side, and the vectors involved in this sum are a basis, so this is impossible. This proves the linear independence (and disjointness of the union).

[3] To well-order a set is, in effect, an invocation of the Axiom of Choice, and should not be taken lightly, even if it is useful or necessary. See the last section in this chapter.

To prove the spanning property, use the fact that

$$\{e_\gamma : \gamma < \alpha\} \cup \{f_\beta : \beta \notin j(A^{<\alpha})\}$$

is a basis. That is, given $v \in V$, there are field elements x_γ and y_δ such that

$$v = \sum_{\gamma<\alpha} x_\gamma e_\gamma + + \sum_{\delta \notin j(A^{<\alpha})} y_\delta f_\delta$$

Since $b_\beta \neq 0$ above, we can express f_β in terms of e_α, by

$$f_\beta = b_\beta^{-1} e_\alpha - \sum_{\gamma<\alpha} a_\gamma \cdot e_\gamma + \sum_{\delta \notin j(A^{\leq\alpha})} b_\delta \cdot f_\delta$$

Thus, we can replace f_β by this expression to express v as a linear combination of

$$\{e_\gamma : \gamma \leq \alpha\} \cup \{f_\beta : \beta \notin j(A^{\leq\alpha})\}$$

proving the spanning. By transfinite induction there exists an injection of A to B. ///

[6.0.2] **Remark:** We could make the invocation of Well-Ordering more explicit: if there were *no* injection $A \longrightarrow B$ as indicated, by Well-Ordering let α be the *first* element in A such that there is no such injection on $A^{<\alpha}$. Then the same discussion yields a contradiction.

We use the *Axiom of Choice* in the guise of the *Well-Ordering Principle*: we *assume* that any set can be *well-ordered*. From the theory of ordinals and well-orderings any well-ordered set is isomorphic (as well-ordered set) to a unique ordinal. From the theory of ordinals, any two ordinals are comparable, in the sense that one is an *initial segment* of the other. Thus, putting these things together, any two sets A, B are comparable in size, in the sense that either A injects to B, or B injects to A.

---

7. Equivalents of the Axiom of Choice

There are several statements which are all logically equivalent to each other, and often used to prove *existence* when only existence is required, and no object must be explicitly exhibited. These are **Zorn's Lemma**, **Hausdorff Maximality Principle**, **Well-Ordering Principle**, and **Axiom of Choice**. Here we describe these assertions in the context of *naive* set theory, in the style of the discussion above, rather than *formal* or *axiomatic* set theory. [4]

The *Axiom of Choice* or *Zermelo's postulate* asserts that, given a set of sets

$$\{S_i \ : \ i \in I\}$$

with (not necessarily mutually disjoint) non-empty sets S_i (indexed by a *set* I), there exists a set of *choices* s_i, one from each S_i. That is, there *exists* a choice set

$$C \ = \ \{s_i \ : \ i \in I\} \qquad \text{with } s_i \in S_i \text{ for all indices } i \in I$$

[4] In the late nineteenth and early twentieth centuries, it was unclear whether or not one could expect to prove these assertions from first principles. Further, some mathematicians felt that one or more of these assertions was *obviously* true, while others felt uneasy to varying degrees about invocation of them. In the early 1930's Kurt Gödel proved that the Axiom of Choice is *consistent* (in the Zermelo-Frankel first-order axiomatization) with the other axioms of set theory. In 1963, Paul Cohen proved that the Axiom of Choice was *independent* of the other axioms. In fact, Gödel also proved that the *Continuum Hypothesis* is consistent. This is the hypothesis that there are no cardinals between the countable and the cardinality of the reals. Cohen also proved that the Continuum Hypothesis is independent.

This is intuitively obvious for *finite* sets I, but less obviously clear for *infinite* sets of sets. Sometimes this is stated in the form that there is a *choice function* f on the index set I such that $f(i) \in S_i$.

The *Well-ordering Principle* asserts that every set can be well-ordered. More precisely, the assertion is that, given a set S, there is a bijection of S to an *ordinal*.

To state *Zorn's lemma* some preparation is needed. In a poset X, a *chain* is a *totally* ordered subset. An *upper bound* for a totally ordered subset Y of a poset X is an element $b \in X$ (not necessarily in Y) such that $y \leq b$ for all $y \in Y$. A *maximal* element $m \in X$ is an element of X such that, for all $x \in X$, $m \leq x$ implies $m = x$. Then Zorn's lemma asserts that *every poset in which every chain has an upper bound contains at least one maximal element.*

The *Hausdorf maximality principle* asserts that in any poset, every totally ordered subset is contained in a *maximal* totally ordered subset. Here a *maximal* totally ordered subset is what it sounds like, namely, a totally ordered subset such that any strictly larger subset fails to be totally ordered. A seemingly weaker, but equivalent, form is the assertion that *every poset contains a maximal totally ordered subset.*

We give a representative proof.

Proof: (Axiom of Choice implies the Well-ordering Principle.) Fix a set X. Let c be a choice function on the set of subsets of X. Try to define a function f on ordinals α by transfinite induction, by

$$f(\alpha) \;=\; c\big(X - \{f(\beta) : \text{ordinals } \beta < \alpha\}\big)$$

where for two sets X, A

$$X - A \;=\; \{x : x \in X,\ x \notin A\}$$

This definition *fails* if-and-when

$$X - \{f(\beta) : \text{ordinals } \beta < \alpha\}\big) \;=\; \phi$$

Let us show that each function so defined (as long as we have not run out of elements of X to hit) is *injective*. Indeed, for ordinals $\alpha > \beta$, consider the definition

$$f(\alpha) \;=\; c\big(X - \{f(\gamma) : \text{ordinals } \gamma < \alpha\}\big)$$

The set of values removed from X to choose a value for $f(\alpha)$ includes $f(\beta)$, so necessarily $f(\alpha) \neq f(\beta)$. If at any point

$$X - \{f(\beta) : \text{ordinals } \beta < \alpha\}\big) \;=\; \phi$$

then f gives a surjection from $\{\beta) :\ \beta < \alpha\}$ to X, which we have just shown is injective, giving a well-ordering of X. Thus, it suffices to show that it is *impossible* that

$$X - \{f(\beta) : \text{ordinals } \beta < \alpha\}\big) \;\neq\; \phi$$

for all ordinals α. Indeed, if this were so, then the transfinite induction proceeds uninterrupted, and we have an injective map f from *all* ordinals to X. But the collection of all ordinals is a *class*, not a set, so cannot be injected to any set, contradiction. That is, at some point the transfinite induction fails, and we have the desired well-ordering. ///

Exercises

14.[7.0.1] Show that the Well-Ordering Principle implies the Axiom of Choice.

14.[7.0.2] Show that an arbitrary poset is isomorphic, as a poset, to a set of *sets*, partially ordered by set inclusion.