VI : Infinite constructions in set theory

In elementary accounts of set theory, examples of finite collections of objects receive a great deal of attention for several reasons. For example, they provide relatively simple illustrations of the abstract formal concepts in the subject. However, Cantor's original motivation for studying set theory involved ***infinite*** collections of objects, and the real breakthrough of set theory was its ability to provide a framework for studying infinite collections and limits that were previously difficult or out of reach.

We shall begin with a variation on the material in Section **III.3,** describing unions and intersections of ***indexed families*** of sets; a typical example of this sort is a sequence of sets $\mathbf{A_n}$**,** where **n** runs through all positive integers. In the second section we define a notion of (possibly infinite) ***Cartesian product for*** such ***indexed families.*** This definition has some aspects that may seem unmotivated, and therefore we shall also describe an axiomatic approach to products such that **(*i*)** there is essentially only one set – theoretic construction satisfying the axioms **(*ii*)** the construction in these notes satisfies the axioms. In the next two sections we shall present Canto's landmark results on comparing infinite sets, including proofs of the following:

1. There is a $\mathbf{1 - 1}$ correspondence between the nonnegative integers $\mathbf{N}$ and the integers $\mathbf{Z}$**.**
2. There is a $\mathbf{1 - 1}$ correspondence between the nonnegative integers $\mathbf{N}$ and the rational numbers $\mathbf{Q}$**.**
3. There is **NO** $\mathbf{1 - 1}$ correspondence between the nonnegative integers $\mathbf{N}$ and the real numbers $\mathbf{R}$**.**

We should note that a few aspects of Cantor's discoveries (in particular, the first of the displayed statements) had been anticipated by Galileo.

Section 5 is a commentary on the impact of set theory, and Section 6 looks at generalizations of finite induction and recursion for sets that are larger than the natural numbers $\mathbf{N}$**.** The latter is included mainly as background for the sake of completeness.

VI.1 : Indexed families and set – theoretic operations

(Halmos, §§ 4, 8 – 9; Lipschutz, §§ 5.3 – 5.4)

One can summarize this section very quickly as follows: In Unit **III** we introduced several ways of constructing a third set out of two given ones, and in this section we

shall describe similar ways of constructing a new set out of a more or less arbitrary list of other ones.

We have frequently considered finite and infinite sequences of sets having the form $\mathbf{A_n}$ where the indexing subscript **n** runs through some finite or infinite set **S** of nonnegative integers. Formally such a sequence of sets corresponds to a function for which the value at a given integer **n** in **S** is equal to $\mathbf{A_n}$. We can generalize this as follows:

Definition. Let **I** be a set. An ***indexed family of sets with indexing set*** **I** is a function from **I** to some other set **X;** very often **X** is the set **P(Y)** of subsets of some other set **Y.** Such an indexed family is usually described by notation such as $\mathbf{\{ A_i \}_{i \in I}}$. In such cases **I** is generally called the ***index set***, while $\mathbf{I(i) = A_i}$ is the ***mapping*** or (Halmos' terminology) ***family,*** and $\mathbf{A_i}$ is the ***element belonging to the index value*** **i,** which is sometimes also called the $\mathbf{i^{th}}$ ***element*** or ***term of the indexed family.***

Given any sort of mathematical objects (***e.g.,*** partially ordered sets), one can define an indexed family of such objects similarly.

As indicated on page 34 of Halmos, in mathematical writings the notation for an indexed family is often abbreviated to $\mathbf{\{A_i\},}$ and this is described by the phrase, **"**unacceptable but generally accepted way of communicating the notation and indicating the emphasis**."** A more concise description would be a "slight abuse of language." Such an abbreviation should only be used if the indexing set it obvious from the context (for example, a subscript of **n** almost always denotes an integer) or its precise nature is relatively unimportant and there is no significant danger that the notation will be misinterpreted.

Subfamilies. An indexed family $\mathbf{\{B_i\}_{i \in J}}$ is a ***subfamily*** of a family of $\mathbf{\{ A_i \}_{i \in I}}$, if and only if **J** is a subset of **I** and for all **i** in **J** we have $\mathbf{B_i = A_i}$.

