Set Theory – An Overview

Gary Hardegree
Department of Philosophy
University of Massachusetts
Amherst, MA 01003

1. Membership

Fundamental to set theory is the notion of *membership* – sets have members, also called *elements*. To express the relation of membership, we use a stylized epsilon symbol '$\in$' (for 'element'). In particular, we write

$$a \in S$$

to say that a is a member of S, and we write

$$a \notin S$$

to say that a is *not* a member of S.

Notice that we adopt a fairly common informal notational convention – to use lower case Roman-Italic letters to denote "points", and to use upper case Roman-Italic letters to denote sets whose elements are points. For sets that have sets as members, we use script letters or some other gaudy font. The following formula illustrates this convention.

$$a \in B \;\;\&\;\; B \in \mathbb{C}$$

Note, however, that the following is equally legitimate.

$$a \in b \;\;\&\;\; b \in c$$

2. Extensionality

Sets have members, just like clubs. But a club is not identified *with* its membership, nor even *by* its membership – two different clubs can have the very same membership. By contrast, it is fundamental to the notion of a set that two different sets *cannot* have the same membership. This is known as the Principle of Extensionality, which may be succinctly formulated as follows.

for any set A, for any set B: $\forall x(x\in A \leftrightarrow x\in B) \rightarrow A=B$	[extensionality]

3. The Empty Set

We said earlier that sets have members, but there is an exception. Specifically, set theory postulates the existence of a set with no members.

there is a set S such that: $\sim\exists x[x \in S]$	[empty set]

In virtue of the Principle of Extensionality, there can be at most one set with no members. Therefore, given that there is at least one such set, there is exactly one such set. It is fittingly called *the* empty set, and is denoted $\varnothing$.

4. Simple Sets; Singletons, Doubletons, etc.

The customary way to denote a set with just a few elements is to list the elements, then surround the list with curly braces, as in the following examples.

{Mozart}
{Mozart, Jupiter}
{Mozart, Jupiter, 41}

Expressions of this basic form can be informally defined as follows.

$\{a\}$	$=_{df}$	the set whose only element is a
$\{a,b\}$	$=_{df}$	the set whose only elements are a and b
$\{a,b,c\}$	$=_{df}$	the set whose only elements are a, b, and c
etc.		

These definitions can be formally summarized in the following principles.

$$\forall x\,(\,x \in \{a\} \leftrightarrow x=a\,)$$
$$\forall x\,(\,x \in \{a,b\} \leftrightarrow.\ x=a \vee x=b\,)$$
$$\forall x\,(\,x \in \{a,b,c\} \leftrightarrow.\ x=a \vee x=b \vee x=c\,)$$
etc.

Terminology:

$\{a\}$ is called the *singleton* (*unit set*) of a.

$\{a,b\}$ is called the *doubleton* (*unordered-pair*) of a and b.

$\{a,b,c\}$ is called the *tripleton* (*unordered triple*) of a, b, and c.

etc.

Next, we note that set theory officially postulates that $\{a\}$, $\{a,b\}$, $\{a,b,c\}$, etc., are all well-defined sets, provided a, b, and c are well-defined.[1] This means that set theory postulates the existence of infinitely-many sets, including the following, just for starters.

$\varnothing$, $\{\varnothing\}$, $\{\{\varnothing\}\}$, $\{\{\{\varnothing\}\}\}$, etc.

Since it is important to appreciate how many sets are alluded to by the above list, we informally demonstrate that (1) the above list has infinitely-many entries, and (2) the above list contains *no* duplicates![2]

Informal Proof:

(1) We begin by noting that, every entry contains two more tokens than its immediate predecessor – except for the first entry, which has no immediate predecessor. Supposing (as we should!) that string σ1 cannot be identical to string σ2 if they contain a different number of tokens, we conclude that that no entry is identical to its immediate predecessor. Furthermore, by mathematical induction,[3] we can establish that no entry is identical to *any* entry preceding it. It follows that there are infinitely-many entries in the above list.

(2) First, note the following.

$\varnothing \neq \{\varnothing\}$

This can be seen by observing that, whereas $\varnothing$ has no elements, $\{\varnothing\}$ has exactly one element. We can similarly argue all the following.

$\varnothing \neq \{\varnothing\}$; $\varnothing \neq \{\{\varnothing\}\}$; $\varnothing \neq \{\{\{\varnothing\}\}\}$; etc.

Based on this information, we can next argue that

$\{\varnothing\} \neq \{\{\varnothing\}\}$

---

[1] Officially, these infinitely-many postulates are accomplished in an elegant mathematical manner, which does not particularly concern us here. For the interested, the relevant official axioms are the Axiom of Pairs, and the Axiom of Unions. See Section 37.

[2] For example, a list can have 3 entries but only refer to 2 individuals – **if** there is a pair of duplicate entries.

[3] Unlike scientific induction, mathematical induction is really deduction, and provides the rigorous logical underpinnings of all reasoning that appeals to the following inference step – "…and so forth (*ad infinitum*)".

Suppose otherwise – that $\{\varnothing\} = \{\{\varnothing\}\}$. Then, since both sets are singletons, their respective elements – – i.e., $\varnothing$ and $\{\varnothing\}$ – must be identical. But we have already argued that this is not so. We can similarly argue all the following.

$$\{\varnothing\} \neq \{\{\varnothing\}\};\ \{\varnothing\} \neq \{\{\{\varnothing\}\}\};\ \{\varnothing\} \neq \{\{\{\varnothing\}\}\};\ \text{etc.}$$

We now have established the pattern for a general proof by mathematical induction, the upshot of which is that the above list contains no duplicates.

:end of proof

5. Pure and Impure Sets

Sets can be constructed entirely from the empty set, as in the above examples. Or they can be constructed from an underlying universe of "ur-elements",[4] which are presumed not to be sets, as in the following (from earlier).

{Mozart, Jupiter, 41}

This corresponds to a distinction between *pure* sets, which ultimately-contain only sets, and *impure* sets, which ultimately-contain non-sets.[5] Whereas formal set theory typically concentrates on pure sets, informal set theory considers mostly impure sets. Formal semantics typically operates in the realm of informal/impure set theory. In particular, linguistic symbols are usually regarded as ur-elements, as are many of the objects designated by these symbols (e.g., persons, places, events).

6. Set-Abstraction

The simple-set notation of the previous sections is unworkable for most sets, whose elements cannot be so easily listed. For this reason, a more concise notation is employed – set-abstraction, whose basic form is

$$\{\nu : \mathbb{F}\}$$

where ν is a variable, and $\mathbb{F}$ is a formula. The following are simple examples.

{x : x is happy}
{x : x is happy and x is virtuous}
{x : the mother of x is taller than x}

The intuitive idea is quite simple – $\{\nu{:}\mathbb{F}\}$ consists of exactly those things that satisfy the condition described by the formula $\mathbb{F}$.[6] For example, {x : x is happy} consists of exactly those things that satisfy the condition of being-happy.

The following is the official explicit definition.

$\{\nu{:}\mathbb{F}\}$ $=_{\text{df}}$ $\iota S\forall\nu(\nu \in S \leftrightarrow \mathbb{F})$ [S not free in $\mathbb{F}$]		[set-abstract]

Here, the symbol ‘ɩ’ (upside-down iota) is the definite-description-operator, informally defined as follows.

$$\iota\nu\mathbb{F} \ =_{\text{df}}\ \text{the } \nu \text{ such that } \mathbb{F}$$

---

[4] The morpheme ‘ur’, which comes from German, means ‘original’ (**ur**iginal?) or ‘first’.

[5] The term ‘ultimately-contain’ can be technically defined if we wish [it is the transitive closure of the converse of the element-relation]. Technical definitions aside, the basic idea is quite simple. For example, the set $\{\{\varnothing\}\}$ has one element – $\{\varnothing\}$ – which in turn has one element – $\varnothing$. So, $\varnothing$ is ultimately-contained in $\{\{\varnothing\}\}$. On the other hand, $\varnothing$ is not an element of $\{\{\varnothing\}\}$, since $\{\{\varnothing\}\}$ contains just $\{\varnothing\}$, and $\varnothing\neq\{\varnothing\}$, as seen earlier. A set that contains ($\ni$) every item it ultimately-contains is called a *transitive set*; the set $\{\varnothing, \{\varnothing\}\}$ is an example.

[6] It’s a little more complicated, if $\mathbb{F}$ contains free variables other than ν, but we need not go into that here.

7. Set-Abstract Conversion

The official definition of set-abstraction employs the description-operator. However, the description-operator is almost never employed in set theory. Rather, it usually gets hidden under an associated principle of set-abstract conversion.

$\forall \nu(\nu \in \{\nu : \mathbb{F}\} \leftrightarrow \mathbb{F})$	[set-abstract conversion]

Here, ν is any variable, and $\mathbb{F}$ is any formula. The following is a simple example.

$$\forall x(x \in \{x : x \text{ is happy}\} \leftrightarrow x \text{ is happy})$$

This says that something is an element of $\{x : x \text{ is happy}\}$ if and only if it is happy.

8. Propriety and Impropriety Among Set-Abstracts

Even if we never explicitly employ the description-operator, we should keep it in mind. The reason is that it reminds us of a potentially very serious logical problem. Specifically, definite-description expressions ("descriptions" for short) can be *proper* or *improper*, according to the following definition.

$\iota\nu\mathbb{F}$ is proper	$=_{df}$	the formula $\mathbb{F}$ is satisfied by exactly one individual in the relevant salient domain of discourse

Additionally, it is fairly common to adopt the following principle about the denotation of descriptions.

> If the formula $\mathbb{F}$ is satisfied by exactly one individual in the relevant salient domain of discourse, then the description $\iota\nu\mathbb{F}$ denotes that individual;
>
> otherwise, $\iota\nu\mathbb{F}$ denotes nothing.

For example, the description ‘the dog’ ($\equiv$ ‘$\iota x[x$ is a dog]’) is proper if and only if there is exactly one dog in the relevant salient domain of discourse. For example, if a family has exactly one dog, then in the usual familial contexts, ‘the dog’ is a proper description which refers to that dog. On the other hand, if the family has no dog, or has more than one dog, then ‘the dog’ is improper in many of these contexts.[7]

Now, the same applies to set-abstracts – some are proper, and some are not. It was originally naïvely assumed that all set-abstracts are proper, which amounts to the following principle of what is often called "naïve set theory".

> every property has an extension;
> i.e., for any property $\mathbb{P}$, there is a set of things that have $\mathbb{P}$;
> alternatively, for any formula $\mathbb{F}$, there is a set of things satisfying $\mathbb{F}$.

Although this principle seems intuitively obvious, it is logically disastrous, as first shown by Russell.[8] In light of Russell’s discovery, known as *Russell’s Paradox*, modern set theory is

---

[7] We say ‘many’ here, since one can easily envisage circumstances in which the relevant salient familial domain is different from the "usual" one.

[8] Russell showed that the abstract ‘$\{x : x \notin x\}$’ is not proper, since assuming it is proper immediately yields a contradiction when we apply set-abstract conversion (see Section 7) – namely, $\{x : x \notin x\} \in \{x : x \notin x\} \leftrightarrow \{x : x \notin x\} \notin \{x : x \notin x\}$.

obsessively careful about postulating the propriety of set-abstracts, and the existence of the corresponding sets.[9]

9. Simple Sets Revisited

Notice that simple sets – $\{a\}$, $\{a,b\}$, etc – can be defined using set-abstraction, as follows.

