

An Introduction to Axiomatic Set Theory

A variation by Arnold Neumaier of the Categorical version by Toby Bartels,
based on an original by Joseph R. Mileti*

2009 October 31

“No one shall expel us from the paradise that Cantor has created.” —David Hilbert

“I wouldn’t dream of trying to expel anyone from this paradise. I would try to do something quite different: I would try to show that it is not a paradise — so that you leave of your own accord.” —Ludwig Wittgenstein

“Few mathematicians would leave Cantor’s paradise for the vision of a hell as presented in your Categorical version, full of artificial baggage but lacking the familiar lattice structure of sets.” —Arnold Neumaier

[*** This is not my (Arnold Neumaier’s) view of set theory, but my attempt to make Toby Bartel’s categorical version intelligible to me by adapting the terminology. Traditionally, sets form a lattice under containment, a very useful structural property destroyed by the present approach. This shows that what was called “sets” by Toby Bartels has nothing to do with the usual conception of sets. Indeed, what he calls subsets are the things in his theory closest to the standard notion of a set. Therefore, in order to better match the traditional usage of terms, I replaced all occurrences of “set” by “object” and of “function” by “morphism” (which were synonymous pairs in the original), “improper set” by “set of elements”, “underlying set” by “shadow”, all occurrences of “subset of” by “set in”, and all other occurrences of “subset” by “set”. I also put the defined words in bold face for easier reading, and redefined the theme command to produce more reasonable output. I also commented on Sections 1-3, but then gave up. Too much extra shadow stuff is introduced, needed to enforce the unnatural moral code. ***]

We work in a first-order dependently typed language. (One could also use an untyped language with equality and a single ternary predicate with the intended meaning that the composite of two terms equals the third, but that would make the axioms more complicated

*The original, dated 2007 February 6, was found at <http://www.math.uchicago.edu/~milet/teaching/math278/settheory.pdf> and is used without permission or consultation. Toby reserves no legal copyright or patent rights to this work, which is intended only as an illustration of principle. [*** The principle demonstrated seems to be how to mess up mathematics by imposing unnatural moral laws. Arnold ***]

to write down and the axioms of elementary category theory nearly unrecognisable. Also, it would necessarily allow equality of sets, which is evil, although that would be fairly harmless for our purposes here.) We have: [*** I improved the formulation below ***]

- terms of type \mathcal{S} (for **objects**);
- a constant term \bullet (the **point**) of type \mathcal{S} ;
- for each pair of terms X and Y of type \mathcal{S} , terms of type $\mathcal{F}_{X,Y}$ (for **morphisms** from X to Y), and an equality predicate for terms of this type $\mathcal{F}_{X,Y}$; and
- for each triple of terms X , Y , and Z of type \mathcal{S} , term f of type $\mathcal{F}_{X,Y}$ and term g of type $\mathcal{F}_{Y,Z}$, a term $g \circ f$ (the **composite** of g after f) of type $\mathcal{F}_{X,Z}$.

When introducing a variable in these formal axioms, it is of type \mathcal{S} by default and of type $\mathcal{F}_{X,Y}$ if introduced with “: $X \rightarrow Y$ ”. We use “: X ” as an abbreviation of “: $\bullet \rightarrow X$ ”; a term of type $\mathcal{F}_{\bullet,X}$ is an **element** of X . For additional redundancy, capital letters will be used always and only [*** this is not true, see after Definition 25 ***] for variables of type \mathcal{S} . Binary operators, both logical and set-theoretic, will associate to the right; of course, set-theoretic operators bind more tightly than logical operators.

Axiom of Associativity:

$$\forall W, \forall X, \forall Y, \forall Z, \forall f: W \rightarrow X, \forall g: X \rightarrow Y, \forall h: Y \rightarrow Z, (h \circ g) \circ f = h \circ (g \circ f)$$

Axiom of Identities:

$$\forall X, \exists i: X \rightarrow X, \forall Y, (\forall f: X \rightarrow Y, f = f \circ i) \wedge \forall f: Y \rightarrow X, f = i \circ f$$

These first axioms do not use \bullet ; the language without \bullet , together with just these axioms, is the foundation of **elementary category theory**.

Axiom (Scheme) of Choice: For each formula ϕ with a variable of type $\mathcal{F}_{\bullet,X}$, a variable of type $\mathcal{F}_{\bullet,Y}$, and possibly additional variables $\vec{\chi}$ whose types may depend on X and Y , we have the axiom

$$\forall X, \forall Y, \forall \vec{\chi}, (\forall a: X, \exists b: Y, \phi(a, b, \vec{\chi})) \Rightarrow \exists f: X \rightarrow Y, \forall a: X, \phi(a, f \circ a, \vec{\chi})$$

Axiom of Extensionality:

$$\forall X, \forall Y, \forall f: X \rightarrow Y, \forall g: X \rightarrow Y, (\forall a: X, f \circ a = g \circ a) \Rightarrow f = g$$

Axiom of the Point:

$$\forall u: \bullet, \forall v: \bullet, u = v$$

Axiom of Products:

$$\begin{aligned} \forall X, \forall Y, \exists C, \exists p: C \rightarrow X, \exists q: C \rightarrow Y, \forall a: X, \forall b: Y, \exists u: C, \\ a = p \circ u \wedge b = q \circ u \wedge \forall v: C, a = p \circ v \Rightarrow b = q \circ v \Rightarrow u = v \end{aligned}$$

Axiom of Power Objects:

$$\begin{aligned} \forall X, \exists P, \exists M, \exists e: M \rightarrow X, \exists s: M \rightarrow P, \forall D, \forall i: D \rightarrow X, \\ \exists u: P, (\forall a: X, (\exists b: D, a = i \circ b) \Leftrightarrow \exists c: M, a = e \circ c \wedge u = s \circ c) \wedge \\ \forall v: P, (\forall a: X, (\exists b: D, a = i \circ b) \Leftrightarrow \exists c: M, a = e \circ c \wedge v = s \circ c) \Rightarrow u = v \end{aligned}$$

Axiom of Infinity:

$$\exists N, \exists z: N, \exists s: N \rightarrow N, \forall a: N, \neg(z = s \circ a) \wedge \forall b: N, s \circ a = s \circ b \Rightarrow a = b$$

Axiom (Scheme) of Separation: For each formula ϕ with a variable of type $\mathcal{F}_{\bullet, X}$ and possibly additional variables $\vec{\chi}$ whose types may depend on X , we have the axiom

$$\forall X, \forall \vec{\chi}, \exists S, \exists i: S \rightarrow X, (\forall a: S, \forall b: S, i \circ a = i \circ b \Rightarrow a = b) \wedge \forall a: X, \phi(a, \vec{\chi}) \Leftrightarrow \exists b: S, i \circ b = a$$

Axiom (Scheme) of Collection: For each formula ϕ with a variable of type $\mathcal{F}_{\bullet, X}$, a variable of type \mathcal{S} , and possibly additional variables $\vec{\chi}$ whose types may depend on X , we have the axiom

$$\forall X, \forall \vec{\chi}, (\forall a: X, \exists B, \phi(a, B, \vec{\chi})) \Rightarrow \exists U, \exists p: U \rightarrow X, \forall a: X, \exists B, \phi(a, B, \vec{\chi}) \wedge \exists i: B \rightarrow U, (\forall y: B, \forall z: B, i \circ y = i \circ z \Rightarrow y = z) \wedge \forall y: U, a = p \circ y \Leftrightarrow \exists z: B, y = i \circ z$$

1 Morphisms

We first establish some basic set theoretic facts carefully from the axioms.