Indexed unions and intersections

Given a set **C,** in Unit **III** we considered the union **$(C),** which is the collection of all **x** such that $\mathbf{x \in A}$ for some $\mathbf{A \in C}$, and we introduced the usual ways of writing these sets as $\mathbf{\cup \{A \mid A \in C\}}$ or $\mathbf{\cup_{A \in C} A}$. If we have an indexed family of sets $\mathbf{\{ A_i \}_{i \in I}}$, then the ***indexed union***

$$\bigcup_{i \in I} A_i,$$

will refer to the union of the collection $\mathbf{\{ B \mid B = A_i}$ for some $\mathbf{i \in I}\}$. Recall that here **I** is a set, and $\mathbf{A_i}$ is a set for every $\mathbf{i \in I}$. In the case that the index set **I** is the set of natural numbers, one also uses notation is analogous to that of infinite series:

$$\bigcup_{i=1}^{\infty} A_i.$$

Similarly, given a ***nonempty*** set **C** (recall the extra condition is important**!**)**,** in Unit **III** we considered the intersection of the sets in **C,** which is the set of all **x** such that **x ∈ A** for every **A ∈ C,** and we similarly introduced the analogous ways of writing these sets as $\cap${**A | A ∈ C**} or $\cap_{A \in C}$ **A.** If we have an indexed family of sets $\{A_i\}_{i \in I}$**,** then we also have the corresponding indexed intersection

$$\bigcap_{i \in I} A_i.$$

As one might expect, this will be the intersection of the collection **{ B | B =** A_i for some **i ∈ I** }. As before, in the case that the index set **I** is the set of natural numbers, one also uses notation is analogous to that infinite series:

$$\bigcap_{i \in I} A_i.$$

These indexed unions and intersections satisfy analogs of the basic formal properties of ordinary unions and intersections which are stated formally on pages 35 – 36 of Halmos.

Numerous properties of unions and intersections of indexed families are developed in the exercises

VI.2 : Infinite Cartesian products

(Halmos, § 9; Lipschutz, §§ 5.4, 9.2)

We have already considered **n – *fold Cartesian products*** of **n** sets $X_1, \ldots, X_n$**:**

$$X_1 \times \ldots \times X_n = \{ (x_1, \ldots, x_n) \mid x_1 \in X_1 \ \& \ \ldots \ \& \ x_n \in X_n \}$$

At least intuitively, this construction can be identified with $(X_1 \times \ldots \times X_{n-1}) \times X_n$**.** We shall not attempt to make this precise here because one can easily do so using the discussion below for general Cartesian products.

Infinite products. For the most common mathematical applications, finite products suffice. However, for some purposes — in particular, many graduate courses in mathematics — it is necessary to define the general Cartesian product over an arbitrary (possibly infinite) collection of sets. Typical examples of this sort arise in the study of infinite sequences.

Definition. Let **I** be an arbitrary index set, and let $\{X_i \mid i \in I\}$ be a collection of sets indexed by **I.** The ***general Cartesian product*** of the indexed family $\{X_i \mid i \in I\}$ is denoted by symbolism such as

$$\prod \{X_i \mid i \in I\} \quad \text{or} \quad \prod_{i \in I} X_i$$

and is formally specified as follows:

$$\prod_{i \in I} X_i = \{f : I \to \bigcup_{i \in I} X_i \mid (\forall i)(f(i) \in X_i)\}$$

In other words, the general product is the set of all functions defined on the index set **I** such that the value of the function at a particular index **i** is an element of X_i. Since functions are determined by their values at the points of their domains, it follows that the element **f** in the general Cartesian product is completely determined by the indexed family of elements $f(i) \in X_i$. In a sense to be made precise later in this section, these elements $x_i = f(i)$ generalize the coordinates of an ordered pair **(x, y)** in the usual Cartesian product of two sets.

We have already noted that the collection of functions from one set to another is always a set, and this yields the corresponding result for general Cartesian products.