$\{a\}$	$=_{df}$	$\{x : x{=}a\}$	[singleton]
$\{a,b\}$	$=_{df}$	$\{x : x{=}a \vee x{=}b\}$	[doubleton]
$\{a,b,c\}$	$=_{df}$	$\{x : x{=}a \vee x{=}b \vee x{=}c\}$	[tripleton]
etc.			

Notice also that the empty set $\varnothing$ can be defined using set-abstraction as follows.

$$\varnothing \;=_{df}\; \{x : x \neq x\} \qquad \text{[empty set]}$$

10. Specific Set-Abstraction

Set-abstracts of the form $\{\nu{:}\mathbb{F}\}$ might be called "generic set-abstracts", which are contrasted with "specific set-abstracts", which have the form

$$\{\nu \in \tau : \mathbb{F}\}$$

where ν is a variable, τ is a singular term (in which ν does not occur free), and $\mathbb{F}$ is a formula. The following is the official definition in terms of generic set-abstraction.

$$\{\nu \in \tau : \mathbb{F}\} \;=_{df}\; \{\nu : \nu \in \tau \;\&\; \mathbb{F}\} \qquad \text{[specific set-abstract]}$$

Specific-abstraction is useful in two ways. First, it can be used to make explicit the underlying domain. For example, a generic abstract such as

$$\{x : x \text{ is even}\}$$

in some sense presupposes a tacitly understood domain of discourse (the natural numbers, the integers, the rational numbers, etc.). On the other hand, a specific abstract such as

$$\{x \in \mathbb{N} : x \text{ is even}\}$$

explicitly declares the domain to be the set $\mathbb{N}$ (= the set of natural numbers; see Section 26).

Second, specific abstraction fits very nicely with a fundamental principle of modern set theory – the Axiom of Separation (see Section 37.). In particular, this axiom can be understood as saying that:

for any set S, for any condition $\mathbb{F}$, the set $\{x \in S : \mathbb{F}\}$ is well-defined

In other words, specific-abstracts are always proper, unlike generic abstracts, which are not always proper.

[9] Notwithstanding the obsessions of set-theorists, we in the lay public can *usually* go about our business in a nonchalant manner. *Most* of the sets we need are well-defined.

11. Generalized Set-Abstraction

There is another frequently used set-abstraction technique, whose general form is

$$\{\tau : \mathbb{F}\}$$

where τ is a singular term, and $\mathbb{F}$ is a formula. The following are simple examples.

{ the mother of x : x is a composer }
{ $x \times y$: x is odd & y is odd }

The first is the set of mothers of composers. The second is the set of numbers obtainable by multiplying two odd numbers. More generally, we have the following official definition.

$\{\tau : \mathbb{F}\} =_{df} \{\nu : \exists\nu_1 \ldots \exists\nu_k(\mathbb{F} \,\&\, \nu = \tau)\}$		[generalized set-abstract]
here, $\nu_1, \ldots, \nu_k$ are the free variables common to τ and $\mathbb{F}$, and ν is any variable distinct from these.		

So, for example:

{the mother of x : x is a composer} = $\{y : \exists x(x$ is a composer & $y =$ the mother of $x)\}$
$\{x \times y : x$ is odd & y is odd$\}$ = $\{z : \exists x \exists y(x$ is odd & y is odd & $z = x \times y)\}$

12. Simple Set-Theoretic Operations

Various set-theoretic operations can also be defined using set-abstracts. The following are a few examples.[10]

$A \cap B$	$=_{df}$	$\{x : x \in A \,\&\, x \in B\}$	[simple intersection]
$A \cup B$	$=_{df}$	$\{x : x \in A \vee x \in B\}$	[simple union]
$A - B$	$=_{df}$	$\{x : x \in A \,\&\, x \notin B\}$	[set-difference]
$A + B$	$=_{df}$	$\{x : x \in A$ xor[11] $x \in B\}$	[Boolean sum]
$\bigcap \mathbb{C}$	$=_{df}$	$\{x : \forall Y(Y \in \mathbb{C} \rightarrow x \in Y)\}$	[general intersection]
$\bigcup \mathbb{C}$	$=_{df}$	$\{x : \exists Y(Y \in \mathbb{C} \,\&\, x \in Y)\}$	[general union[12]]

Whereas simple intersection/union applies to pairs of sets, and therefore to finite collections by inductive generalization [e.g., $A \cup (B \cup C)$], general intersection/union applies to arbitrary collections of sets, including infinitely-large collections. Notice that simple intersection/union is a special case of general intersection/union, in light of the following theorems.

$$\bigcap\{A,B\} = A \cap B$$
$$\bigcup\{A,B\} = A \cup B$$

---

[10] It is commonplace to give formal definitions using informal notation. For example, in these definitions, the variables 'A', 'B', and '$\mathbb{C}$' serve as meta-linguistic variables ranging over singular terms of the underlying language. Also, schematic definitions frequently contain tricky provisos ("catches"); here, the catch is that x and Y do not occur free in A, B, or $\mathbb{C}$.

[11] Here, 'xor' is exclusive-disjunction – P xor Q $=_{df}$ (P & ~Q) $\vee$ (Q & ~P).

[12] The existence of unions is postulated by the Axiom of Unions. See Section 37.

13. Inclusion and Exclusion

Set A is said to be *included* in set B precisely when every element of A is also an element of B. For example, the set of college sophomores is included in the set of college students, which is included in the set of students. There are a number of alternative ways of expressing this; the following are synonymous.

A is *included* in B	B *includes* A
A is a *subset* of B	B is a *superset* of A

The following introduces the official notation, which presupposes that A and B are sets.[13]

$$A \subseteq B \quad =_{\text{df}} \quad \forall x(x{\in}A \rightarrow x{\in}B) \qquad \text{[inclusion]}$$

Notice that the empty set $\varnothing$ is (trivially) included in every set. Notice also that every set is automatically included in itself. Finally, notice that the Principle of Extensionality can be rewritten as follows.

$$A \subseteq B \;\;\&\;\; B \subseteq A \;\; .{\rightarrow}\;\; A = B$$

This suggests a further notion – A is *properly included* in B if and only if A is included in B, but B is not included in A. The following is the official definition.[14]

$$A \subsetneq B \quad =_{\text{df}} \quad A \subseteq B \;\&\; {\sim}[B \subseteq A] \qquad \text{[proper inclusion]}$$

Alternative terminology exists; the following are synonymous.

A is *properly included* in B	B *properly includes* A
A is a *proper subset* of B	B is a *proper superset* of A

Note that, in light of the Principle of Extensionality, we have the following theorem.

$$A \subset B \;\; {\leftrightarrow}. \;\; A \subseteq B \;\&\; A \neq B$$

In other words, a proper subset of a set S is any subset of S other than S itself.

Next, sets A and B are said to be *disjoint* precisely when they have no elements in common. There is alternative terminology; the following are synonymous.

$\{A,B\}$ is/are a *disjoint pair*[15]
A is *disjoint from* B
A *excludes* B

The following is the official definition.

$$A \perp B \;\; =_{\text{df}} \;\; {\sim}\exists x(x{\in}A \;\&\; x{\in}B\} \qquad \text{[disjoint]}$$

On the other hand, a collection $\mathbb{C}$ of sets is said to be *pair-wise disjoint* precisely if every pair included in $\mathbb{C}$ is/are disjoint. Formally:

$$\perp(\mathbb{C}) \quad =_{\text{df}} \quad \forall X{\in}\mathbb{C}, \forall Y{\in}\mathbb{C} : X \perp Y \qquad \text{[pair-wise disjoint]}$$

---

[13] The catch here is that the variable x cannot occur free in A or B.

[14] Our official definition is in accordance with the general mathematical practice of defining a strict-ordering based on an ordering as follows: $x{<}y =_{\text{df}} x{\leqslant}y \;\&\; y{\nleqslant}x$. This way of defining it does not presuppose that $\leqslant$ is anti-symmetric ($x{\leqslant}y \;\&\; y{\leqslant}x \;.{\rightarrow}\; x{=}y$).

[15] The singular/plural (is/are) distinction is based on whether we treat a set as a singular object or as a plurality. Technically, we must treat a set as a singular object, but informally we also often treat a set as the corresponding plurality. Context usually indicates which usage is in effect.

The *definiens* in the above definition employs our first use of *set-specific-quantification*. It is read so naturally that we hardly even notice it.

for any X *in* $\mathbb{C}$, for any Y *in* $\mathbb{C}$: X is disjoint from Y

The official definition of set-specific-quantification is fairly obvious.[16]

$\forall\nu{\in}\tau{:}\mathbb{F} \quad =_{\text{df}} \quad \forall\nu(\nu \in \tau \rightarrow \mathbb{F})$
$\exists\nu{\in}\tau{:}\mathbb{F} \quad =_{\text{df}} \quad \exists\nu(\nu \in \tau \ \& \ \mathbb{F})$
Here, $\mathbb{F}$ is any formula, τ is any singular term, and ν is any variable not free in τ.

14. Power Sets

Given any set A, we can consider all the subsets of A, which we can then collect into a set of its own, called the *power set* of A, and formally defined as follows.[17]

$\wp(A) \quad =_{\text{df}} \quad \{X : X \subseteq A\}$	[power set]

In other words, X is an element of $\wp(A)$ if and only if X is a subset of A.

Examples:

A	$\wp(A)$
$\varnothing$	$\{\varnothing\}$
$\{0\}$	$\{ \varnothing, \{0\} \}$
$\{0,1\}$	$\{ \varnothing, \{0\}, \{1\}, \{0,1\} \}$
$\{0,1,2\}$	$\{ \varnothing, \{0\}, \{1\}, \{2\}, \{0,1\}, \{0,2\}, \{1,2\}, \{0,1,2\} \}$

Here, numerals are used intuitively to denote numbers. See Section 26 for a formal presentation.

Note that the power set $\wp(A)$ invariably has greater power (i.e., size) than A (see Section 35). For example, if A has k-many elements, then $\wp(A)$ has 2^k-many elements.

15. Ordered-Pairs

The elements of a set are not ordered. For example, the set

{Jupiter, Saturn}

and the set

{Saturn, Jupiter}

are one and the same set, in virtue of the Principle of Extensionality. If we want to sort the elements of the pair {Jupiter, Saturn} so that one of them is first and the other is second, we need the notion of *ordered-pair*.

There are basically two approaches to ordered-pairs.

---

[16] Set-specific-quantification a special case of a more general form of specific-quantification, given by the following definition.

$\forall\mathbb{F}_1\mathbb{F}_2 \ =_{\text{df}} \ \forall\nu_1\ldots\forall\nu_k(\ \mathbb{F}_1 \rightarrow \mathbb{F}_2\ )$

$\exists\mathbb{F}_1\mathbb{F}_2 \ =_{\text{df}} \ \exists\nu_1\ldots\exists\nu_k(\ \mathbb{F}_1 \ \& \ \mathbb{F}_2\ )$

here, $\nu_1, \ldots, \nu_k$ are the free variables common to $\mathbb{F}_1$ and $\mathbb{F}_2$.