Definition 1: An **identity** of an object X is a morphism $i: X \rightarrow X$ that satisfies the property in the Axiom of Identities.

Proposition 2: For any object A , there exists a unique identity of A , denoted id_A (or simply id if the context is clear).

Proof: By the Axiom of Identities, some identity i exists. Let j also be an identity of A ; then

$$i = i \circ j = j.$$

■

Definition 3: Let A and B be objects, and let $f: A \rightarrow B$ and $g: B \rightarrow A$ be morphisms. Then g is an **inverse** of f if $f \circ g = \text{id}_B$ and $g \circ f = \text{id}_A$.

Proposition 4: If $f: A \rightarrow B$ has an inverse, then it is unique and denoted f^{-1} .

Proof: Let g and h both be inverses of f . Then

$$g = g \circ \text{id}_B = g \circ f \circ h = (g \circ f) \circ h = \text{id}_A \circ h = h.$$

■

Lemma 5: A morphism $g: B \rightarrow A$ is an inverse of $f: A \rightarrow B$ if and only if $x = g \circ f \circ x$ for every element $x: A$ and $y = f \circ g \circ y$ for every element $y: B$.

Proof: If g is an inverse of f , then

$$x = \text{id}_A \circ x = (g \circ f) \circ x = g \circ f \circ x$$

and the condition for y is similar. For the converse,

$$\text{id}_A \circ x = x = g \circ f \circ x = (g \circ f) \circ x$$

for all $x: A$, so $\text{id}_A = g \circ f$ by Extensionality, and the condition for y is similar. ■

This lemma is how we usually prove that morphisms are inverses.

Next we break bijectivity into injectivity and surjectivity.

Definition 6: Let A and B be objects, and let $f: A \rightarrow B$. Then f is **injective** (or an **injection**) if for every element $x: A$ and element $y: A$, we have $x = y$ whenever $f \circ x = f \circ y$.

Definition 7: Let A and B be objects, and let $f: A \rightarrow B$. Then f is **surjective** (or a **surjection**) if for every element $i: B$ there is an element $j: A$ such that $i = f \circ j$.

Definition 8: Let A and B be objects, and let $f: A \rightarrow B$. Then f is **bijective** (or a **bijection**) if it has an inverse.

Definition 9: Let A and B be objects.

1. We write $A \preceq B$ to mean that there is an injection $f: A \rightarrow B$.
2. We write $A \approx B$ to mean that there is a bijection $f: A \rightarrow B$.

Proposition 10: Let A , B , and C be objects.

1. If $A \preceq B$ and $B \preceq C$, then $A \preceq C$.
2. $A \approx A$.
3. If $A \approx B$, then $B \approx A$.
4. If $A \approx B$ and $B \approx C$, then $A \approx C$.

Proposition 11: A morphism $f: A \rightarrow B$ is bijective if and only if it is both injective and surjective.

Proof: Suppose that f is bijective. Let x and y be elements of A such that $f \circ x = f \circ y$; then

$$x = \text{id}_A \circ x = (f^{-1} \circ f) \circ x = f^{-1} \circ f \circ x = f^{-1} \circ f \circ y = (f^{-1} \circ f) \circ y = \text{id}_A \circ y = y,$$

so f is injective. Let x be an element of B ; then

$$f \circ f^{-1} \circ x = (f \circ f^{-1}) \circ x = \text{id}_B \circ x = x,$$

so f is surjective.

Conversely, suppose that f is injective and surjective. Apply the Axiom of Choice to B , A , and a statement that $x = f \circ y$:

$$(\forall x: B, \exists y: A, x = f \circ y) \Rightarrow \exists g: B \rightarrow A, \forall x: B, x = f \circ g \circ x$$

Since f is surjective, the hypothesis holds, and we have g . If x is an element of B , then of course $x = f \circ g \circ x$; if y is an element of A , then $f \circ y = f \circ g \circ f \circ y$, so $y = g \circ f \circ y$ since f is injective. Therefore, g is an inverse of f . ■

The above construction of the inverse of f is an example of a general principle which we will use from now on.

Theorem 12: Morphism Comprehension Schema. *Let A and B be objects, and let $\phi(x, y, \vec{\chi})$ be a property [*** what is a property? Do you mean a formula? The same undefined term occurs elsewhere. ***] of elements of A , elements of B , and optionally other variables. Suppose that, for each element $x: A$, there exists a unique element $y: B$ such that $\phi(x, y, \vec{\chi})$ holds. Then there exists a unique morphism $f: A \rightarrow B$ such that, for every $x: A$ and $y: B$, $y = f \circ x$ if and only if $\phi(x, y, \vec{\chi})$ holds.*

Proof: Apply the Axiom of Choice to A , B , and ϕ :

$$(\forall x: A, \exists y: B, \phi(x, y, \vec{\chi})) \Rightarrow \exists f: A \rightarrow B, \forall x: A, \phi(x, f \circ x, \vec{\chi})$$

The hypothesis is satisfied, so we have f ; we must check that it has the desired property and is unique. If $y = f \circ x$, then certainly $\phi(x, y, \vec{\chi})$ holds; if, conversely, $\phi(x, y, \vec{\chi})$ holds, then $y = f \circ x$ by the condition of unicity. If g is a morphism satisfying the desired property of f and x is an element of A , then $\phi(x, g \circ x, \vec{\chi})$ holds, so $f \circ x = g \circ x$; by Extensionality, $f = g$. ■

If we have an expression $\psi(x, \vec{\chi})$ for an element of B , then let $\phi(x, y, \vec{\chi})$ state that $y = \psi(x, \vec{\chi})$; then we denote the morphism above by $(\psi(x, \vec{\chi}))_{x:A}$, in which x is a dummy variable. In particular,

$$(\psi(x, \vec{\chi}))_{x:A} \circ w = \psi(w, \vec{\chi})$$

for every element $w: A$; conversely,

$$(f \circ x)_{x:A} = f.$$

2 Cartesian Products and Ordered Pairs

We first show that the point \bullet is given by a universal property before turning our attention to products.

Definition 13: *Let A be an object.*

1. A is **inhabited** if there is an element of A .
2. A is a **subsingleton** if any two elements of A are equal.
3. A is a **singleton** if A is an inhabited subsingleton.

[*** Here would most naturally fit the later stuff about the empty object. ***] ■

Proposition 14: *The point is a singleton.*

Proof: It is inhabited by its identity id_\bullet , and the Axiom of the Point states precisely that it is a subsingleton. ■

Proposition 15: *Let A be a singleton, and let G be any object. Then there is a unique morphism from $G \rightarrow A$, denoted $!_G$ when A is the point.*

Proof: Apply Morphism Comprehension to G , A , and an always true statement; this applies precisely because A is a singleton. ■

This has an immediate converse, taking G to be the point.