Proposition 1. *Let* **I** *be an arbitrary index set, and let* $\{X_i \mid i \in I\}$ *be a family of sets indexed by* **I.** *Then the general Cartesian product of the indexed family* $\{X_i \mid i \in I\}$ *is also a set.*

Proof. As noted in the paragraph preceding the statement of the proposition, the collection of all functions from the set **I** to the union $X = \cup \{X_i \mid i \in I\}$ is a set. By definition, the general Cartesian product is contained in this set, and therefore it is also a set.■

An **n –** tuple can be viewed as a function on **{1, 2, ... , n}** that takes its value at **i** to be the i^{th} element of the **n –** tuple. Hence, when **I** is **{1, 2, ... , n}** this definition coincides with the definition for the finite case.

One particular and familiar infinite case arises when the index set is the set **N** of natural numbers; this is just the set of all infinite sequences with the i^{th} term in its corresponding set X_i. An even more specialized case occurs when all the factors X_i involved in the product are the same, in which case the construction has an interpretation as **"**Cartesian exponentiation.**"** Then the big union in the definition is just the set itself, and the other condition is trivially satisfied, so this is just the set of ***all*** functions from **I** to **X,** which is the object we have previously called X^I.

In the ordinary Cartesian product of two sets, an element is completely specified by its coordinates, and the same is true for our general definition.

Proposition 2. *Let* **I** *be an arbitrary index set, and let* $\{X_i \mid i \in I\}$ *be a collection of sets indexed by* **I,** *and let* **x** *and* **y** *be elements of the Cartesian product of the indexed family* $\{X_i \mid i \in I\}$. *Then* $x = y$ *if and only if* $x_i = y_i$ *for all* **i.**

This follows immediately from the definition of the elements of the Cartesian product as functions defined on the indexing set.■

Formal characterizations of large products

For many purposes it is more convenient to look at large Cartesian products in terms of their functional behavior rather than their set – theoretic construction. In effect, this amounts to giving an axiomatic characterization of such products; from this viewpoint the main point of the previous construction is that it establishes the existence of an object which satisfies the axioms.

Definition. Let $\{ \mathbf{X}_j \}$ be an indexed family of sets with indexing set **J.** An ***abstract direct product*** of the indexed family $\{ \mathbf{X}_j \}$ is pair $\mathbf{(P, \{ p_j \})}$, where **P** is a set and $\mathbf{\{ p_j \}}$ is an indexed family of functions from $\mathbf{p_j : P \to X_j}$ such that the following ***Universal Mapping Property*** holds:

> **[UMP]** *Given an arbitrary set* **Y** *and functions* $\mathbf{f_j : Y \to X_j}$ *for each* **j,** *there is a* ***unique*** *function* $\mathbf{f : Y \to P}$ *such that* $\mathbf{p_j f = f_j}$ *for each* **j.**

Footnote. Such characterizations of mathematical constructions by universal mapping properties are fundamental to a topic in the foundations of mathematics known as ***category theory*****,** which was developed by S. Eilenberg (1913 – 1998) and S. MacLane (1909 – 2005)**.** This subject may be described as an abstract study of functions in mathematics, and among other things it can be used as alternative to set theory for constructing the logical foundations of mathematics (compare the comments at the beginning of Section **IV.3**). We shall not formally discuss the history, motivations and applications of category theory in these notes, but we shall give some online references for such topics. The first reference is a general discussion, the next few give some information about R. Carnap (1891 – 1970), a philosopher whose term ***functor*** was adopted to describe a fundamental concept of category theory, and the final reference is a summary of the main ideas from a slightly more advanced viewpoint.

http://plato.stanford.edu/entries/category-theory/

http://www.iep.utm.edu/c/carnap.htm

http://en.wikipedia.org/wiki/Rudolf_Carnap

http://www.rbjones.com/rbjpub/philos/history/rcp000.htm

http://math.ucr.edu/~res/math205A/categories.pdf

Universal mapping properties like **[UMP]** generally turn out to characterize mathematical constructions uniquely up to a suitably defined notion of equivalence. For our abstract definition of direct products, here is a formal statement of the appropriate uniqueness result.