[17] The relevant postulate is called "the axiom of powers" or "the power set axiom". See Section 37.

(1) We can treat ordered-pairs as an additional primitive concept of set theory.
(2) We can define ordered-pairs in terms of ordinary (un-ordered) sets.

In either case, we need to make precise the criterion of individuation, which is summarized as follows.

Criterion of Individuation for Ordered-pairs:
An ordered-pair is individuated entirely by its components; to know the identity of an ordered-pair is simply to know which component is first and which component is second.

This can be stated more formally as follows.

If p_1 and p_2 are ordered-pairs, then:

$$p_1 = p_2 \;\leftrightarrow.\; 1^{st}(p_1) = 1^{st}(p_2) \;\&\; 2^{nd}(p_1) = 2^{nd}(p_2)$$

where

$1^{st}(p) =_{df}$ the first component of P
$2^{nd}(p) =_{df}$ the second component of P

The notation for ordered-pairs varies from author to author. Officially, we use round parentheses in accordance with the following informal definition.[18]

$(a,b) =_{df}$ the ordered-pair p such that $1^{st}(p) = a$ and $2^{nd}(p) = b$

Earlier, we claimed that ordered-pairs might be treated as derivative objects, defined in terms of un-ordered sets. The usual definition (originally due to von Neumann) goes as follows.

$$(a,b) =_{df} \{ \{a\}, \{a,b\} \}$$

With this definition, we can prove the fundamental (individuation) principle about ordered-pairs – namely:

$$(a_1, a_2) = (b_1, b_2) \;\leftrightarrow.\; a_1 = b_1 \;\&\; a_2 = b_2$$

This is the good news. The bad news is that the above definition also yields some spurious theorems, including:[19]

$$\{a\} \in (a,b);\ \{a,b\} \in (a,b);\ a \notin (a,b);\ b \notin (a,b)$$

At this point, we have to bite the bullet – the question is which one. Do we admit an extra primitive concept into set theory; or, do we live with a few spurious theorems. I

[18] Notice that our notation for ordered-pairs is nearly absolutely minimal. To see this, let us consider a few alternatives. First, we introduce an ordered-pair symbol, preferably not already in wide use in logic – say '%'. Next, we decide whether we write "%ab" or "ab%" or "a%b"? The advantage of prefix and postfix formatting is that no parentheses are required; the disadvantage is that few humans can read complex combinations of them. The advantage of infix formatting is that humans can read them; the disadvantage is that parentheses are required. So, officially "a%b" is really "(a%b)". But notice that the latter differs from "(a,b)" by merely substituting "%" for ",". But, surely, a comma is more minimal than '%'. Nevertheless, since most uses of ordered pairs do not involve reiteration, we don't really need parentheses, so it may occasionally be useful to use a primitive symbol (for example, some sort of slash) to notate ordered pairs.

[19] This sort of thing is common in set theory – in an attempt to reduce mathematical concepts to just a few, one usually ends up positing a few spurious facts about various mathematical objects. If we think of set theory as proposing *models* of mathematical entities, then it is important to keep in mind that, generally, a model has both positive analogy and negative analogy. For example, a model airplane has wings, like a real airplane (positive analogy), but a model airplane is (usually) made of cheap plastic, unlike a real airplane (negative analogy).

propose that we "bracket" this question, since nothing we propose later depends upon our choice.[20]

The ontological issues do not stop here, though. In addition to ordered-pairs, we need ordered-triples, ordered-quadruples, etc. However, our approach is to derive all of these concepts (and many others!) from the concept of ordered-pair. So we really only have one agonizing choice at the moment – whether or not to treat ordered-pairs as primitive.

16. The Cartesian-Product

Once we have the notion of ordered-pair, we can define numerous other set-theoretic notions. The first one is the *Cartesian product*.

René Descartes (a.k.a. Cartesius) observed that the Euclidean plane can be coordinatized by the real numbers. This means that every spatial point can be labeled, i.e., assigned coordinates, which are simply ordered-pairs of real numbers. Beginning with this observation, Descartes showed how the Euclidean plane can be replaced by (reduced to) the set of ordered-pairs of real numbers. Thus came into being the field of analytic geometry, which was essential to the development of the differential and integral calculi, which were essential to the development of physics.

Indeed, the reduction of geometry to analysis is so natural that we are led to identify the Euclidean points *with* (not simply *by*) their Cartesian coordinates.

The set of ordered-pairs of real numbers is a prominent and historically important example of a Cartesian product. In particular, it is the Cartesian product of the set $\mathbb{R}$ of real numbers with itself. More generally, for any sets A and B, we can form their Cartesian product in accordance with the following definition.

$$A \times B \;=_{\mathrm{df}}\; \{\, (x,y) : x \in A \;\&\; y \in B \,\} \qquad \text{[Cartesian-product]}$$

Notice that this is our first official use of generalized set-abstraction (see Section 11).

[20] Note, however, that the formal theory of sets we officially describe (see Section 37) treats ordered-pairs as derivative.

17. Relations

Set-theoretic relations[21] are intended to be the extensions of dyadic (2-place) predicates, in pretty much the same way that ordinary sets are extensions of monadic (1-place) predicates. For example, the extension of the predicate 'is taller than' is the set of ordered-pairs each of which is such that its first component is taller than its second component. In other words:

the extension of 'is taller than' $=$ $\{ p : 1^{\text{st}}(p)$ is taller than $2^{\text{nd}}(p) \}$

or:

the extension of 'is taller than' $=$ $\{ (x,y) : x$ is taller than $y \}$

The notion of binary set-theoretic relation (or simply relation[22]) is officially defined as follows.

R is a relation $=_{\text{df}}$ R is a set of ordered-pairs[23]		[relation]

Next, we say that a *bears* relation R to b precisely when the ordered-pair (a,b) is an element of R; in other words:

a bears R to b $=_{\text{df}}$ $(a,b) \in R$

It is customary to abbreviate the 3-place predicate '…bears…to…' in the starkest manner possible.

aRb $=_{\text{df}}$ a bears R to b $=_{\text{df}}$ $(a,b) \in R$

alternatively:

Rab $=_{\text{df}}$ a bears R to b $=_{\text{df}}$ $(a,b) \in R$[24]

Associated with every relation are three inter-related sets, called its *domain*, *range*, and *field*, which are officially defined as follows.

$$\begin{aligned} \textit{domain}(R) &=_{\text{df}} \{1^{\text{st}}(p) : p \in R\} = \{ x : \exists y[xRy] \} \\ \textit{range}(R) &=_{\text{df}} \{2^{\text{nd}}(p) : p \in R\} = \{ y : \exists x[xRy] \} \\ \textit{field}(R) &=_{\text{df}} \textit{domain}(R) \cup \textit{range}(R) \end{aligned}$$

With these notions in hand, we can define 'from', 'to', and 'on'.

R is a relation *from* A if and only if:

$\textit{domain}(R) \subseteq A$

R is a relation *to* B if and only if:

$\textit{range}(R) \subseteq B$

R is a relation *on* A if and only if:

$\textit{field}(R) \subseteq A$

---

[21] More properly *binary* (2-place) relations. Without a salient modifier, 'relation' means 'binary relation'. We discuss the more general category of k-place relations later.

[22] See note 21.

[23] Note: in general, a set of Φ's is a set every element of which is a Φ.

[24] Since this notation looks exactly like predicate-subject-object notation in predicate logic, there is likely to be some confusion. In the context of predicate logic, 'Rab' has two subjects 'a' and 'b', and one predicate 'R'. In the context of set theory, 'R' is not a 2-place predicate, but one of the three "subjects" of the covert 3-place predicate '…bears…to…'.

A relation *from* A *to* B is then a relation that is both *from* A and *to* B. Notice that this is equivalent to:

$$R \subseteq A \times B$$

Notice also that a relation *on* A is simple a relation from A to A (i.e., $R \subseteq A \times A$).

18. Functions

From the point of view of set theory, a function is a special kind of relation,[25] the following being the official definition.

A relation R is said to be a *function* precisely if it satisfies the following condition.

$$\forall x \forall y \forall z (xRy \;\&\; xRz \;.\!\rightarrow y{=}z)$$

In other words, a function is a relation R in which no single thing bears R to two or more things.

Examples:

The set of ordered pairs (x,y) satisfying:	is a function?
$x^2 + y^2 = 4$	no
$y = x^2$	yes
y parents x	no
y fathers x	yes

It is customary (but hardly universal) to use lower case letters to denote functions. It is also customary to employ *function-argument-value* notation. In the paradigm[26]

$$f(a)$$

which is read "f of a",

f is the *function;*
a is the *argument*;
$f(a)$ is the *value* of f at a.

Officially, we have the following explicit definition.

$f(a) \;=_{\text{df}}\; \iota x[afx]$ [function-application]

[25] A function *simpliciter* is a unary (1-place) function. We will discuss more general k-place functions later.

[26] This is the "analytic" paradigm, as in (mathematical) analysis. In the "algebraic" paradigm, not widely used anymore, but once used primarily by algebraists, one writes "af" rather than "$f(a)$". Whether to use prefix or postfix notation corresponds in ordinary language to whether we say (e.g.) "the mother of a" or "a's mother".

In other words $f(a)$ is the unique individual to which a bears relation f.[27] However, the definite-description notation seldom arises in practice, but is rather hidden under the following conversion principle, which serves as an implicit definition.

$\forall x \forall y (\, y = f(x) \leftrightarrow xfy \,)$	[function-conversion]

Next, just as with relations, we can define 'from', 'into', and 'on'.[28]

f is a function *from* A if and only if:

$$domain(f) = A$$ [29]

f is a function *into* B if and only if:

$$range(f) \subseteq B$$

f is a function *on* A if and only if:

f is a function from A into A.

Notation: $f{:}A\to B$ $=_{\text{df}}$ f is a function from A into B.

Notation: $A\to B$ $=_{\text{df}}$ set of all functions from A into B.

Notice that our notational convention yields the following simple biconditional.

$$f{:}A\to B \;\leftrightarrow\; f \in A\to B$$

By way of closing this section we note that every relation has its "very own" function. Specifically, we have the following theorem.

(t) Let A and B be sets, and let R be a relation from A to B. Then there is an associated function fun_R from A into the power set $\wp(B)$ such that:

$$fun_R(a) = \{\, b\in B : aRb \,\}$$

Since this is semantically useful, a simple example might be worthwhile. Consider the friendship-relation restricted to humans. I.e.,

$$F =: \{\, (x,y) : y \text{ is a friend of } x \text{ (both being humans)} \,\}$$

Now, presumably F is not itself a function, since a given person might have two or more friends. On the other hand, there is an associated friendship-function fun_F defined as follows.

$$fun_F(x) = \{y : y \text{ is a friend of } x\}$$

[27] As with the distinction between binary relations and two-place predicates, there is a distinction between unary functions and 1-place function *signs*. In elementary logic, the expression '$f(a)$' consists in a function sign 'f' applied to a singular term 'a'. By contrast, in set theory, the expression '$f(a)$', 'f ' is not a function sign, but a singular term just like 'a'. The resulting complex noun phrase '$f(a)$' then results by applying a covert 2-place function sign – something like 'the result of applying…to…'.
[28] Notice the slight variation in wording – a function from A "into" B is a relation from A "to" B.
[29] Notice that, because of this condition, a function from A is not merely a relation from A that is also a function. It is key that A is the domain of the function/relation; every element of A must be related to something.