Proposition 16: *If A and B are singletons, then there is a unique bijection from A to B .*

Proof: Since B is a singleton, we have a morphism $f: A \rightarrow B$; since A is a singleton, we have a morphism $g: B \rightarrow A$. Since a morphism from A to A must be unique, $g \circ f = \text{id}_A$; similarly, $f \circ g = \text{id}_B$. Therefore, f is a bijection. Since any morphism from A to B must be unique, certainly this is true of bijections. ■

Because of this, we speak of “the” singleton, although a complete justification of that terminology relies on a metatheorem that we will not prove here. [*** but may be at least state it, so that one knows what was meant? ***]

Definition 17: *Let A and B be objects. A **Cartesian product** of A and B is any object C , called the **shadow** of the product, together with morphisms $p: C \rightarrow A$ and $q: C \rightarrow B$, called the **projections** of the product on A and B , such that, for every element $x: A$ and $y: B$, there exists a unique $u: C$ such that $x = p \circ u$ and $y = q \circ u$.*

Note that a Cartesian product is not a single object in our language, but rather three objects: an object and two morphisms. So when we write “for every Cartesian product” or “there is a Cartesian product”, formally this involves three quantifiers. It is a common abuse of terminology, however, to refer to the shadow as the Cartesian product itself and suppress mention of the projections.

Proposition 18: *Let A and B be objects. Then there exists a Cartesian product of A and B .*

Proof: Apply the Axiom of Products to A and B . ■

It is convenient, in any particular context, to pick a Cartesian product if needed, once and for all within that context. [*** But in Proposition 21, more than one is needed. Strange moral, made necessary by having too many copies of everything around. ***] Then the shadow will be denoted $A \times B$, with the projections denoted $\pi_{A,B}: A \times B \rightarrow A$ and $\rho_{A,B}: A \times B \rightarrow B$; we suppress the subscripts when the context makes them clear. An alternative formulation of the axioms of object theory could make these basic terms of the language, as with \bullet . Conversely, \bullet could be removed from the language by making the Axiom of the Point an existential statement and beginning every other axiom (except Associativity and Identities) with it.

Cartesian products are also given by a universal property.

Proposition 19: *Let A , B , and G be objects. Then, for every Cartesian product $A \times B$ of A and B , and [*** The original formulation said: “fix a Cartesian product of A and B ”. I thought this was automatically fixed once and for all, by your moral. Why then fix it again? I improved the present formulation. But later I didn’t fix the formulation. ***] given morphisms $x: G \rightarrow A$ and $y: G \rightarrow B$, there is a unique morphism from G to $A \times B$, denoted (x, y) , such that $\pi \circ (x, y) = x$ and $\rho \circ (x, y) = y$.*

Proof: Apply Morphism Comprehension to G , $A \times B$, and a statement that the values in A and B match:

$$\pi \circ u = x \circ i \wedge \rho \circ u = y \circ i$$

(where i is an arbitrary element of G and u is the desired unique element of $A \times B$). The Axiom of Products, applied to $x \circ i$ and $y \circ i$, gives us a unique u satisfying the condition above, so we have a unique morphism $(x, y): G \rightarrow A \times B$ such that, for every $i: G$, we have $\pi \circ (x, y) \circ i = x \circ i$ and $\rho \circ (x, y) \circ i = y \circ i$. By Associativity and Extensionality, $\pi \circ (x, y) = x$ and $\rho \circ (x, y) = y$, as desired. Now let $e: G \rightarrow A \times B$ be any morphism such that $\pi \circ e = x$ and $\rho \circ e = y$. Given any element $i: G$, we have $\pi \circ e \circ i = (\pi \circ e) \circ i = x \circ i$ and similarly $\rho \circ e \circ i = y \circ i$, so $e \circ i$ satisfies the requirement for u in the Axiom of Products. But the same requirement is met by $(x, y) \circ i$, so $(x, y) \circ i = e \circ i$; by Extensionality, $(x, y) = e$. ■

Conversely, any $A \times B$, π , and ρ that satisfy this universal property comprise a product of A and B , as is immediately seen by taking G to be the point. In this case, we have the pairing $(x, y): A \times B$ of elements $x: A$ and $y: B$.

Corollary 20: *Given $x: G \rightarrow A$, $y: G \rightarrow B$, $c: G \rightarrow A$, and $d: G \rightarrow B$ and fixing a Cartesian product of A and B , we have $(x, y) = (c, d)$ if and only if $x = c$ and $y = d$.*

Proof: If $(x, y) = (c, d)$, then of course $x = \pi(x, y) = \pi(c, d) = c$ and $y = d$ similarly. If $x = c$ and $y = d$, then (c, d) satisfies the defining property of (x, y) , so they are equal by the preceding result. ■

Proposition 21: *Let A and B be objects, and fix a Cartesian product of A and B . If C , with $p: C \rightarrow A$ and $q: C \rightarrow B$, is also a Cartesian product of A and B , then there is a unique bijection $f: C \rightarrow A \times B$ such that $p = \pi \circ f$ and $q = \rho \circ f$.*

Proof: We already have a unique morphism f , which is (p, q) ; we need only check that f is a bijection. Given elements x and y of C such that $f \circ x = f \circ y$,

$$p \circ x = (\pi \circ f) \circ x = \pi \circ f \circ x = \pi \circ f \circ y = (\pi \circ f) \circ y = p \circ y$$

and similarly $q \circ x = q \circ y$; since C , with p and q , is a Cartesian product, we have $x = y$; therefore, f is injective. Given an element x of $A \times B$, we have (since C is a Cartesian product) an element $y: C$ such that $p \circ y = \pi \circ x$ and $q \circ y = \rho \circ x$. This is the defining property of $f \circ y$, so $x = f \circ y$; therefore, f is surjective. ■

Because of this, we also speak of “the” Cartesian product of A and B . [*** This introduces an abuse of notation. Are we still allowed to consider two distinct Cartesian products, as in Proposition 21, or is this against the moral? ***]

3 Sets

If U is an object, then we may wish to consider a “set” in U , that is object which is only “part” of U , whose elements correspond to some but not all of the elements of U . (Of course, we will also accept the extreme cases of “none” and “all” in addition to “some but not all”.) The simplest way to specify such a thing is by a morphism $i: A \rightarrow U$; the elements of U that we want are those that i assigns to some element of A . It is possible to *define* a set to be such a morphism (or an injective such morphism so that the corresponding element of A is unique); however, there are two ways in which this is inconvenient: we wish to compare sets (considering them equal if the same elements of U belong to them), but we have no direct way to compare morphisms with different sources; and we wish to gather all sets in U into a single “object of all sets” in U , but objects and most morphisms are not elements of objects in our formalism. It is possible to get around the former inconvenience (by simply defining what we mean for two sets to be “equal”), but we need the Axiom of Power Objects to avoid the latter inconvenience; and having done so, it is simplest to define a set in U to *be* an element of the object of all sets in U . [*** namely the power object ***]

Definition 22: *Let U be an object. A **power object** of U is an object P (the **shadow** of the power object), together with a [*** dummy ***] object M and a [*** dummy ***] morphism $e: M \rightarrow U \times P$, such that, given any object A and morphism $i: A \rightarrow U$, there is a unique element b of P such that, for all $x: U$, there exists an element of A mapped to x by i if and only if there exists an element of M mapped to (x, b) by e .*

[*** It seems that a power object is not a typed object but a composite thing that cannot be expressed as an object in your formal language. ***]

In other words, just as $i: A \rightarrow U$ specifies a part of U as described earlier, so $e: M \rightarrow U \times P$ specifies a part of $U \times P$; an element b of P also specifies a part of U , consisting of those $x: U$ such that (x, b) belongs to the part of $U \times P$ specified by e , and we want a unique such element of P to correspond to any part of U given by any $i: A \rightarrow U$. That is the idea behind this definition.