Theorem 3. (Uniqueness of Direct Products). *Let* $\mathbf{\{ X_j \}}$ *be an indexed family of sets with indexing set* ***J****, and suppose that* $\mathbf{(P, \{ p_j \})}$ *and* $\mathbf{(Q, \{ q_j \})}$ *are direct products of the indexed family* $\mathbf{\{ X_j \}}$**.** *Then there is a* ***unique*** $\mathbf{1 - 1}$ *correspondence* $\mathbf{h : Q \to P}$ *such that* $\mathbf{p_j h = q_j}$ *for all* **j.**

Proof. **(**)** First of all, we claim that a function $\mathbf{T : P \to P}$ is the identity if and only if $\mathbf{p_j T = p_j}$ for all **j,** and likewise $\mathbf{S : Q \to Q}$ the the identity if and only if $\mathbf{q_j S = q_j}$ for all **j.** These are immediate consequences of the Universal Mapping Property, for in the first case we have $\mathbf{p_j T = p_j 1_X = p_j}$ for all **j,** and in the second we have the corresponding equations $\mathbf{q_j S = q_j 1_Y = q_j}$ for all **j.**

Since **(P,** { $\mathbf{p_j}$ } **)** is a direct product, the Universal Mapping Property implies there is a unique function $\mathbf{h : Q \to P}$ such that $\mathbf{p_j\, h = q_j}$ for all **j,** and likewise since **(Q,** { $\mathbf{q_j}$ } **)** is a direct product, there also exists a unique function $\mathbf{k : P \to Q}$ such that $\mathbf{q_j\, k = p_j}$ for all **j.** We claim that **h** and **k** are inverse to each other; this is equivalent to the pair of identities $\mathbf{h\, k = 1_Q}$ and $\mathbf{k\, h = 1_P}$**.**

To verify these identities, first note that for all **j** we have

$$\mathbf{p_j\, 1_X = p_j = q_j\, k = p_j\, h\, k}$$

for all **j** and similarly

$$\mathbf{q_j\, 1_Y = q_j = p_j\, h = q_j\, k\, h}$$

for all **j.** By the observations in the first paragraph of the proof, it follows that $\mathbf{k\, h = 1_P}$ and $\mathbf{h\, k = 1_Q}$**.**■

We now need to show that the axiomatic description of direct products is valid for the product construction described above. However, before doing so we verify that the ordinary Cartesian product of two sets also satisfies this propery.

Proposition 4. *If* **A** *and* **B** *are sets and* $\mathbf{p_A}$ *and* $\mathbf{p_B}$ *denote the standard coordinate projections from* $\mathbf{A \times B}$ *to* **A** *and* **B** respectively*, then* $\mathbf{(A \times B;\ p_A,\ p_B)}$ *is a direct product in the sense described above.*

Proof. We need to verify the Universal Mapping Property. Suppose that $\mathbf{f : C \to A}$ and $\mathbf{g : C \to B}$ are functions. Then we may define a function $\mathbf{H : C \to A \times B}$ by the formula $\mathbf{H(c) = (\, f(c),\, g(c)\,),}$ and by construction this function satisfies $\mathbf{p_A H = f}$ and $\mathbf{p_B H = g.}$ To conclude the proof we need to prove there is a unique function of this type, so assume that $\mathbf{K : C \to A \times B}$ also satisfies $\mathbf{p_A K = f}$ and $\mathbf{p_B K = g.}$ Now write $\mathbf{K(c) = (a,\, b),}$ and note that $\mathbf{a = p_A K(c) = f(c)}$ and $\mathbf{b = p_B K(c) = g(c).}$ Thus we have $\mathbf{K(c) = (\, f(c),\, g(c)\,) = H(c).}$ Since **c** was arbitrary it follows that **H = K.**■

Theorem 5. *Let* { $\mathbf{X_j}$ } *be an indexed family of sets with indexing set* **J,** *let*

$$\mathbf{\prod \{X_j \mid j \in J\}} = \mathbf{\prod_{j \in J} X_j}$$

be the generalized Cartesian product defined above, and for $\mathbf{k \in J}$ *let*

$$\mathbf{p_k : \prod \{X_j \mid j \in J\} \to X_k}$$

be the coordinate projection map such that $\mathbf{p_k(\, f\,) = f(k)}$ *for all* **k.** *Then the system*

$$\mathbf{\left(\prod_{j \in J} X_j,\ \{\, p_j \,\}\right)}$$

Is a direct product of the indexed family { $\mathbf{X_j}$ }**.**

The following **"**associativity property**"** of the ordinary Cartesian product will be useful in the proof of the theorem.