At this point, we note that it is occasionally convenient to abuse notation slightly, and use the same letter 'F' for both the original relation and the derived function, so that the following is intelligible.

$$F(a) = \{x : aFx\}$$

Here, the first occurrence of 'F' denotes the function derived from the relation denoted by the second occurrence of 'F'.

The notation is fairly easy to read in simple cases; for example – $F(a)$ **are** the friends of a.

19. Characteristic-Functions

In the previous section, we noted that it is occasionally useful to informally identify (conflate) a relation with its derivative function In this section, we consider another semantically-useful function, and an associated semantically-convenient conflation.

Let A be a subset of set $\mathbb{U}$. Then associated with A is a function $\mathbb{X}_A$ from $\mathbb{U}$ into the set $\{\mathsf{T},\mathsf{F}\}$ of truth-values,[30] defined so as to satisfy the following condition.

$$\begin{aligned} \mathbb{X}_A(a) &= \mathsf{T} \quad \text{if} \quad a \in A \\ &= \mathsf{F} \quad \text{if} \quad a \notin A \end{aligned}$$

The function $\mathbb{X}_A$ is called the *characteristic-function* of A (relative to $\mathbb{U}$).[31]

Now, it is often convenient to informally identify (conflate) a set A with its characteristic-function $\mathbb{X}_A$ (granted a prior-understood universe $\mathbb{U}$). We signal this as follows.

$$S \approx \mathbb{X}_S$$

By making this informal identification, we can use function-notation and membership-notation interchangeably – as the context demands. For example, supposing $\mathbb{U}$ = the set of composers, we can intelligibly write the following.

{Mozart, Bach}(Bach)	=	T	because	Bach $\in$ {Mozart, Bach}
{Mozart, Bach}(Beethoven)	=	F	because	Beethoven $\notin$ {Mozart, Bach}

Now, the set {Mozart, Bach} is not a function *per se*, so we cannot meaningfully apply it to Bach, or Beethoven, or anything for that matter. On the other hand, we can apply its associated characteristic-function $\mathbb{X}_{\{\text{Mozart, Bach}\}}$, which we informally identify with the set {Mozart, Bach}.

More generally, we have the following equivalence .

$$S(a) = \mathsf{T} \;\leftrightarrow\; a \in S$$

This is understood so that the first occurrence of 'S' refers to the characteristic-function of the set, and the second occurrence refers to the set itself.

By way of concluding this section, we illustrate how our two informal identifications (conflations) can be combined. In particular, let 'F' ambiguously denote the friendship-

[30] For formal semantic reasons, we introduce as genuine primitives the truth-values T ("the true") and F ("the false"). Note that, since T and F have no particular set-theoretic use, set theory customarily defines characteristic-functions in terms of the numbers 0 (in place of F) and 1 (in place of T). The numbers in turn are defined as sets; for example, $0 =_{df} \varnothing$; $1 =_{df} \{\varnothing\}$. See Section 26. Another approach reverse-engineers the truth-values, treating "the true" as the number 1, and "the false" as the number 0. The problem with this approach is that it is (1) ontologically ridiculous, and (2) artificially mathematical. I prefer keeping formal semantic primitives, and reverse-engineering characteristic-functions.

[31] The letter '$\mathbb{X}$' is meant to be a stylized upper case letter chi, which is short for 'characteristic'.

relation and the associated friendship-function. Combining this with our ambiguous use of set-names, we can intelligibly write the following.

$$[F(a)](b) = \mathsf{T} \;\leftrightarrow\; aFb$$

Notice that this says, in effect, that b is an element of the set of a's friends if and only if b is a friend of a.

20. Inversion, Composition, and Restriction

In the current section, we examine a few ways of modifying relations, and hence functions. For various reasons, however, the concepts are subtly adjusted when moving from the relational context to the functional context.

1. Relational-Inversion

First, every relation R has a *converse* (or *inverse*), denoted R^{-1}, and defined as follows.

$$R^{-1} =_{\text{df}} \{ (x,y) : yRx \}$$

In other words:

$$x \text{ bears } R^{-1} \text{ to } y \;\leftrightarrow\; y \text{ bears } R \text{ to } x.$$

For example, the ancestor-relation is the converse of the descendent-relation.

$$x \text{ is an ancestor of } y \;\leftrightarrow\; y \text{ is a descendent of } x$$

2. Functional-Inversion

Since every function is a relation, it can be inverted *qua* relation. The problem is that the converse of a function need not be a function. For example,

$$\{ (x,y) : y \text{ fathers } x \} \qquad \{ (x,y) : y \text{ mothers } x \}$$

are both functions, but their respective converses are not. Everyone has exactly one father and exactly one mother, but fathers and mothers may have more than one child.

3. Relational-Composition

Inversion is a natural one-place operation on relations. There is also a natural two-place operation, called *composition*, defined as follows.

$$R \circ S =_{\text{df}} \{ (x,y) : \exists z[xRz \;\&\; zSy]\}$$

or:

$$x[R \circ S]y \;\leftrightarrow\; \exists z[xRz \;\&\; zSy]$$

or:

$$x \text{ bears } R \circ S \text{ to } y \;\leftrightarrow\; x \text{ bears } R \text{ to something that bears } S \text{ to } y.$$

Kinship relations offer countless examples of composition and inversion. In the following we adopt a stilted dialect in which we convert every relational phrase into a simple transitive verb (a process one might call "verbing"). With this *caveat* in mind, consider the following examples of kinship.

x aunts y	$\leftrightarrow$	x sisters someone who parents y
x uncles y	$\leftrightarrow$	x brothers someone who parents y
x grandparents y	$\leftrightarrow$	x parents someone who parents y
x great-grandparents y	$\leftrightarrow$	x parents someone who parents someone who parents y

Notice that the last example subtly illustrates an important algebraic fact about relational-composition – it is *associative*, which is to say that the following condition is satisfied.

$$(R\circ S)\circ T \;=\; R\circ(S\circ T)$$

On the other hand, relational-composition is not *commutative*, which is to say that generally:

$$R\circ S \;\neq\; S\circ R$$

For example, the following are not equivalent.

x sisters someone who parents y
x parents someone who sisters y

While we are on this topic, we note the following interesting algebraic fact.

$$(R\circ S)^{-1} \;=\; S^{-1}\circ R^{-1}$$

For example, the following are equivalent (restricted to women).[32]

x mothers someone who aunts y
y nieces someone who childs x

4. Functional-Composition

Since functions are relations, they can be composed, although the notation is altered. The basic idea is simple; if f and g are functions, then the composite-function, denoted $f\circ g$, satisfies the following condition.

$$[f\circ g](x) \;=\; f\,(\,g(x)\,)$$

Notice that this notation is exactly the reverse of the notation for relational-composition.[33] Also notice that the above condition provides only a *desideratum*, not a definition. The definition is provided as follows.

$$f\circ g \quad =_{\text{df}} \quad \{\,(x,y) : \exists z(\,xgz \;\;\&\;\; zfy\,)\,\}$$

Notice that, since every function is also a relation, this definition is in direct conflict with our earlier definition of relational-composition. This is a problem that we solve pragmatically. In particular, we adopt the following strategy.

When functions are written relationally, we use relational-composition. When functions are written functionally, we use functional-composition.

The following illustrates both clauses of this principle in a single sentence.

$$x\,[f\circ g]\,y \;\;\leftrightarrow\;\; y = [g\circ f](x)$$ [34]

5. Relational-Restriction

By way of concluding this section, we introduce one further concept – *restriction* – which we have already used informally. We now make it official.

$$R|A \quad =_{\text{df}} \quad R\cap(A\times A)$$

“$R|A$” is read “R restricted to A”. When we restrict a relation R to set A, we "filter out" elements not in A; i.e.,

$$x \text{ bears } R|A \text{ to } y \;\;\leftrightarrow\;\; x \text{ bears } R \text{ to } y \;\; \textbf{and} \;\; x, y\in A$$

[32] The notion of restriction is used informally here; it is entirely natural. Nevertheless, we provide a formal account later in this section.

[33] If we adopted algebraist notation, writing “xf” instead of the customary “$f(x)$”, this problem would not arise.

[34] The syntactic transformation implicitly alluded to in this principle seems very natural – read each side of the biconditional both left to right and right to left.

6. Functional-Restriction

Functions are relations, and can accordingly be restricted, although there is a slight adjustment.

$$f|A \;=_{df}\; f \cap (\, A \times range(f)\,)$$

Notice how this differs from the corresponding definition for relational-restriction. In functional-restriction, only the domain (the "from") is restricted, not the co-domain (the "to"); by contrast, in relational-restriction, both the "from" and the "to" are restricted. As with composition, the notational conflict can be handled pragmatically.

21. Ancestors

If we think of relational composition as a form of multiplication, then it is natural to use exponent notation, which yields the following definitions.[35]

$$R^2 \;=_{df}\; R \circ R$$
$$R^3 \;=_{df}\; R \circ R \circ R$$
$$R^4 \;=_{df}\; R \circ R \circ R \circ R$$
etc.

Notice that, since relational composition is associative, we can drop parentheses.

Considering all these relations at once, we can conceptualize another very important set-theoretic notion – that of *ancestor*. Let's begin with the concrete example that motivates the terminology. To say that a is an ancestor of b is to say that one of the following is true, where P is the parents-relation.

aPb	i.e.,	a parents b
aP^2b	i.e.,	a parents someone who parents b
aP^3b	i.e.,	a parents someone who parents someone who parents b
etc.		

Set theory is not concerned with the ancestor-relation *per se*. Rather, it is interested in the above construction for an arbitrary relation R. In particular, starting with any relation R, we can form the corresponding ancestral-relation, which we denote R^{a}, informally defined as follows.

$$xR^{a}y \;\leftrightarrow\; xRy \vee xR^2y \vee xR^3y \vee \ldots$$

or:

$$R^{a} \;=_{df}\; \bigcup\{R, R^2, R^3, \ldots\}$$

22. Special Types of Relations

Oftentimes, we are interested in relations with special properties, to which we give special names. The following is a list of a few of the basic types of relations. In what follows, R is presumed to be a relation, although the definitions do not officially demand this.

R is Φ in A	
Φ	**Definition**
reflexive	$\forall x \in A : xRx$
irreflexive	$\forall x \in A : \sim xRx$
transitive	$\forall xyz \in A : xRy \;\&\; yRz \;.\!\rightarrow xRz$

[35] Unfortunately, this notation conflicts with later notation for set exponentiation. We will depend upon the context to determine which usage is intended.

intransitive	$\forall xyz \in A : xRy \;\&\; yRz \;.\!\rightarrow \sim xRz$
symmetric	$\forall xy \in A : xRy \rightarrow yRx$
asymmetric	$\forall xy \in A : xRy \rightarrow \sim yRx$
antisymmetric	$\forall xy \in A : xRy \;\&\; yRx \;.\!\rightarrow x{=}y$
strongly-connected	$\forall xy \in A : xRy \vee yRx$
weakly-connected	$\forall xy \in A : x{=}y \vee xRy \vee yRx$

So, for example, R is *reflexive* in A if and only if every element in A bears R to itself (note the *reflexive* pronoun here)..