Proposition 23: *Let U be an object. Then there exists a power object of U .*

Proof: Apply the Axiom of Power Objects to U . The e in the definition above is constructed by pairing from the e and s in the axiom. ■

The same formal remarks apply to power objects as apply to Cartesian products. In particular, we find it convenient to pick a power object of U , denote its shadow by $\mathcal{P}(U)$, and refer to that object itself as the power object by abuse of terminology. [*** This definition of an abused power object is far removed from intuition, but required by the unnatural moral. The naturality of ordinary set theory is completely gone! ***]

Definition 24: *Let U be an object and fix a power object $\mathcal{P}(U)$ of U . A **set** in U is an element of $\mathcal{P}(U)$. [*** We may call U the **context** of S . ***]*

[*** It seems that the concept of a set in U is already ambiguous, depending on the choice of a power object. Very unnatural, and difficult for an automatic reader! In particular, sets corresponding to different power objects of the same object U may have the same elements, against your moral! ***]

There is usually no need to refer to M and e , [*** which proves that your moral generates lots of artificial baggage ***] except through the following concept:

Definition 25: Let U be an object and fix a power object $\mathcal{P}(U)$. If x is an element of U and A is a set in U , then x **belongs to** A , denoted $x \in A$, if there exists an element of M that is assigned (x, A) by e (where M and e are as in the definition of power object above).

[*** Suppose I want to create the set $A \times B$ inside of $U \times V$. It requires a lot of extra baggage to do that in your morally pure environment. I guess, few ordinary mathematicians will like your moral code. ***]

Here we have introduced a font shift; while before now we have used lowercase letters for elements and other morphisms but uppercase letters only for objects, now [*** against the earlier injunction! ***] we are using an uppercase letter for an element of a power object. In general, we will use uppercase letters for morphisms to power objects (including elements of them), and even fancier fonts for elements of power objects of power objects and the like. This notation reminds us that a set corresponds to an entire object (equipped with a morphism to U), as follows:

Definition 26: Let U be an object, and let A be a set in U . Then a **presentation** of A consists of an object S (called the **shadow**) and an injective morphism $i: S \rightarrow U$ (called the **inclusion**) such that, for every $x: U$, $x \in A$ if and only if there exists an element of S that is assigned x by i .

[*** Thus, a set **has** a shadow, and **belongs to** its context. ***]

Proposition 27: Let U be an object, and let A be a set in U . Then there exists a presentation of A .

Proof: Apply the Axiom of Separation to U and a statement that x belongs to A :

$$\exists S, \exists i: S \rightarrow U, (\forall a: S, \forall b: S, i \circ a = i \circ b \Rightarrow a = b) \wedge \forall a: X, x \in A \Leftrightarrow \exists b: S, i \circ b = a$$

This is precisely the result that we want. ■

We normally pick a presentation for each set once and for all. [*** This seems to require the axiom of **global** choice! ***] The shadow of the presentation of A may be denoted $|A|$, but it is more common to abuse notation by denoting it A again. [*** But then $x \in A$ iff $x : A$ iff $x : |A|$ iff $x \in ||A|$, although the first and the last statement in this chain are not equivalent. Thus you generate harmful ambiguity! ***] Since the actual A is a morphism (an element of $\mathcal{P}(U)$) while $|A|$ is an object, there should be no confusion. We will denote the inclusion as $\iota_A: |A| \rightarrow U$, or simply $\iota_A: A \rightarrow U$ if we abuse notation.

Proposition 28: Let U be an object and A a set in U ; fix a presentation of A . For each element x of U , if $x \in A$, then there exists a unique element c of $|A|$ such that $x = \iota_A \circ c$.

Proof: By definition of presentation, there exists such a c . It is unique because ι_A is an injection. ■

We may denote this unique element of $|A|$ by x^A , but it is more common to refer to it as x again. Together, the above abuses of notation allow us to avoid referring explicitly to presentations of sets, instead thinking of a set in U as an object whose elements simply “are” some of the elements of U . (Actually, since the typing declaration “ $: A \rightarrow U$ ” has no literal meaning when A is not literally an object, this is not really an abuse of notation at all if we simply *define* it to refer to $|A|$ instead. Similarly, any reference to $x:U$ as if it were an element of $|A|$ has no literal meaning, so we may define it to refer to x^A when that exists. The same goes for the other alleged abuses of notation introduced below.)

Theorem 29: Set Comprehension Schema. *Let U be an object and let $\phi(x, \vec{x})$ be a property of elements of U and optionally other variables. Then there exists a unique set A of U such that, for every element y of U , $y \in A$ if and only if $\phi(y, \vec{x})$ holds.*

Proof: Apply the Axiom of Separation to U and ϕ :

$$\exists S, \exists i: S \rightarrow U, (\forall a: S, \forall b: S, i \circ a = i \circ b \Rightarrow a = b) \wedge \forall a: X, \phi(x, \vec{x}) \Leftrightarrow \exists b: S, i \circ b = a$$

Then apply the defining property of $\mathcal{P}(U)$ to S and i to get the desired result. ■

We denote this set A by $\{x:U \mid \phi(x, \vec{x})\}$, in which x is a dummy variable. That is,

$$y \in \{x:U \mid \phi(x, \vec{x})\} \Leftrightarrow \phi(y, \vec{x});$$

conversely,

$$\{x:U \mid x \in A\} = A.$$

Definition 30: *Let U be an object. The set of elements of U is $\{x:U \mid \top\}$, where \top is a universally true statement.*

It is common to abuse notation by referring to the set of elements in U as U again. To keep our abuses from conflicting, we choose a presentation whose shadow really is U : [*** Everything seems to be choice- and presentation-dependent. Are you sure that one can always make these choices consistently without running into conflicts? It seems to me that you get now the same conflict as pointed out earlier. ***]

Proposition 31: *The identity $\text{id}:U \rightarrow U$ is a presentation of the set of elements in U .*

Proof: For every $x:U$, $\text{id} \circ x = x$. ■

Definition 32: *Let U be an object, and let A and B be sets in U . Then A is **contained** in B , denoted $A \subseteq B$, if every element of U [*** “of U ” is superfluous ***] that belongs to A also belongs to B .*

Proposition 33: *Let A be a set in an object U . Then A is contained in the set of elements in U ; that is, $A \subseteq U$.*

Proof: If $x:U$ such that $x \in A$, then of course $x \in U$ since that is always true. ■

Proposition 34: (extensionality for sets) *Let A and B be sets in an object U . Then $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.*

Proof: If $A = B$, then of course $x \in A$ if and only if $x \in B$. Conversely, suppose that $A \subseteq B \subseteq A$. Since $x \in A$ if and only if $x \in B$, we have

$$A = \{x:U \mid x \in A\} = \{x:U \mid x \in B\} = B.$$

[*** I moved a remark from here into the above proposition ***] ■

Definition 35: *Let U be an object. The **empty set** in the context U , denoted \emptyset_U , is*

$$\{x:U \mid \perp\},$$

where \perp is a universally false statement.