Lemma 6. Let **A, B** *and* **C** *be sets. Then there is a canonical* $1-1$ *correspondence* **T** *from* **(A × B) × C** *to* **A × (B × C)** *defined by the formula*

$$T((a, b), c) = (a, (b, c))$$

for all $a \in A$, $b \in B$, *and* $c \in C$.

***Proof of Lemma* 6. (**)** The formula for **T** is given in the lemma; we need to show this map is $1-1$ and onto. To see that it is $1-1$, suppose that

$$T((a, b), c) = T((x, y), z).$$

By construction this means that $(a, (b, c)) = (x, (y, z))$. Since ordered pairs are equal if and only if their respective coordinates are equal, it follows that we have $a = x$ and $(b, c) = (y, z)$. The second equation then implies $b = y$ and $c = z$, and from these we conclude that $((a, b), c) = ((x, y), z)$. Therefore the mapping **T** is $1-1$. To see that it is onto, note that every element of the codomain has the form $(a, (b, c))$ for suitable choices of **a, b** and **c,** and by the definition of **T** each such element belongs to the image of **T.**■

***Proof of Theorem* 5. (***)** All we need to do is verify the Universal Mapping Property. Suppose that we are given functions $f_j : Y \to X_j$ for each **j.**

For each **j** let G_j denote the subset of all (j, y, x) in $\{ j \} \times (Y \times X_j)$ such that (y, x) lies in the graph of f_j. Denote the union $\cup_j X_j$ of all the sets X_j by **X,** and let $G \subset J \times (Y \times X)$ be the union $\cup_j G_j$. Let $G' \subset (J \times Y) \times X$ denote the image of the set **G** under the associativity map in the lemma. **CLAIM:** For each (j, y) there is a unique **x** such that the object $((j, y), x)$ belongs to G'. This follows immediately from the fact that each f_j is a function.

Consider now the $1-1$ correspondence

$$J \times (Y \times X) \to J \times (X \times Y) \to (J \times X) \times Y \to Y \times (J \times X)$$

which takes $((j, y), x)$ to $((y, j), x)$. The middle step of this map is the associativity map in the lemma, and the outside steps merely transpose the coordinates in the appropriate ordered pairs. Let G^* denote the image of **G** under this mapping, and for each y in Y let G_y^* denote the intersection of G^* with the set $\{ y \} \times (J \times X)$. By the final two sentences of the preceding paragraph, it follows that G_y^* is the graph of a function H_y from **J** to **X,** and in fact the assumption on the functions f_j imply that H_y is the graph of a function such that $H_y(j)$ belongs to f_j for each **j.** The definition of the general Cartesian product then implies that H_y defines an element of the product $\prod\{X_j \mid j \in J\}$. By construction we have $H_y(j) = f_j(y)$, and this verifies the projection identities for the function we have constructed, proving the existence of a function from **Y** into the general Cartesian product with the required properties.

We now need to prove uniqueness. Suppose that **H** and **K** are functions from **Y** into the product which satisfy the basic projection identities. The latter imply that $H_y(j) = f_j(y)$ and $K_y(j) = f_j(y)$ for all **j** and **y.** But the latter equations mean that **H** and **K** define the

same functions from **J** to **X** for each **y,** so that $H_y = K_y$ for all **y,** which in turn implies that **H = K.**■