Every predicate defined above is a two-place predicate, and expresses a relation between R and A. We can define an associated one-place predicate by specifying A to be the field of R. This yields a corresponding series of definitions, founded on the following paradigm.

$$R \text{ is } \Phi \quad =_{\text{df}} \quad R \text{ is } \Phi \text{ in } \textit{field}(R)$$

In addition to the basic properties of relations listed above, there are various combinations that are important, of which we list just a few.

R **is a(n):**	**if and only if** R **is:**
quasi-order relation	reflexive and transitive
partial-order relation	reflexive, transitive, and anti-symmetric
strict-partial order relation	irreflexive and transitive (and hence asymmetric)
linear-order relation	reflexive, transitive, anti-symmetric, and strongly connected
strict-linear-order relation	irreflexive, transitive, and weakly connected.
equivalence relation	reflexive, symmetric, and transitive.

Note that an alternative terminology for order-relations is frequently used; for example, a partial-order relation is also called a *partial-ordering*, or a *partially-ordered set*. Also, notice that every predicate in this list is a one-place predicate. The corresponding two-place predicate is defined in accordance with the following paradigm.

$$R \text{ is a } \Phi \text{ relation on } A \quad =_{\text{df}} \quad R \text{ is a } \Phi \text{ relation, and } \textit{field}(R) = A$$

23. Tree-Orderings

Among partial-order relations are a special type called *tree-order relations*, or tree-structures, or simply *trees*. What makes a tree-ordering special among partial-orderings is that a tree-ordering only branches in one direction (at most).

Tree-ordered structures occur frequently in nature. First of all, "real" trees exemplify this structure – one branch can split into two smaller branches, but two branches never combine into a single larger branch. Another biological example of tree-ordering are evolutionary trees; it is a fundamental thesis of cladistics (mathematical evolutionary taxonomy) that branching occurs in one direction only – one taxonomic group can split into two groups, but two groups cannot combine into a single group. For example, although Reptiles and Humans may have a common ancestor, they do not, and never will, have a common descendent. Another abstract example are possible-world-histories, which can be viewed as tree-ordered in the sense that, although we have alternative futures available to us, we do not have alternative pasts available to us. Natural river systems are another example of a tree-structure. Characteristically, a river only branches in one direction – up-stream; it never branches down-stream. Two rivers can feed into one river, but one river cannot feed into two rivers.[36]

Curiously, family trees are not trees in the technical sense, because people have both common ancestors and common descendents. On the other hand, if one traces one's own ancestral lineage (one's own family tree), then the resulting structure is indeed a tree – *provided* there is no incest![37] On the other (third!) hand, if we all descended from a small group of (say, two) ur-humans, then there *had* to be incest! So one has a family *tree* up to the point one encounters incest.

Mathematically, a tree-ordering may be defined as follows.

A tree-ordering is a relation, $\leqslant$, satisfying the following conditions.

$$x \leqslant x$$
$$x \leqslant y \;\&\; y \leqslant z \;.\!\rightarrow\; x \leqslant z$$
$$x \leqslant y \;\&\; y \leqslant x \;.\!\rightarrow\; x = y$$
$$x \leqslant y \;\&\; x \leqslant z \;.\!\rightarrow.\; y \leqslant z \vee z \leqslant y$$

Notice that the first three conditions are characteristic of partial-order relations. One could equally well write this definition using the converse relation $\geqslant$. By using $\leqslant$, we have in effect chosen to have trees branch downward in the manner of syntactic-trees, rather than upward in the manner of botanical-trees, or evolutionary-trees (evolutionary time proceeds upward in analogy with the fossil record).

24. Grammatical-Trees

The above definition describes a genus – trees. Grammatical-trees are a species of this genus, involving special features. First, we define the domination-relation as follows.

$$x \text{ dominates } y \;=_{\text{df}}\; x > y \;=_{\text{df}}\; y \leqslant x \;\&\; \sim[x \leqslant y]$$

Notice that the domination-relation satisfies the following conditions.

[36] A river delta is an exception to this principle, but usually all the parts of a delta are named the same. So at least nominally the branching principle is maintained.

[37] Incest is broadly understood to mean the following – a descends from b and c ($b \neq c$), who in turn both descend from d.

$\sim[x > x]$
$x > y \;\&\; y > z \;.\!\rightarrow\; x > z$
$x > z \;\&\; y > z \;.\!\rightarrow\!.\; x > y \;\vee\; y > x \;\vee\; x = y$

Second, every syntactic-tree is finite. Third, every syntactic tree has a single common ancestral node (usually an S-node), as follows.

$\exists x \forall y [y \leqslant x]$

Finally, if we are so inclined, we can specify that syntactic trees possess a "left-right" structure in addition to an "up-down" structure. This requires adding a further order-relation, say $\leqq$, together with the following postulate.

$\leqq$ is a linear-order relation

Exactly how the left-right relation is supposed to interact with the up-down relation (e.g., whether branches are allowed to cross) is not entirely settled.

25. Multi-Place Functions and Relations

So far, we have only discussed *unary* functions and *binary* relations. Since formal semantics requires multi-place functions and relations as well, we need a formal account of these concepts. There are several alternatives available. According to one alternative:

we define ordered-triples in terms of ordered-pairs,
    and on the basis of these we define 3-place predicates,
        and on the basis of *these* we define 2-place functions;
we define ordered quadruples in terms of ordered triples,
    and on the basis of these we define 4-place relations,
        and on the basis of *these* we define 3 place functions;
etc.

According to a more elegant alternative, we first define the natural numbers (0, 1, 2, etc.), and then use these items to define sequences, and in particular k-tuples, and then use *these* items to define k-place relations and k-place functions. Since this is more elegant, and since it is more easily generalized, and since we need to define the natural numbers anyway, this is the alternative we choose.

26. The Natural Numbers

We can treat the natural numbers as primitive objects (ur-elements), or we can treat them as pure sets. It doesn't really matter to semantics which approach we adopt, but it is helpful for later set-theoretic constructions to define them as pure sets. In particular, later numerical notation officially presupposes the following definition.

A natural number is identified with the set of all its predecessors;
in other words:

$0 =_{\text{df}} \varnothing$ [since 0 has no predecessors]
$1 =_{\text{df}} \{\varnothing\}$ [since 1 has exactly one predecessor – 0 ($\varnothing$)]
$2 =_{\text{df}} \{\varnothing, \{\varnothing\}\}$ [since 2 has exactly 2 predecessors – 0($\varnothing$) and 1($\{\varnothing\}$)]
etc.

$\mathbb{N} =_{\text{df}} \{0, 1, 2, \ldots\}$[38]

[38] This is informal, but rest assured that it can be made formally exact. Suffice it to say that $\mathbb{N}$ is the set that contains all and only natural numbers. The existence of this set is guaranteed by the Axiom of Infinity.

27. *K*-Tuples; Finite Sequences

Once we have the natural numbers, we can *officially* introduce all manner of numerical notions and notation. The following is our first example.

> Where k is a natural number, a *k-tuple* is, by definition, any function whose domain is k.
> A *finite sequence* is, by definition, a k-tuple, for some number k.

By themselves, these definitions are not very helpful in understanding the set-theoretic nature of k-tuples (finite sequences). They only become helpful when combined with a standard notational convention, which we now present.

Recall that it is common to display a set by listing its elements and surrounding the resulting list by curly-braces, as in the following tripleton.

{Mozart, Bach, Beethoven}

It is equally common to display a k-tuple by listing its components *in order* and surrounding the resulting list by corner-brackets, as in the following 3-tuple.

⟨Mozart, Bach, Beethoven⟩

The chief difference between a tripleton and a 3-tuple is that, whereas the latter is *sorted*, the former is not.

As officially defined, a 3-tuple is a function whose domain is 3. So, what function is the 3-tuple ⟨Mozart, Bach, Beethoven⟩? Well, it's a 3-tuple, so its domain is 3, which is the set {0,1,2}, so the function in question assigns an item to each element of {0,1,2}. What item? Well there is an obvious order to the elements of {0,1,2}, which is presumably inherited by the elements of ⟨Mozart, Bach, Beethoven⟩. Specifically, the function in question is the following pairing.[39]

0 → Mozart
1 → Bach
2 → Beethoven

More generally, we have the following notational convention.

> $\langle\sigma_0, \sigma_1, \ldots, \sigma_{k\text{-}1}\rangle$ $=_{\text{df}}$ the function f such that $domain(f) = k$,
> and such that
> for any $i \in k$, $f(i) = \sigma_i$.

Notice that, in the above definition, subscript notation is an informal part of the metalanguage. We can also officially introduce subscript notation into the object language as a mere notational variant of function-application notation. The following are the relevant definitions.

> Let σ be a k-tuple; then: $\sigma_i =_{\text{df}} \sigma(i)$
>
> $\langle\sigma_0, \ldots, \sigma_{k\text{-}1}\rangle =_{\text{df}} \sigma$

In summary, if a function σ has natural number k as its domain, in which case σ qualifies as a k-tuple, then we can display σ by listing its values in numerical order, and enclosing the list in corner brackets. For example, the 3-tuple

[39] It is fairly common, and natural, to display a relation/function as a "graph", in which the arrows mark which ordered-pairs are elements of the relation/function.

$\langle$Mozart, Bach, Beethoven$\rangle$

is officially the function σ such that[40]

$$\begin{aligned}\sigma_0 &=_{df} \sigma(0) = \text{Mozart}\\ \sigma_1 &=_{df} \sigma(1) = \text{Bach}\\ \sigma_2 &=_{df} \sigma(2) = \text{Beethoven}\end{aligned}$$

By way of illustrating the notation, observe that the following are intelligible, and are in fact true.

$$\begin{aligned}\langle\text{Mozart, Bach, Beethoven}\rangle(0) &= \text{Mozart}\\ \langle\text{Mozart, Bach, Beethoven}\rangle(1) &= \text{Bach}\\ \langle\text{Mozart, Bach, Beethoven}\rangle(2) &= \text{Beethoven}\end{aligned}$$

28. Index-Shifting

The notation presented in the previous section is the orthodox notation. According to this notation, the 1st component of sequence σ is σ_0, the 2nd component is σ_1, and generally, the i-th component of σ is $\sigma_{i\text{-}1}$.

The intuitive discrepancy between ordinal numbers and subscripts is annoying to many, including the present author. Accordingly, we find it convenient to adopt an alternative notation according to which the i-th component of a sequence σ is σ_i, not $\sigma_{i\text{-}1}$. This can be accomplished either formally or informally. Specifically, we can go back and re-do all our official definitions[41]; or, we can introduce a few more informal identities. We do a combination of the two. First, we adjust the official definition of corner-bracket notation, as follows.

$\langle\sigma_1, \ldots, \sigma_k\rangle$	$=_{df}$	the sequence whose first component is σ_1, …, whose second component is σ_2, …, and whose k-th component is σ_k.