[*** Actually, the empty set depends both on the choice of U and of its power object, but the notation does not reflect this. Unlike Bourbaki, you need abuses of language and notation from the beginning! ***] The empty set may also be denoted \emptyset when the context is clear. ■

Proposition 36: *The shadow of the empty set in any context U has no elements.*

Proof: If it has an element x , then we have $\iota \circ x \in \emptyset_U$, which is false. ■

We can pick some object, such as the point, and let “the” **empty object**, denoted \emptyset (with no subscript), be the shadow of its empty set. [*** why pick here a particular object, if elsewhere your moral insists on having everything ambiguous? ***] ■

Definition 37: *Let U be an object, and let a be an element of U . Then the **singleton set** $\{a\}$ is*

$$\{x:U \mid x = a\}.$$

If a and b are elements of U , the **unordered pair** $\{a, b\}$ is

$$\{x:U \mid x = a \vee x = b\}.$$

And so on.[*** The phrase “and so on” is formally undefined. You haven’t natural numbers yet! ***] ■

Proposition 38: *Let a be an element of an object U . Then $\{a, a\} = \{a\}$.*

Proof: Given an element x of U , $x \in \{a, a\}$ if and only if $x = a$ or $x = a$, which happens if and only if $x = a$, which happens if and only if $x \in \{a\}$. By extensionality of sets, $\{a, a\} = \{a\}$. ■

Proposition 39: *Let a and b be elements of an object U . Then $\{a, b\} = \{b, a\}$.*

Proof: Given an element x of U , $x \in \{a, b\}$ if and only if $x = a$ or $x = b$, which happens if and only if $x = b$ or $x = a$, which happens if and only if $x \in \{b, a\}$. By extensionality of sets, $\{a, b\} = \{b, a\}$. ■

Definition 40: Let U be an object, and let A be a set in U . The **complement** of A , denoted $-A$, is

$$\{x: U \mid \neg x \in A\}.$$

Definition 41: The **relative complement** of A in B , denoted $B - A$, is $-A \cap B$.

Recall that a set in U is an element of $\mathcal{P}(U)$. We can also consider sets in $\mathcal{P}(U)$.

Definition 42: Let U be an object. A **collection of sets** of U is an element of $\mathcal{P}(\mathcal{P}(U))$.

Definition 43: Let U be an object, and let \mathcal{S} be a collection of sets in U . The **intersection** of \mathcal{S} , denoted $\bigcap \mathcal{S}$, is

$$\{x: U \mid \forall A: \mathcal{P}(U), A \in \mathcal{S} \Rightarrow x \in A\}.$$

Dually, the **union** of \mathcal{S} , denoted $\bigcup \mathcal{S}$, is

$$\{x: U \mid \exists A: \mathcal{P}(U), A \in \mathcal{S} \wedge x \in A\}.$$

Proposition 44: 1. $\bigcap \mathcal{P}(U) = \emptyset_U$.

2. $\bigcup \mathcal{P}(U) = U$.

3. $\bigcap \emptyset_{\mathcal{P}(U)} = U$.

4. $\bigcup \emptyset_{\mathcal{P}(U)} = \emptyset_U$.

Definition 45: Let f be a morphism from A to B . The **range** of f is a set in B :

$$\text{ran } f = \{y: B \mid \exists x: A, y = f \circ x\}.$$

While we are discussing power objects, let us use them to perform another important construction on objects: the disjoint union.

Definition 46: Let A and B be objects. A **disjoint union** of A and B is any object C , called the **shadow** of the disjoint union, together with injective morphisms $i: A \rightarrow C$ and $j: B \rightarrow C$, called the **natural inclusions** of A and B into the disjoint union, such that, for every element $z: C$, either there exists an element $x: A$ such that $z = i \circ x$ or there exists an element $y: B$ such that $z = j \circ y$, but not both.

Like a Cartesian product (and similar to a power object), a disjoint union is not a single object but an object and two morphisms. Unlike Cartesian products and power objects, we have not included an axiom stating the existence of disjoint unions. This is because we can already prove it.

Proposition 47: Let A and B be objects. Then there exists a disjoint union of A and B .

Proof: Consider the Cartesian product of the power objects, and consider its set

$$\{z: \mathcal{P}(A) \times \mathcal{P}(B) \mid (\exists x: A, z = (\{x\}, \emptyset_B)) \vee \exists y: B, z = (\emptyset_A, \{y\})\}.$$

Let C be the shadow of a presentation of this set. We define $i: A \rightarrow C$ and $j: B \rightarrow C$ respectively as $((\{x\}, \emptyset_B)^C)_{x:A}$ and $((\emptyset_A, \{y\})^C)_{y:B}$. If $(\{x\}, \emptyset)_B = (\{y\}, \emptyset)_B$, then $\{x\} = \{y\}$ and then $x = y$, so i is injective (and similarly for j). Given $z: C$, either $z = (\{x\}, \emptyset_B)$ or $z = (\emptyset_A, \{y\})$, so all that remains is to prove that z cannot be both at once. But if it is, then $\{x\} = \emptyset_A$, so $x \in \emptyset_A$, which is false. ■

There are other ways to prove that a disjoint union of A and B exists; however we do it, we pick one and denote its shadow $A \uplus B$. The natural inclusions may be denoted $\iota_{A,B}: A \rightarrow A \uplus B$ and $\kappa_{A,B}: B \rightarrow A \uplus B$.

Proposition 48: *Let A and B be objects, and consider the ranges of $\iota_{A,B}$ and $\kappa_{A,B}$, which are sets in $A \uplus B$.*

1. $\text{ran } \iota \cap \text{ran } \kappa = \emptyset$.
2. $\text{ran } \iota \cup \text{ran } \kappa = A \uplus B$.

[*** \cap and \cup are undefined ***] ■

Proof: 1. Let z be an element of $A \uplus B$. If $z \in \text{ran } \iota \cap \text{ran } \kappa$, then $z = \iota \circ x$ for some $x: A$ and $z = \kappa \circ y$ for some $y: B$, but this is impossible. Thus, $\text{ran } \iota \cap \text{ran } \kappa = \emptyset$.
 2. Let z be an element of $A \uplus B$. Then either $z = \iota \circ x$ for some $x: A$ or $z = \kappa \circ y$ for some $y: B$; either way, $z \in \text{ran } \iota \cup \text{ran } \kappa$. Thus, $\text{ran } \iota \cup \text{ran } \kappa = A \uplus B$. ■

This explains the origins of the term “disjoint union”. [*** a long, artificial road to what takes a few lines in standard set theory! ***] ■

4 Relations

Definition 49: *Let A and B be objects, and fix a Cartesian product $A \times B$. A **binary relation**, or simply **relation**, between A and B is a set in $A \times B$. A relation on A is a relation between A and itself, that is a set in $A \times A$. Given a relation R between A and B , if $x: A$ and $y: B$ are such that $(x, y) \in R$, then we write $x \sim_R y$ or simply $x R y$.*

We typically construct relations using a modified form of Set Comprehension:

Theorem 50: Relation Comprehension Schema. *Let A and B be objects and let $\phi(x, y, \vec{x})$ be a property of elements of A , elements of B , and optionally other variables. Then there exists a unique relation R between A and B such that, for every element x of A and y of B , $x R y$ if and only if $\phi(x, y, \vec{x})$ holds.*

Proof: Let R be

$$\{z: A \times B \mid \phi(\pi \circ z, \rho \circ z, \vec{x})\}.$$

The rest reduces to Set Comprehension. ■

We denote this relation R by $\{(x, y): A \times B \mid \phi(x, y, \vec{x})\}$, in which x and y are dummy variables. That is,

$$(u, v) \in \{(x, y): A \times B \mid \phi(x, y, \vec{x})\} \Leftrightarrow \phi(u, v, \vec{x});$$

conversely,

$$\{(x, y): A \times B \mid x R y\} = R.$$

As relations are a special case of sets, all of the operations (intersection, union, etc) that apply to sets also apply to relations. However, we also have some additional notions.