Technical note. Our definition of function differs from that of Halmos (we are including the codomain as part of the structure)**.** Because of this, the first sentence in the exercise on page 37 of Halmos must be modified to as follows in order to match our formulation: Instead of saying that the sets in question are equal we need to say that there is a $1-1$ correspondence between them. More precisely, if **J** is an index set, with $\{X_j \mid j \in J\}$ a collection of sets indexed by **J** and for each $j \in J$ we are given a subset A_j of X_j**,** then according to Halmos' definition we know that

$$\prod\{A_j \mid j \in J\} \quad \textit{is a subset of} \quad \prod\{X_j \mid j \in J\}$$

but in our formulation one only has the following weaker statement, which is completely adequate for all practical purposes:

Proposition 7. *In the setting above, let* e_j *denote the inclusion mapping from* A_j *to* X_I**.** *Then there is a* ***unique canonical*** **1 – 1** ***mapping***

$$e : \prod\{A_j \mid j \in J\} \rightarrow \prod\{X_j \mid j \in J\}$$

such that for each element **a** *of the domain and each indexing variable* **j** *we have the coordinate identity* $e(a)_j = e_j(a_j)$**.**

This mapping is often denoted by $\prod\{e_j \mid j \in J\}$ or more simply by $\prod e_j$**.**

Using the map **e** we may naturally identify the domain with the elements of the codomain such that for each **j,** the j^{th} coordinate lies in A_j**.**

***Proof.* (*)** Usually the fastest way of proving such a result is to apply the Universal Mapping Property, and doing so will also give us an opportunity to illustrate how the latter is used in mathematical work.

Let $\{p_j\}$ denote the family of coordinate projection maps for $\prod\{X_j \mid j \in J\}$**,** and similarly let $\{q_j\}$ denote the corresponding coordinate projection maps for the other product $\prod\{A_j \mid j \in J\}$**.** For each indexing variable **k,** define a mapping

$$f_k : \prod\{A_j \mid j \in J\} \rightarrow X_k$$

by setting f_k equal to the composite $e_k q_k$**.** The Universal Mapping Property then implies the existence of a unique function

$$e : \prod\{A_j \mid j \in J\} \rightarrow \prod\{X_j \mid j \in J\}$$

such that for each $j \in J$ we have $p_j e = e_j q_j$**.** This is equivalent to the condition on coordinates, so all that remains is to verify that **e** is a $1-1$ mapping**.** Since elements of a Cartesian product are determined by their coordinates, the latter reduces to showing that if **e(x) = e(y),** then for each $j \in J$ we have $x_j = y_j$**.** Let **J** be fixed but arbitrary, and consider the following string of equations which follows from **e(x) = e(y):**

$$e_j(x_j) = e(x)_j = e(y)_j = e_j(y_j)$$

Since the inclusion map $\mathbf{e_j}$ is $\mathbf{1-1}$ by construction, it follows that $\mathbf{x_j = y_j}$. Since **j** was arbitrary, this means that all the corresponding coordinates of **x** and **y** are equal and consequently that **x = y,** proving that **e** is also a $\mathbf{1-1}$ mapping.■

Applications of the Universal Mapping Property

We shall conclude this section with a few examples illustrating the use of the Universal Mapping Property for products to answer some basic questions. We shall begin with a version of the recursive property for finite Cartesian products mentioned at the beginning of this section.

Proposition 8. *Let* **A, B, C** *be sets. Denote the projections from* $\mathbf{(A \times B) \times C}$ *to* $\mathbf{A \times B}$ *and* **C** *by* $\mathbf{p_{1,2}}$ *and* $\mathbf{p_3}$ *respectively, and for* $\mathbf{i = 1}$ *or* **2** *let* $\mathbf{p_i}$ *denote the projection of* $\mathbf{A \times B}$ *to* **A** *and* **B** *respectively. Define mappings* $\mathbf{q_i}$ *by* $\mathbf{q_i = p_i p_{1,2}}$ *for* $\mathbf{i = 1}$ *or* **2,** *and* $\mathbf{q_3 = p_3}$. *Then the system* $\mathbf{((A \times B) \times C, \{q_1, q_2, q_3\})}$ *satisfies the Universal Mapping Property for products.*