Since this definition uses subscripts informally in the metalanguage, it is not inconsistent with our other definitions; the inconsistency arises in the following principle.

$$\langle\sigma_1, \ldots, \sigma_k\rangle(i) = \sigma_i$$

e.g.,

$$\begin{aligned}\langle\text{Mozart, Bach, Beethoven}\rangle(1) &= \text{Mozart}\\ \langle\text{Mozart, Bach, Beethoven}\rangle(2) &= \text{Bach}\\ \langle\text{Mozart, Bach, Beethoven}\rangle(3) &= \text{Beethoven}\end{aligned}$$

Irrespective of how we implement this change, we henceforth think of most sequences as starting with index 1, although there are occasions in which starting with index 0 is also convenient.

29. Cartesian-Exponentiation (Powers)

Just as we can "multiply" two sets A and B, in the sense of forming their Cartesian-product $A\times B$, we can "repeatedly multiply" a single set A, or "raise" it to a power. This process, called *Cartesian-exponentiation*, might plausibly be defined as follows.

$$\begin{aligned}A^1 &=_{df} A\\ A^2 &=_{df} A\times A\\ A^3 &=_{df} A\times A\times A\\ A^4 &=_{df} A\times A\times A\times A\end{aligned}$$

etc.

[40] We speak a bit loosely here. There are infinitely-many functions satisfying the conditions under consideration. By ‘the function such that Φ’ we usually mean ‘the smallest function such that Φ’.

[41] For example, we can officially define a k-tuple to be a function from the set $(k+1)-\{0\}$.

There are two problems with this approach, the first one fairly minor, the other more serious.

(1) Cartesian-multiplication, which this approach presupposes, is not associative – generally, $(A \times B) \times C \neq A \times (B \times C)$; so we are not warranted in using the notation in the manner we did in the above definitions.

(2) This approach does not easily generalize to arbitrary forms of exponentiation. Recall that ordinary exponentiation is defined for arbitrary real numbers.[42]

For this reason, we adopt the usual official definition of Cartesian-exponentiation.

$A^B =_{df}$ the set of all functions from B into A	[set-exponentiation]
i.e., $A^B =_{df} B \rightarrow A$	

Now, the obvious question is – how is Cartesian-exponentiation, so defined, even remotely like exponentiation? Less rhetorically, how does Cartesian-exponentiation relate to Cartesian-multiplication; for example, what is the relation between A^2 and $A \times A$?

Let us explore the latter question. First, by definition,

$$A^2 = \text{ the set of functions from } 2 \text{ into } A$$

But a function whose domain is 2 [i.e., $\{0,1\}$, i.e., $\{\varnothing, \{\varnothing\}\}$] is what we call a 2-tuple. So, using our earlier notation for 2-tuples, we can write the following.

$$A^2 = \{ \langle x,y \rangle = x \in A \ \& \ y \in A \}$$

On the other hand, according to the definition of Cartesian-product, we have

$$A \times A = \{ (x,y) = x \in A \ \& \ y \in A \}$$

Thus, the only difference between A^2 and $A \times A$ is that, whereas A^2 consists of 2-tuples of elements of A, $A \times A$ consists of ordered-pairs of elements of A.

Now, officially 2-tuples are not the same things as ordered-pairs; generally, we have:

$$\langle a,b \rangle \neq (a,b)$$

2-tuples are defined as functions, which are defined in terms of ordered-pairs.[43] Nevertheless, we can *informally* identify them, so that:

$$\langle a,b \rangle \approx (a,b)$$

in which case we have the following further informal identity.

$$A^2 \approx A \times A$$

More generally, we have the following.

$$A^0 = 1 \ [=_{df} \{\varnothing\}]$$
$$A^1 \approx A$$
$$A^2 \approx A \times A$$
$$A^3 \approx A \times (A \times A)$$
etc.

Finally, we give the official definition of k-fold Cartesian power.[44]

$A^k =_{df} k \rightarrow A$ [= the set of all k-tuples of elements of A]

[42] For example, 5^{π} is perfectly well-defined, even though it is impossible to understand it as "5 × … × 5 … π-many times"!

[43] Officially, $(a,b) = \{ \{a\}, \{a,b\} \}$, whereas $\langle a,b \rangle = \{ (0,a), (1,b) \}$. The latter expression can be further expanded, which is left as an exercise for the reader.

[44] Admittedly, "Cartesian power" sounds like a slogan promoting a metaphysics that derives from Descartes.

30. Multi-Place Functions and Relations Revisited

We are now in position to define polyadic (multi-place) functions and relations.

Where k is any natural number,

R is a k-place relation on A	$=_{\text{df}}$	$R \subseteq A^k$
f is a k-place function from A	$=_{\text{df}}$	f is a (unary) function from A^k
f is a k-place function on A	$=_{\text{df}}$	f is a (unary) function from A^k into A

For example, a k-place relation on A is a set of k-tuples of elements of A, and a k-place function on A is a function that assigns an element of A to every k-tuple of elements of A.

Notice that we officially allow k-place relations and k-place functions for any natural number k. This raises a few natural questions. What is a 0-place relation, a 0-place function, a 1-place relation? The following are some of the relevant informal identifies.

(1)	a 2-place relation on A	$\approx$	a binary relation on A
(2)	a 1-place function from A into B	$\approx$	a unary function from A into B
(3)	a 1-place relation on A	$\approx$	a subset of A
(4)	a 0-place function from A into B	$\approx$	an element of B
(5)	a 0-place relation on A	$\approx$	an element of $\{ \varnothing, \{\varnothing\} \}$ $[=\{0,1\}]$

(1)-(3) may be obtained from the following informal identifies.

$$\langle a,b\rangle \approx (a,b)$$
$$\langle a\rangle \approx a$$

Concerning (4), we note that a 0-place function from A into B is, by definition, a function from A^0 into B. But A^0 is the set of all functions from 0 (i.e., $\varnothing$) into A, but there is only one such function, the trivial function $\varnothing$. So, $A^0 = \{\varnothing\}$. So suppose f is a function from A^0 into B. Then f assigns a value in B to $\varnothing$, and to nothing else. Accordingly, we can informally identify the function f with $f(\varnothing)$ $[= f\langle\rangle]$.

Concerning (5), a 0-place relation on A is, by definition, a subset of A^0, which as we saw in connection with (4) is the set $\{\varnothing\}$. Now, $\{\varnothing\}$ has exactly two subsets – namely $\varnothing$ and $\{\varnothing\}$. Accordingly, a 0-place relation on A must be either $\varnothing$ or $\{\varnothing\}$ (i.e., 0 or 1).

31. Anadic Functions

Before continuing, we note the following synonyms.

monadic	unary	1-place
dyadic	binary	2-place
triadic	ternary	3-place
etc.		

Also, we note the following terms.

polyadic	2-place or 3-place or …
anadic	"any"-place

More specifically, to say that f is a *polyadic* function from A into B is to say that f is a 2-place, or 3-place, or …, function from A into B. This can be written officially as follows.

$$f \in \bigcup\{A^2 \rightarrow B, A^3 \rightarrow B, \ldots\} = \bigcup\{A^k \rightarrow B : k = 2, 3, \ldots\}$$

The term ‘polyadic’ must be carefully distinguished from the other term ‘anadic’. The prefix ‘an’ means ‘without’, so an anadic function is one without "adicity", which is to say that it is not *specifically* 1-place, 2-place, or any *particular* place. This can be set-theoretically written as follows.[45]

$$f \in A^{a} \rightarrow B$$

where

$$A^{a} =_{df} \bigcup\{A^{1}, A^{2}, \ldots\}$$

In other words, f assigns a value in B to every 1-tuple, every 2-tuple, every 3-tuple, etc., in A.

32. Indexed Sets; Families

Recall that a finite sequence (k-tuple) is a function from k. This notion can be considerably generalized – to the notion of *indexed set* (*family*). Whereas a finite sequence is a function whose domain is a particular natural number k, an indexed set (family) is a function whose domain is an arbitrary set I of "indices". Materially speaking, then, an indexed set is just a function. So what is the big deal? As with finite sequences, the usefulness comes not in the identity conditions, but in the corresponding notation – indexed sets (families) are not a new *thing*, but a new *notation*.

As we propose to implement it,[46] the notation for indexed sets parallels set-abstract notion, in the same way that k-tuple notation parallels simple-set notation. The following is the simple, and most commonly used, form

$$\langle \sigma_i : i \in I \rangle$$

which is a special case of the following more abstract form,

$$\langle \tau : \mathbb{F} \rangle$$

where τ is any singular term, and $\mathbb{F}$ is any formula. The following is the official definition.[47]

$\langle \tau : \mathbb{F} \rangle \quad =_{df} \quad \{ \langle \langle \nu_1, \ldots, \nu_k \rangle, \tau \rangle : \mathbb{F} \}$
where $\langle \nu_1, \ldots, \nu_k \rangle$ is a non-repeating sequence of the free variables common to τ and $\mathbb{F}$

Note that, as a special case, we have the following definition of the simple form.

$\langle \sigma_i : i \in I \rangle \quad =_{df} \quad \{ \langle i, \sigma_i \rangle : i \in I \}$

Note the following re-description of the *definiens*.

the function, with domain I, that assigns σ_i to each i in I

Next, we note the following very important notion..

An *infinite sequence* is, by definition, a function from $\mathbb{N}$.

This is the orthodox line. If we are averse to having item number 1 be the second component, we can liberalize the definition as follows.

---

[45] This condition imposes a finitely-many restriction on the input arguments. We could correspondingly describe a *completely-anadic* function from A into B simply as a function from $\wp(A)$ into B; such a function assigns a value in B to every subset of B.

[46] There exists a variety of ways of notating indexed sets; we find our way best!

[47] Notice that we hereby adopt the informal convention of writing "$\langle a,b \rangle$" in place of the official "(a,b)". Having painstakingly climbed the ladder from sets to ordered pairs, to relations, to functions, to 2-tuples, we follow Wittgenstein's advice and kick the ladder out from under us!

An *infinite sequence* is, by definition, a function from $\mathbb{N}$,
or a function from $\mathbb{N} - \{\varnothing\}$

Infinite sequences can be notated in a variety of ways, including the following.

$\langle\sigma_0, \sigma_1, \ldots\rangle$	$\langle\sigma_1, \sigma_2, \ldots\rangle$
$\langle\sigma_i : i = 0, 1, 2, \ldots\rangle$	$\langle\sigma_i : i = 1, 2, 3, \ldots\rangle$[48]
$\langle\sigma_i : i \in \mathbb{N}\rangle$	$\langle\sigma_i : i \in \mathbb{N}+\rangle$[49]

33. Family Values

In order to illustrate how one might use families (indexed sets), we show how to notate various functions using indexing notation. For example, consider the addition-function, Σ, defined on the set of natural numbers as follows.

$$\Sigma =_{df} \{ \langle \langle x,y\rangle, z \rangle : x, y \in \mathbb{N} \ \& \ z=x+y \}$$ [50]

In other words, for any $x, y \in \mathbb{N}$, we have:

$$\Sigma\langle x,y\rangle = x+y$$

Writing Σ in family-notation, we have:

$$\Sigma = \langle x+y : x, y \in \mathbb{N}\rangle$$

Before continuing, it is very important to appreciate the difference between

the **family**: $\langle x+y : x, y \in \mathbb{N}\rangle$

and

the **set**: $\{x+y : x, y \in \mathbb{N}\}$

which look superficially quite similar. Whereas the former is a function, the latter is the set of all the output values of this function; in other words, it is what we have previously called the *range* of the function. Notice in particular, that

$$\{x+y : x, y \in \mathbb{N}\} = \mathbb{N}$$

since every natural number can be obtained by adding two natural numbers together. On the other hand,

$$\langle x+y : x, y \in \mathbb{N}\rangle \neq \mathbb{N}$$

since the addition-function is surely not identical to the set of natural numbers.