Definition 51: Let R be a relation between A and B .

1. The **domain** of R is a set in A :

$$\text{dom } R = \{x: A \mid \exists y: B, x R y\}.$$

2. The **range** of R is a set in B :

$$\text{ran } R = \{y: B \mid \exists x: A, x R y\}.$$

Here are some important special kinds of relations:

Definition 52: Let A and B be objects, and let R be a relation between A and B .

1. R is **morphismal** if, for all elements $x: A$ and $y, z: B$, $y = z$ if $x R y$ and $x R z$.
2. R is **entire** if, for every element $x: A$, there is an element $y: B$ such that $x R y$.

Morphismal entire relations correspond precisely to morphisms:

Definition 53: Let A and B be objects, and let $f: A \rightarrow B$ be a morphism. Then the **graph** of f , denoted Γ_f , is

$$\{(x, y): A \times B \mid y = f \circ x\}.$$

Theorem 54: 1. The graph of every morphism is a morphismal entire relation.
2. Every morphismal entire relation is the graph of a unique morphism.

Proof: 1. Let $f: A \rightarrow B$ be a morphism. If $x \Gamma_f y$ and $x \Gamma_f z$, then $y = f \circ x$ and $z = f \circ x$, so $y = z$. Given any $x: A$, $x \Gamma_f f \circ x$ because $f \circ x = f \circ x$.
2. Let $R: \mathcal{P}(A \times B)$ be a morphismal entire relation. Given $x: A$, there is a unique $y: B$ such that $x R y$; by Morphism Comprehension, there is a unique morphism $f: A \rightarrow B$ such that $y = f \circ x$ iff $x R y$. In other words, there is a unique morphism $f: A \rightarrow B$ such that $\Gamma_f = R$. ■

Proposition 55: Let f be a morphism from A to B . Then the range of f equals the range of the graph of f .

Proof: Let y be an element of B . If $y \in \text{ran } f$, then $y = f \circ x$ for some $x: A$, so $x \Gamma_f y$; thus, $y \in \text{ran } \Gamma_f$. Conversely, if $y \in \text{ran } \Gamma_f$, then $x \Gamma_f y$ for some $x: A$, so $y = f \circ x$; thus, $y \in \text{ran } f$. ■

Proposition 56: *Let A and B be objects, and let R be an entire relation between A and B . Then there exists a morphism $f: A \rightarrow B$ such that $\Gamma_f \subseteq R$.*

Proof: Apply the Axiom of Choice to A , B , and a statement that $x R y$:

$$(\forall x: A, \exists y: B, x R y) \Rightarrow \exists f: A \rightarrow B, \forall x: A, x R y$$

The hypothesis is satisfied because R is entire, so we have f . If $x \Gamma_f y$, then $y = f \circ x$ and so $x R y$; thus, $\Gamma_f \subseteq R$. ■

Note that the morphism provided by this proposition is not unique. This proof is our *only* application of the Axiom of Choice other than through Morphism Comprehension. It is possible to modify the Axiom of Choice to the Axiom of Unique Choice, in which the b that appears in the hypothesis is required to be unique. Then Morphism Comprehension can still be proved, but the previous proposition cannot; however, adopting this proposition as an additional axiom allows one to prove the more general Axiom of Choice. (Thus, this proposition may also be called the “axiom of choice”.)

Definition 57: *Let R be a relation on an object A .*

1. R is **reflexive** if for all $x: A$, we have $x R x$.
2. R is **symmetric** if for all $x, y: A$, if $x R y$ then $y R x$.
3. R is **asymmetric** if for all $x, y: A$, if $x R y$ then it is not the case that $y R x$.
4. R is **antisymmetric** if for all $x, y: A$, if $x R y$ and $y R x$, then $x = y$.
5. R is **transitive** if for all $x, y, z: A$, if $x R y$ and $y R z$, then $x R z$.
6. R is **connected** if for all $x, y: A$, either $x R y$, $x = y$, or $y R x$.

Definition 58: *Let A be an object.*

1. A **partial ordering** on A is a relation on A that is transitive and asymmetric.
2. A **linear ordering** on A is a partial ordering on A that is connected.
3. A **well-ordering** on A is a linear ordering on A such that for every inhabited set S of A , there exists $m: S$ such that for all $x: S$, either $m = x$ or $m R x$.

Incidentally, some parts of the last two definitions are not really what we want if we use intuitionistic logic, but we won’t worry about that here.

5 The Natural Numbers and Induction

We have not yet used the Axiom of Infinity.

Definition 59: *A **Dedekind system** consists of an object I , an injective morphism $s: I \rightarrow I$, and an element $z: I$ such that $z \notin \text{ran } s$.*

Proposition 60: *There exists a Dedekind system.*

Proof: This is precisely what the Axiom of Infinity states. ■

Definition 61: Let I (with $z: I$ and $s: I \rightarrow I$) be a Dedekind system, and let A be a set in I . We say that A is **inductive** if $z \in A$ and, whenever $x \in A$, $s \circ x \in A$.

Proposition 62: In any Dedekind system, the set of elements is inductive.

That is obvious; the real point is the converse:

Definition 63: A **Peano system** is a Dedekind system whose only inductive set is the set of elements.

Theorem 64: There exists a Peano system.

Proof: Let I be any Dedekind system. Let \mathcal{J} be the collection of inductive sets in I (given by the Axiom of Separation). Let N be the intersection $\bigcap \mathcal{J}$; we will give N (or rather $|N|$, to be pedantic) the structure of a Peano system. Since $z \in A$ for every $A \in \mathcal{J}$, we have $z \in N$. Given $x \in N$, $x \in A$ for every $A \in \mathcal{J}$, so $s \circ x \in A$ for every such A , proving that $s \circ x \in N$. Therefore, z (pedantically, z^N) and $(s \circ x)_{x \in N}$ (pedantically, $((s \circ \iota_N \circ x)^N)_{x:|N|}$) make N into a Dedekind system.

Now suppose that A is an inductive set in N ; let B be the set $\{x: I \mid x \in N \wedge x^N \in A\}$ of I . We claim that B is an inductive set in I : we have $z \in B$ because $z \in N$ and $z^N \in A$; if $x \in B$, then $x^N \in A$ and then $(s \circ x)^N \in A$, so $s \circ x \in B$. Therefore, $B \in \mathcal{J}$, so $N \subseteq B$. Given $x: N$, we have $\iota_N \circ x \in B$, so $x \in A$; therefore, A is the set of elements in N , as desired. ■

Definition 65: We now pick a Peano system; we denote its shadow \mathbf{N} . A **natural number** is an element of \mathbf{N} . The natural number **zero**, denoted $0_{\mathbf{N}}$ or simply 0 , is the element of \mathbf{N} given by its chosen structure as a Dedekind system; similarly, the **successor morphism**, denoted $S_{\mathbf{N}}$ or simply S , is the chosen morphism $\mathbf{N} \rightarrow \mathbf{N}$. If x is a natural number, then the **successor** of x is $S \circ x$; we also define 1 as $S \circ 0$, 2 as $S \circ 1$, etc.