Proof. Suppose that $\mathbf{f_1 : D \to A}$, $\mathbf{f_2 : D \to B}$, $\mathbf{f_3 : D \to C}$ are functions. By the Universal Mapping Property for twofold products there is a unique function $\mathbf{f_{1,2} : D \to A \times B}$ such that $\mathbf{p_i f_{1,2} = f_i}$ for $\mathbf{i = 1, 2.}$ Similarly, there is a unique function $\mathbf{f : D \to (A \times B) \times C}$ such that $\mathbf{p_{1,2} f = f_{1,2}}$ and $\mathbf{p_3 f = f_3}$. Since $\mathbf{q_3 = p_3}$, clearly $\mathbf{q_3 f = f_3}$. Furthermore, if $\mathbf{i = 1, 2}$ then $\mathbf{q_i f = p_i p_{1,2} f = p_i f_{1,2} = f_i}$, proving the existence part of the Universal Mapping Property.

To prove uniqueness, suppose that the projections of $\mathbf{h, k : B \to (A \times B) \times C}$ onto the sets **A, B, C** are equal to the mappings $\mathbf{f_i}$. We first claim that the projections of **h** and **k** onto $\mathbf{A \times B}$ are equal. The projections of **h** and **k** onto $\mathbf{A \times B}$ satisfy $\mathbf{q_i h = f_i = q_i k}$ for $\mathbf{i = 1}$ or **2,** and thus by the Universal Mapping Property for twofold products it follows that $\mathbf{p_{1,2} h = p_{1,2} k.}$

By assumption we also have $\mathbf{q_3 h = f_3 = q_3 k,}$ and hence by the Universal Mapping Property for the twofold product $\mathbf{(A \times B) \times C}$ it follows that $\mathbf{h = k.}$■

Here is another example, which is also a good illustration of proving that a mapping is bijective.

Proposition 9. *Let* **A, B, C** *be sets.*

(1) *There is a unique mapping* **T** *from* $\mathbf{(A \times B) \times C}$ *to* $\mathbf{(C \times A) \times B}$ *such that* $\mathbf{T(x, y, z) = (z, x, y)}$ *for all* **x, y, z.**

(2) *The mapping* **T** *is bijective, and if* $\mathbf{A = B = C}$ *the inverse is given by* $\mathbf{T \circ T.}$

Proof. By the Universal Mapping Property for products there is a unique mapping **T** from $\mathbf{(A \times B) \times C}$ to $\mathbf{(C \times A) \times B}$ such that $\mathbf{p_1 T = p_3}$, $\mathbf{p_2 T = p_1}$, and $\mathbf{p_3 T = p_2}$. By construction, such a map satisfies $\mathbf{T(x, y, z) = (z, x, y)}$ for all **x, y, z.**

We first show that **T** is injective. If $\mathbf{T(x, y, z) = T(x', y', z'),}$ then by definition of **T** we have $\mathbf{(z, x, y) = (z', x', y')}$ and the latter implies $\mathbf{x = x', y = y',}$ and $\mathbf{z = z'.}$ Next we prove

that **T** is surjective. To solve the equation **T(x, y, z) = (u, v, w)** we need to find **(x, y, z)** so that **(z, x, y) = (u, v, w).** Clearly **x = v, y = w, z = u** gives a solution, so that map is surjective as claimed.

If we have **A = B = C** then $\mathbf{T^{-1}(u, v, w) = (x, y, z)}$ implies **(z, x, y) = (u, v, w),** so that $\mathbf{T^{-1}(u, v, w) = (v, w, u).}$ But the latter is equal to **T(w, u, v) = T ∘ T(u, v, w),** and therefore $\mathbf{T^{-1} = T \circ T}$ as required**.**■

V.3 : Transfinite cardinal numbers

(Halmos, §§ 22 – 23; Lipschutz, §§ 6.1 – 6.3, 6.5)

Early in his work on infinite sets, Cantor considered the problem of comparing the relative sizes of such sets. Specifically, given two infinite sets, the goal is to determine if one has the same size as the other or if there are different orders of infinity such that one set is of a lower order than the other. Many of Cantor's results were entirely unanticipated, and ultimately his findings led mathematicians to make major changes to their perspectives on infinite objects. In several respects the material in this section is the central part of these notes.