Next, we show that the addition-function Σ, which is a two-place function, can be re-notated as a family of one-place functions. To accomplish this, we first make the following list of one-place functions.

$\langle 0+y : y \in \mathbb{N}\rangle$	[takes a number and adds it to 0]
$\langle 1+y : y \in \mathbb{N}\rangle$	[takes a number and adds it to 1]
$\langle 2+y : y \in \mathbb{N}\rangle$	[takes a number and adds it to 2]
$\langle 3+y : y \in \mathbb{N}\rangle$	[takes a number and adds it to 3]
etc.	

We next "collect" this list into the following infinite sequence (orthodox definition!)

$$\big\langle \langle 0+y : y \in \mathbb{N}\rangle, \langle 1+y : y \in \mathbb{N}\rangle, \langle 2+y : y \in \mathbb{N}\rangle, \ldots \big\rangle$$

Finally, we observe that this sequence can be re-notated as follows.

$$\big\langle \langle x+y : y \in \mathbb{N}\rangle : x \in \mathbb{N} \big\rangle$$

---

[48] The notation "$i = 1, 2, 3, \ldots$" is short for "$i = 1 \vee i = 2 \vee i = 3 \vee \ldots$"

[49] $\mathbb{N}+$ is the set of positive natural numbers – i.e., all the natural numbers except 0.

[50] Although it may go without saying, the notation "$x, y \in \mathbb{N}$" is short for "$x \in \mathbb{N} \ \& \ y \in \mathbb{N}$".

Now, suppose we use 'Σ' ambiguously for both the original two-place function and the corresponding family of one-place functions. We then have the following.

$$\Sigma\langle x\rangle\langle y\rangle \;=\; x+y \;=\; \Sigma\langle x,y\rangle$$ [51]

Or using subscript notation for the family, we have:

$$\Sigma_x\langle y\rangle \;=\; x+y \;=\; \Sigma\langle x,y\rangle$$

34. The Generalized Cartesian-Product

Earlier, we considered the k-fold Cartesian power of a set, which was defined as follows.

$$A^k \;=_{\mathrm{df}}\; k\!\rightarrow\! A \quad [= \text{the set of all } k\text{-tuples of elements of } A]$$

This technique can be considerably generalized. As a first step, we can multiply any finite sequence of sets. The following is the official definition.

> Let $\langle A_1, \ldots, A_k\rangle$ be a finite sequence of sets. Then
>
> $$\times\langle A_1, \ldots, A_k\rangle \;=_{\mathrm{df}}\; \{\, \langle a_1, \ldots, a_k\rangle : a_1 \in A_1, \ldots, a_k \in A_k \,\}$$

In other words, an element of the Cartesian-product of a k-tuple of sets is any k-tuple of respective elements of those sets. Notice that we have the following (strict) identities.

$$A^1 \;=\; \times\langle A\rangle$$
$$A^2 \;=\; \times\langle A,A\rangle$$
$$A^3 \;=\; \times\langle A,A,A\rangle$$
etc.

Next, since a k-tuple of sets $\langle A_1, \ldots, A_k\rangle$ is just a special case of an indexed set $\langle A_i : i \in I\rangle$, we can generalize the notion of Cartesian-product as follows.

> Let $\langle A_i : i \in I\rangle$ be a family of sets. Then
>
> $$\times\langle A_i : i \in I\rangle \;=_{\mathrm{df}}\; \{\, \langle a_i : i \in I\rangle : \text{for any } i \in I,\, a_i \in A_i \,\}$$

In other words, an element of the Cartesian-product of a family of sets is any family of respective elements of those sets.[52]

35. Cardinality

A critical notion of set theory is the notion of *cardinality*; indeed, set theory was originally invented (by Cantor) to characterize cardinality.[53] The cardinality of a set is its size (power). For finite sets, the notion of size is unproblematic; we all know intuitively what it means to say that set S has size k, where k is a natural number. Our intuitions are less reliable when it comes to infinite sets.

Before continuing, perhaps we should ask whether there are any infinite sets; because, if there aren't any, we don't have to worry about how big they are! Now, in Section 4, we

---

[51] At this point, notice that we are use tuple notation even for unary functions; recall $f\langle a\rangle \approx f(a)$.

[52] At this point, if we worry about whether the sets we describe are in fact well-defined, we come face-to-face with one of the more esoteric and controversial postulates of modern set theory – The Axiom of Choice. It has numerous formulations, many pairs of which barely look alike. My favorite formulation goes as follows:

> the Cartesian product of any non-empty family of non-empty sets is non-empty.

[53] This is fairly common in the history of thought – a conceptual problem leads to the invention of an entire branch of mathematics. For example, Newton invented the differential and integral calculus *in order to* solve the problems of planetary mechanics.

demonstrated that there are infinitely-*many* sets. On the other hand, at no point have we demonstrated that there are infinitely-*large* sets. This in fact requires its own special postulate, which is known as the *Axiom of Infinity*. According to one formulation (see Section 37), the Axiom of Infinity claims that there does indeed exist a set that contains the sets mentioned in Section 4. Therefore, since there are infinitely-*many* such sets, there is at least one infinitely-*large* set.

The question then is – how do we compare the sizes of infinite sets? In what follows, we briefly review what Cantor proposed by way of solving this problem.

First, recall the following definition.

> A function *from* A *into* B is, by definition, a function f satisfying the following conditions.
>
> $domain(f) = A$
> $range(f) \subseteq B$

Closely related to the notion of *into* is the notion of *onto*, which is defined as follows.

> A function *from* A *onto* B is, by definition, a function f satisfying the following conditions.
>
> $domain(f) = A$
> $range(f) = B$

Next, recall that a function f can assign the same value (output) to two different arguments (input). For example, although no one has two f's (e.g., fathers), two people can have the same f (e.g., father) . Functions that assign different output to different input are of special interest, and are said to be *one-one (1-1, one-to-one*), which are defined as follows.

> A function f is said to be *one-one* precisely if it satisfies the following condition.
>
> $\forall x \forall y (f(x) = f(y) \;\rightarrow\; x = y)$

Putting *one-one* together with *onto*, we arrive at the notion of a *one-to-one correspondence*,[54] which is officially defined as follows.

> f is said to be a *one-to-one correspondence* between A and B precisely if f is a one-one *function* from A onto B.

Once we have the notion of one-to-one correspondence, we can characterize *sameness-of-size*, also called *equipollence*, denoted $\cong$, in the manner first proposed by Cantor.

> $A \cong B \;\;=_{\text{df}}\;\;$ there is a one-to-one correspondence between A and B

In the case of finite sets, equipollence is intuitively clear. If a set has k-many elements, then it is equipollent to every other set with k-many elements. Indeed, a set A has k-many elements if and only if it is equipollent to the number k (regarded as the set of its k-many predecessors). Officially, a set is said to be *finite* precisely when it is equipollent to

[54] One-to-one correspondences are also called *bijections*. Along the same lines, one-one functions are called *injections*, and onto functions are called *surjections*.

some natural number. On the other hand, a set is said to be *infinite* precisely when it is not equipollent to any natural number. For example, the set ℕ of natural numbers is infinite. A set is said to be *denumerable* precisely when it is equipollent to the set ℕ. Notice that every denumerable set is automatically infinite.

Equipollence among infinite sets is somewhat bizarre. For example, every infinite subset of ℕ is equipollent to ℕ. This is in fact a general characteristic of infinite sets – they have proper subsets with the very same size! Examples of denumerable proper subsets of ℕ include the even numbers, the odd numbers, the prime numbers, just to mention a few. ℕ also has denumerable proper supersets, including the integers, and the rational numbers. Intuitively, it would seem that there are vastly more rational numbers than there are integers; after all, between any two integers there are infinitely-many rational numbers; indeed, between any two rational numbers there are infinitely-many *more* rational numbers. Nevertheless, as Cantor first showed, the rational numbers can be enumerated – i.e., put in a one-to-one correspondence with the natural numbers. So, contrary to initial intuition, there are no more rational numbers than there are natural numbers!

Upon learning that the rational numbers are denumerable, one might naturally conjecture that all infinite sets are denumerable; in other words, infinite is infinite. However, once again, our intuitions are bludgeoned. As Cantor first proved, the power set ℘(ℕ) is *uncountable*, as is the set of irrational numbers, and the set of real numbers. A set is said to be *uncountable* precisely when it is neither finite nor denumerable; otherwise, it is said to be *countable*. Notice that the countably-infinite (i.e., countable and infinite) sets coincide with the denumerable sets.[55]

OK, we have infinitely-many finite sizes, and we have two infinite sizes. Are there any more? Yes indeed! As Cantor first proved, no matter how big a set is, its power set is bigger. So, ℕ is smaller than ℘(ℕ), which is smaller than ℘℘(ℕ), which is smaller than ℘℘℘(ℕ), and so forth.[56]

[55] Note carefully that the terminology in set theory is not completely standardized. In particular, some authors treat 'denumerable' and 'countable' as synonyms, so that a finite set is denumerable according to this usage.
[56] What is worse, there are cardinalities (sizes) that are impossible to construct set-theoretically – the so-called *inaccessible* cardinals. So far, however, the already-staggering universe of set-theoretically constructible objects seems to be more than adequate for formal semantic purposes. But one never knows!

Summary:

Size	Definition
S is **finite**	S is equipollent to k, for some natural number k
S is **infinite**	S is not finite
S is **denumerable**	S is equipollent to the set $\mathbb{N}$ of natural numbers
S is **countable**	S is finite or denumerable
S is **uncountable**	S is not countable
S is **countably-infinite**	S is infinite and countable

36. Ordinal Numbers

In addition to cardinal numbers, there are also ordinal numbers, and just as there are infinite cardinal numbers, there are infinite ordinal numbers. Generally speaking, whereas cardinal numbers pertain to *size* (in particular, how-many), ordinal numbers pertain to *rank* or *relative position*. The distinction is perhaps initially best understood by the words we employ to express them. Whereas the cardinal-*numerals*[57] are 'one', 'two', 'three', etc., the ordinal-*numerals* are 'first', 'second', 'third', etc. This is not a hard and fast distinction, however. For example, when sports fans exclaim "we're number one" they are not using 'one' to say their *size*, but rather their *rank*. So, even 'one' is occasionally used to express an ordinal concept. Similarly, house numbers and room numbers have ordinal-significance rather than cardinal-significance. Some "numbers" – for example, social security numbers, id numbers, and phone numbers – have neither cardinal nor ordinal significance.