If we apply the definition of a Peano system with this notation, we have:

Proposition 66:

1. If $S \circ x = S \circ y$, then $x = y$.
2. Zero is not the successor of any natural number.
3. Suppose that A is a set in \mathbf{N} , $0 \in A$, and for all $x: \mathbf{N}$, if $x \in A$ then $S \circ x \in A$. We then have $A = \mathbf{N}$.

We think that \mathbf{N} (with zero and the successor operation) captures our intuitive idea of the system of natural numbers, and it is now our goal to show how to prove the basic statements about the natural numbers which are often accepted axiomatically. We first define a relation $<$ (along with a few related relations) on \mathbf{N} .

Definition 67: Let A be a set in \mathbf{N} . We say that A is an **upper set** of \mathbf{N} if, whenever $x \in A$, then $S \circ x \in A$.

Definition 68: Let i and j be natural numbers. We define four relations on \mathbf{N} through Relation Comprehension as follows:

1. We say that i is **less than** j , denoted $i <_{\mathbf{N}} j$ or simply $i < j$, if there is an upper set in \mathbf{N} to which j belongs but i does not.
2. We say that i is **at most** j , denoted $i \leq_{\mathbf{N}} j$ or simply $i \leq j$, if j belongs to every upper set in \mathbf{N} to which i belongs, that is if $j < i$ fails.
3. We say that i is **greater than** j , denoted $i >_{\mathbf{N}} j$ or simply $i > j$, if there is an upper set in \mathbf{N} to which i belongs but j does not, that is if $j < i$ holds.
4. We say that i is **at least** j , denoted $i \geq_{\mathbf{N}} j$ or simply $i \geq j$, if i belongs to every upper set in \mathbf{N} to which j belongs, that is if $i < j$ fails.

Lemma 69:

1. There is no $k: \mathbf{N}$ with $k < 0$; that is, $0 \leq k$ for every $k: \mathbf{N}$.
2. There is no $k: \mathbf{N}$ with $S \circ k \leq 0$; that is, $0 < S \circ k$ for every $k: \mathbf{N}$.
3. We always have $i < j$ if and only if $S \circ i < S \circ j$.

Proof:

1. Let A be an upper set in \mathbf{N} . If $0 \in A$, then A is an inductive set in \mathbf{N} , so A is the set of elements, so $k \in A$.
2. Let A be the set in \mathbf{N} consisting of all successors; that is, $A = \{j: \mathbf{N} \mid \exists i: \mathbf{N}, j = S \circ i\}$. Given $k: \mathbf{N}$, we have $S \circ k \in A$ regardless of whether $k \in A$, so A is certainly an upper set. Since $0 = S \circ i$ is never true, we have A as an upper set to which any $S \circ k$ belongs but 0 does not.
3. Let A be an upper set in \mathbf{N} to which j belongs but i does not. Let B be

$$\{k: \mathbf{N} \mid \exists l: \mathbf{N}, l \in A \wedge k = S \circ l\}.$$

Then B is an upper set to which $S \circ j$ belongs but $S \circ i$ does not. (We must use that S is injective here.) Conversely, let B be an upper set in \mathbf{N} to which $S \circ j$ belongs but $S \circ i$ does not. Let A be $\{k: \mathbf{N} \mid S \circ k \in B\}$. Then A is an upper set to which j belongs but i does not. ■

Our primary objective is to show that $<$ is a well-ordering on \mathbf{N} . Due to the nature of the definition of \mathbf{N} , it seems that only way to prove nontrivial results about \mathbf{N} is “by induction”. We state the Step Induction Principle in two forms. The first is much cleaner and seemingly more powerful (because it immediately implies the second and we can quantify over objects but not over formulas), but the second is how one often thinks about induction is used in practice (using “properties” of natural numbers) and will be the only form that we can generalize to the collection of all ordinals.

Theorem 70: Step Induction Schema.

1. Suppose that A is a set in \mathbf{N} , $0 \in A$, and for all $x: \mathbf{N}$, if $x \in A$ then $S \circ x \in A$. We then have $A = \mathbf{N}$.
2. For any formula $\phi(x, \vec{\chi})$ with an element x of \mathbf{N} and optionally other free variables, we have

$$\forall \vec{\chi}, \phi(0, \vec{\chi}) \Rightarrow (\forall x: \mathbf{N}, \phi(x, \vec{\chi}) \Rightarrow \phi(S \circ x, \vec{\chi})) \Rightarrow \forall x: \mathbf{N}, \phi(x, \vec{\chi}).$$

[*** The formula needs an extra pair of parentheses ***] ■

Proof: 1. We have already seen this.

2. Fix the $\vec{\chi}$, and suppose $\phi(0, \vec{\chi})$ and $\forall x: \mathbf{N}, \phi(x, \vec{\chi}) \Rightarrow \phi(S \circ x, \vec{\chi})$. Let A be $\{x: \mathbf{N} \mid \phi(x, \vec{\chi})\}$. Notice that $0 \in A$ and for all $x: \mathbf{N}$, if $x \in A$ then $S \circ x \in A$ by assumption. It follows from part 1 that $A = \mathbf{N}$. Therefore, we have $\forall x: \mathbf{N}, \phi(x, \vec{\chi})$. ■

With the Step Induction Principle in hand, we can begin to prove the basic facts about the natural numbers. Our goal is to prove that $<$ is a well-ordering on \mathbf{N} , but it will take some time to get there. We first give a very simple inductive proof. For this proof only, we will give careful arguments using both versions of Step Induction to show how a usual induction proof can be formalized in either way.

Lemma 71: *There is no $k: \mathbf{N}$ with $k < k$; that is, $k \leq k$ for every $k: \mathbf{N}$.*

Proof: This is obvious from the definition, but we will give two proofs corresponding to the above two versions of the Induction Principle.

1. Let X be $\{k: \mathbf{N} \mid k \leq k\}$, and notice that $0 \in X$ since $0 \leq k$ is always true. Suppose now that $k \in X$, so $k < k$ fails. If $S \circ k < S \circ k$, then $k < k$, which is false, so $S \circ k \in X$. Thus, by Step Induction, we have $X = \mathbf{N}$. Therefore, for all $k: \mathbf{N}$, we have $k \leq k$.
2. Let $\phi(k)$ be the statement that $k \leq k$. We clearly have $\phi(0)$ because $0 \leq 0$. Suppose now for $k: \mathbf{N}$ that $k < k$ fails. If $S \circ k < S \circ k$, then $k < k$, which is false. It follows that $\phi(S \circ k)$ holds. Therefore, by Step Induction, we have $k \leq k$ for all $k: \mathbf{N}$. ■

Our remaining inductive proofs will be given in a more natural relaxed style.

Proposition 72: *$<_{\mathbf{N}}$ is transitive.*

Proof: We prove the result by induction on the third variable. First fix $i, j: \mathbf{N}$ and suppose that $i < j$ and $j < 0$. We know that $j < 0$ is impossible, so there is nothing to prove here. Next fix $k: \mathbf{N}$ and suppose that for all $i, j: \mathbf{N}$, if $i < j$ and $j < k$, then $i < k$; we must prove that, for all $i, j: \mathbf{N}$, if $i < j$ and $j < S \circ k$, then $i < S \circ k$.