What are the ordinal numbers? This is a long story, but suffice it to say that the following can serve as a fairly simple set-theoretic definition of ordinal number.

> an *ordinal number* is, by definition,
> a transitive set every element of which is a transitive set

The following is the affiliated definition of 'transitive set'.

> a set A is said to be *transitive* precisely if every element of A is a subset of A

In symbols the *definiens* goes as follows, which will help explain the terminology 'transitive'.

$$\forall y \{ y{\in}A \rightarrow \forall x \{ x{\in}y \rightarrow x{\in}\mathrm{A} \} \}$$

Notice in particular that the latter is logically equivalent to:

$$\forall x \forall y \{ x{\in}y \;\&\; y{\in}A \;.{\rightarrow}\; x{\in}\mathrm{A} \}$$

What are some examples of transitive sets? Well, for starters, we have a trivial example – the empty set $\varnothing$. It has no elements, so trivially every one of them is a subset of

[57] I take a "numeral" to be any symbol – simple or complex, logographic or phonographic – that expresses a number. So the word 'one' and the Arabic-numeral '1' and the Roman-numeral 'I' are all numerals that express the number 1.

$\varnothing$. Next, we note that $\{\varnothing\}$ is transitive. $\{\varnothing\}$ has exactly one element, and sure enough it is also a subset. More generally, if we take the *successor* of a transitive set, we obtain another transitive set, where the relevant definition is given as follows.

$$A^+ \quad =_{\mathrm{df}} \quad A \cup \{A\}$$

What are some examples of ordinal numbers? Once again, $\varnothing$ is our first example. Every element of $\varnothing$ is a subset, and every element of $\varnothing$ is also transitive, so $\varnothing$ is an ordinal number. We can also prove the following.

if A is an ordinal number, then so is A^+

In this connection, note in particular the following identities.

$$0^+ = \varnothing^+ = \varnothing \cup \{\varnothing\} = \{\varnothing\} = 1$$
$$1^+ = \{\varnothing\}^+ = \{\varnothing\} \cup \{\{\varnothing\}\} = \{\varnothing, \{\varnothing\}\} = 2$$
$$2^+ = 3$$

etc.

This can be summarized as follows.

every natural number (finite cardinal number) is also an ordinal number

Are there any other ordinal numbers? For starters, consider the set $\mathbb{N}$ of all natural numbers. It is fairly routine to demonstrate that $\mathbb{N}$ is also an ordinal number. When $\mathbb{N}$ is thought of as an ordinal number it is given a special name – ω (omega) – which is the first infinite ordinal number. As noted earlier, the successor of any ordinal number is itself an ordinal number. This means we already have the following two infinite sequences of ordinal numbers.

$$0 \quad 0^+ \quad 0^{++} \quad 0^{+++} \quad \ldots$$

$$\omega \quad \omega^+ \quad \omega^{++} \quad \omega^{+++} \quad \ldots$$

It is customary to call these respectively,

$$\omega \quad \omega+1 \quad \omega+2 \quad \omega+3 \quad \ldots$$

Let us refer to this collection as "ω and all its immediate descendants", and let us call the set of natural numbers "0 and all its immediate descendants". An immediate descendant can be reached by counting.

There is an ordinal number that is immediately after 0 and all its descendants. What about ω? Is there an ordinal number immediately after ω and all its descendants? The standard answer is 'yes'. Just as the "limit" of $\langle 0,1,2,\ldots\rangle$ is ω, the "limit" of $\langle \omega,\omega+1,\omega+2,\ldots\rangle$ is $\omega+\omega$, also known as $\omega\times 2$ [also $\omega 2$]. How is $\omega 2$ defined? It is the smallest set containing ω, $\omega+1$, $\omega+2$, etc., or informally:

$$\omega 2 \quad =_{\mathrm{df}} \quad \{\omega, \omega+1, \omega+2, \ldots\}$$

One can demonstrate that this set is an ordinal number that is immediately after ω, $\omega+1$, $\omega+2$, …

Now, $\omega 2$ has a sequence of immediate descendants

$$\omega 2 \qquad \omega 2+1 \qquad \omega 2+2 \qquad \omega 2+3 \qquad \ldots$$

This sequence also has a limit, which is $\omega 2+\omega$ [also known as $\omega\times 3$ and $\omega 3$]. One can prove this is an ordinal number. It too engenders a set of immediate descendants.

$$\omega 3 \qquad \omega 3+1 \qquad \omega 3+2 \qquad \omega 3+3 \qquad \ldots$$

And the limit of this sequence is $\omega 4$, which produces yet another sequence of descendants, whose limit is $\omega 5$, and so forth!

What we have constructed so far can be summarized in the following array.

sequence					**limit**
0	1	2	3	…	ω
ω	$\omega+1$	$\omega+2$	$\omega+3$	…	$\omega 2$
$\omega 2$	$\omega 2+1$	$\omega 2+2$	$\omega 2+3$	…	$\omega 3$
$\omega 3$	$\omega 3+1$	$\omega 3+2$	$\omega 3+3$	…	$\omega 4$
…	…	…	…	…	…

Now things get interesting. Let us jump up an "order of magnitude" and consider the following sequence of limit-ordinals from the previous array.

$$\omega \qquad \omega 2 \qquad \omega 3 \qquad \omega 4 \qquad \ldots$$

This too has a limit – i.e., an ordinal immediately after this set – which is $\omega\times\omega$ – also known as ω^2.

The entire process we have just described can in turn be repeated infinitely-many times to produce the following sequence of limit ordinals.

$$\omega \qquad \omega^2 \qquad \omega^3 \qquad \omega^4 \qquad \ldots$$

This sequence also has a limit, which is ω^ω [omega to the omega-power] which we will henceforth write thus: $\omega\uparrow\omega$.

We can also repeat the entire process infinitely-many times, to produce the following sequence.[58]

$\omega\uparrow\omega$	$\omega\uparrow\omega\uparrow\omega$	$\omega\uparrow\omega\uparrow\omega\uparrow\omega$	…
ω to the ω	ω to the ω to the ω	ω to the ω to the ω to the ω	…

This too has a limit, which is $\omega\uparrow\uparrow\omega$, which is an example of *second-degree exponentiation*.

If there is first-degree and second-degree exponentiation, there is also third degree, fourth-degree, etc. So we get the following infinite sequence.

$$\omega\uparrow\omega \qquad \omega\uparrow\uparrow\omega \qquad \omega\uparrow\uparrow\uparrow\omega \qquad \omega\uparrow\uparrow\uparrow\uparrow\omega \qquad \ldots$$

This sequence also has a limit – $\omega\uparrow^\omega\omega$. Now things seem to have gotten completely out of hand, but you guessed it – the process just keeps on going!

Now, here is the real kicker. Although we have just constructed a staggeringly complex collection of ordinal numbers, all the ordinal numbers we have constructed are countable!

[58] Exponentiation is not associative; $(2\uparrow3)\uparrow4 = 8\uparrow4 = 4096$, but $2\uparrow(3\uparrow4) = 2\uparrow81 = 2.42178 \times 10^{24}$ – a huge difference! We take the missing parentheses to be such that, for example, $2\uparrow3\uparrow4 = (2\uparrow3)\uparrow4$.

37. Appendix: The Axioms of Set Theory

The following is a list of axioms of set theory, in the manner proposed by Zermelo and Fraenkel (what is accordingly called ZF). Note that ZF does not include the Axiom of Choice (see footnote 52). Also, note that ZF is a *pure* set theory (no ur-elements), so the quantifiers range over sets. Note also that ZF is formulated with only one primitive symbol – '$\in$' (membership). Note also that the theory is *herein* formulated as a *second-order* theory – block letters are second-order variables; in particular, '$\mathbb{F}$' is a one-place predicate variable, and '$\mathbb{R}$' is a two-place predicate variable. Using second-order machinery is not critical to the presentation, but simply makes the relevant axioms easier to understand. Finally, note that whenever the existence of a set is postulated, this is signaled by using the variable 's'.

A1.	$\forall x \forall y\{ \forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y \}$	[Extensionality]
A2.	$\exists s {\sim} \exists x[x \in s]$	[Empty Set]
A3.	$\forall \mathbb{F} \forall x \exists s \forall y\{ y \in s \leftrightarrow . \, y \in x \,\&\, \mathbb{F}y \}$	[Separation]
A4.	$\forall x \forall y \exists s \forall z(z \in s \leftrightarrow . \, z = x \vee z = y)$	[Pairs]
A5.	$\forall x \exists s \forall y\{ y \in s \leftrightarrow \exists z(y \in z \,\&\, z \in x) \}$	[Unions]
A6.	$\forall x \exists s \forall y\{ y \in s \leftrightarrow \forall z(z \in y \rightarrow z \in x) \}$	[Power Sets]
A7.	$\exists s\{ \varnothing \in s \;\&\; \forall x(x \in s \rightarrow \{x\} \in s) \}$	[Infinity][59]
A8.	$\forall x\{ \exists y[y \in x] \rightarrow \exists y[y \in x \,\&\, {\sim}\exists z(z \in x \,\&\, z \in y)] \}$	[Regularity]
A9.	$\forall \mathbb{R}\!: \forall x \exists y \forall z(\mathbb{R}xz \leftrightarrow z = y) \rightarrow \forall x \exists s \forall y\{y \in s \leftrightarrow \exists z(z \in x \,\&\, \mathbb{R}zy)\}$	[Replacement]

An Aside:

A7 is written in the manner of Zermelo's original axiom. Notice that, unlike the others, A7 is not written in primitive notation, involving as it does both '$\varnothing$' and set-braces. It could be expanded into primitive notation, according to the following definitions.

$$\varnothing \;=_{\mathrm{df}}\; \iota s {\sim} \exists x[x \in s]$$
$$\{\tau\} \;=_{\mathrm{df}}\; \iota s \forall x(x \in s \leftrightarrow x = \tau) \qquad [x, s \text{ not free in } \tau]$$

However, if we wish to avoid definite-descriptions, and the myriad logical problems that plague them, we are faced with a more subtle task – to formulate "the" Axiom of Infinity in primitive notation.

The following is perhaps the simplest formula in primitive notation that succeeds in postulating an infinitely-large set.

A7*. $\exists s\{ \exists x[x \in s] \;\&\; \forall x(x \in s \rightarrow \exists y(x \in y \,\&\, y \in s) \}$

Note however that, unlike Zermelo's A7, the viability of A7* as an axiom of *infinity* depends upon the Axiom of Regularity – which disallows circular epsilon-chains (e.g., $a \in a$, $a \in b \in a$, $a \in b \in c \in a$, etc.). Without the Axiom of Regularity, A7* only succeeds in postulating the existence of at least one non-empty set – a fact that is not exactly newsworthy!

[59] Recall the sets $\varnothing$, $\{\varnothing\}$, $\{\{\varnothing\}\}$, … We have already shown that this list implicitly alludes to infinitely-many sets (see proof at end of Section 4). The Axiom of Infinity, A7, says that there is at least one set that contains all of them. It follows that there is at least one infinitely-large set. An alternative version of the Axiom of Infinity states that there is a set that contains all the natural numbers – $\varnothing$, $\{\varnothing\}$, $\{\varnothing,\{\varnothing\}\}$, etc.