We prove this last result by induction on j . First fix $i: \mathbf{N}$ and suppose that $i < 0$ and $0 < S \circ k$. We know that $i < 0$ is impossible, so there is nothing to prove here. Next fix $j: \mathbf{N}$ and suppose that for all $i: \mathbf{N}$, if $i < j$ and $j < S \circ k$, then $i < S \circ k$; we must prove that, for all $i: \mathbf{N}$, if $i < S \circ j$ and $S \circ j < k$, then $i < S \circ k$.

We prove this last result by induction on i . First suppose that $0 < S \circ j$ and $S \circ j < S \circ k$; we must prove that $0 < S \circ k$, which we already know. Next fix $i: \mathbf{N}$ and suppose that, if $i < S \circ j$ and $S \circ j < S \circ k$, then $i < S \circ k$; we must prove that, if $S \circ i < S \circ j$ and $S \circ j < S \circ k$, then $S \circ i < S \circ k$.

We prove this last result using the lemma that $S \circ i < S \circ j$ if and only if $i < j$. Suppose that $S \circ i < S \circ j$ and $S \circ j < S \circ k$. Then $i < j$ and $j < k$, so $i < k$ by the first induction hypothesis. Therefore, $S \circ i < S \circ k$, and we are done. ■

Proposition 73: *$<_{\mathbf{N}}$ is asymmetric.*

Proof: Suppose that $i < j$ and $j < i$ for some $i, j: \mathbf{N}$. By transitivity, it follows that $i < i$, contradicting the previous lemma. ■

Proposition 74: $<_{\mathbf{N}}$ is connected.

Proof: We prove the result by induction on the second variable. Even in the base case, we will also use induction on the first variable. This induction's base case is immediate; $0 = 0$. Next fix $i: \mathbf{N}$ and suppose that $i < 0$ or $i = 0$ or $0 < i$; we must prove that $S \circ i < 0$ or $S \circ i = 0$ or $0 < S \circ i$. In fact, we already know that $0 < S \circ i$.

Next fix $j: \mathbf{N}$ and suppose that for all $i: \mathbf{N}$ we have $i < j$ or $i = j$ or $j < i$; we must prove that for all $i: \mathbf{N}$ we have $i < S \circ j$ or $i = S \circ j$ or $S \circ j < i$. Again we use induction on i . For the base case, we know that $0 < S \circ j$. Next fix $i: \mathbf{N}$ and suppose that $i < S \circ j$ or $i = S \circ j$ or $S \circ j < i$; we must prove that $S \circ i < S \circ j$ or $S \circ i = S \circ j$ or $S \circ j < S \circ i$. Again using the lemma that S respects $<$ (and is injective), this is equivalent to that $i < j$ or $i = j$ or $j < i$, which we know by the outer inductive hypothesis. ■

Lemma 75: Let $i, j: \mathbf{N}$. If $i < S \circ j$, then either $i < j$ or $i = j$. (The converse also holds, but we will not need it.)

Proof: The structure is the same as the last proof; we show only the highlights.

1. (both zero): We have $0 = 0$.
2. (zero and a successor): We have $0 < S \circ j$.
3. (a successor and zero): If $S \circ i < S \circ 0$, then $i < 0$, which is impossible.
4. (both successors): If $S \circ i < S \circ S \circ j$, then $i < S \circ j$, so $i < j$ or $i = j$ by the induction hypothesis, so $S \circ i < S \circ j$ or $S \circ i = S \circ j$. ■

In order to finish off the proof that $<$ is a well-ordering on \mathbf{N} , we need a new version of induction. You may have heard it referred to as “Strong Induction”.

Theorem 76: Induction Principle on \mathbf{N} Schema.

1. Suppose that A is a set in \mathbf{N} and for all $x: \mathbf{N}$, if $k \in A$ for all $k < x$, then $x \in A$. We then have $\mathbf{N} = A$.
2. For any formula $\phi(x, \vec{\chi})$ with an element x of \mathbf{N} and optionally other free variables,

$$\forall \vec{\chi}, (\forall x: \mathbf{N}, (\forall k: \mathbf{N}, k < x \Rightarrow \phi(k, \vec{\chi})) \Rightarrow \phi(x, \vec{\chi})) \Rightarrow \forall x: \mathbf{N}, \phi(x, \vec{\chi})$$

Proof: 1. Let B be $\{x: \mathbf{N} \mid \forall k: \mathbf{N}, k < x \Rightarrow k \in A\}$; our hypothesis states that $B \subseteq A$. Notice that $0 \in B$ because there is no $k: \mathbf{N}$ with $k < 0$. Suppose that $x \in B$; we show that $S \circ x \in B$. Suppose that $k < S \circ x$; we show that $k \in A$. By the previous lemma, either $k < x$ or $k = x$. If $k < x$, then $k \in A$ because $x \in B$ and $k < x$; if $k = x$, then $k \in A$ because $x \in B$ and $B \subseteq A$. Therefore, $S \circ x \in B$. By Step Induction, it follows that $\mathbf{N} = B$. Since $B \subseteq A$, it follows that $\mathbf{N} = A$.

2. This follows from part 1 using Separation. Fix the $\vec{\chi}$, and suppose that

$$\forall x: \mathbf{N}, (\forall k: \mathbf{N}, k < x \Rightarrow \phi(k, \vec{\chi})) \Rightarrow \phi(x, \vec{\chi})$$

Let A be $\{x: \mathbf{N} \mid \phi(x, \vec{\chi})\}$. Given $x: \mathbf{N}$, suppose that $k \in A$ for all $k < x$. We then have $\forall k: \mathbf{N}, k < x \Rightarrow \phi(k, \vec{\chi})$, hence $\phi(x, \vec{\chi})$ by assumption, so $x \in A$. It follows from part 1 that $\mathbf{N} = X$. Therefore, we have $\forall x: \mathbf{N}, \phi(x, \vec{\chi})$. ■

It is possible to give a proof of part 2 which makes use of part 2 of the Step Induction Principle, thus avoiding the detour through objects and using only formulas. This proof simply mimics how we obtained part 1 above, but uses formulas everywhere instead of working with objects. Although it is not nearly as clean, when we treat ordinals, there will times when we need to argue at the level of formulas.

Theorem 77: $<_{\mathbf{N}}$ is a well-ordering.

Proof: We have already proved that $<_{\mathbf{N}}$ is a linear ordering. Suppose then that Z is a set in \mathbf{N} and there is no $m \in Z$ such that for all $x \in Z$, either $m = x$ or $m < x$. We show that $Z = \emptyset$. Notice that for every $m \in Z$, there exists $x \in Z$ with $x < m$ by connectedness.

Let Y be $-Z$. We show that $Y = \mathbf{N}$ using the Induction Principle. Suppose then that $m: \mathbf{N}$ is such that $x \in Y$, i.e. $x \notin Z$, for all $x < m$. If $m \notin Y$, we would then have that $m \in Z$, so by the last sentence of the previous paragraph, there exists $x \in Z$ with $x < m$, a contradiction. Therefore, $m \in Y$. Hence, by the Induction Principle, we have that $Y = \mathbf{N}$ and so $Z = \emptyset$.

Therefore, given $Z: \mathcal{P}\mathbf{N}$ with $Z \neq \emptyset$, there exists $m \in Z$ such that for all $x \in Z$, either $m = x$ or $m < x$. It follows that $<_{\mathbf{N}}$ is a well-ordering. ■

6 Recursion

Already everything works the same way.