Definition. If **A** and **B** are sets, we write **|A| = |B|,** and say that *the cardinality of* **A** *is equal to the cardinality of* **B** (or they have the same cardinality, *etc*.) if there is a $\mathbf{1-1}$ onto mapping **f : A → B .**

The relationship **|A| = |B|** is clearly reflexive because the identity on **A** is a $\mathbf{1-1}$ onto map from **A** to itself, and if **|A| = |B|,** then **|B| = |A|** is also true because the inverse of **f** is a $\mathbf{1-1}$ onto mapping from **B** to **A.** Finally, if **|A| = |B|** and **|B| = |C|,** then we also have **|A| = |C|,** for if we have $\mathbf{1-1}$ onto mappings **f : A → B** and **g : B → C** , then the composite **g f** is a $\mathbf{1-1}$ onto mapping from **A** to **C.** In particular, if **X** is a set and we define a binary relation of **"**having the same cardinality**"** on **P(X)** to mean that **|A| = |B|,** then having the same cardinality defines an equivalence relation on **P(X).** In such a setting, the ***cardinal number*** of a subset **A** may be interpreted as the equivalence class of all sets **B** which have the same cardinality as **A.** This relation is actually independent of the choice of set **X** containing **A** and **B,** for if **Y c**ontains **X** then **A** and **B** determine the same equivalence class in **P(X)** if and only if they determine the same equivalence class in **P(Y).**

The restriction to subsets of a given set is awkward, but some restrictive condition is needed and we have chosen one that is relatively simple to state. Initially, many mathematicians and logicians including Cantor, B. Russell and G. Frege (1848 – 1925), attempted to define the cardinal number of a set **X** as the equivalence class of all sets **Y** that can be put into a $\mathbf{1-1}$ correspondence with **X,** but a definition of this type cannot

be made logically rigorous because of the family of all such objects is "too large" to be a set.

Finite and infinite sets

For finite sets, the notion of cardinality has been understood for thousands of years.

Definition. If **n** is a positive integer, then a nonempty set **X** has ***cardinal number equal to*** **n** if there is a $\mathbf{1-1}$ correspondence between **X** and $\mathbf{\{\ 0, \ldots, n-1\ \}}$. By the results of Section **V.3,** it follows that there is at most one **n** such that a set has cardinal number equal to **n.** The definition is extended to nonnegative integers by taking the cardinality of the empty set to be $\mathbf{0}$. We say that a set **X** is ***finite*** if it has cardinal number equal to **n** for some **n** and that **X** is ***infinite*** otherwise.

Cantor's important — and in fact revolutionary — insight was that one can define ***transfinite cardinal numbers*** to measure the relative sizes of infinite sets.

Partial ordering of cardinalities

Definition. If **A** and **B** are sets, we write $\mathbf{|A| \leq |B|}$, and say that *the cardinality of* **A** *is less than or equal to the cardinality of* **B** if there is a $\mathbf{1-1}$ map from **A** to **B.**

The notation suggests that this relationship should behave like a partial ordering (in analogy with finite sets we would like it to be a linear ordering, but reasons for being more modest in the infinite case will be discussed later). It follows immediately that the relation we have defined is ***reflexive*** (take the identity map on a set **A**) and ***transitive*** (given $\mathbf{1-1}$ maps $\mathbf{f : A \to B}$ and $\mathbf{g : B \to C}$, the composite $\mathbf{g\,f}$ is also $\mathbf{1-1}$), but the proof that it is ***antisymmetric*** is decidedly nontrivial:

Theorem 1. (Schröder – Bernstein Theorem.) *If* **A** *and* **B** *are sets such that there are* $\mathbf{1-1}$ *maps* $\mathbf{A \to B}$ *and* $\mathbf{B \to A}$, *then* $\mathbf{|A| = |B|}$.

***Proof.* (**)** We shall give the classic argument from the (third edition of the) book by G. [= Garrett] Birkhoff (1911 – 1996) and S. MacLane (1909 – 2005) cited below; the precise reference is page 340.

> G. Birkhoff and S. MacLane, ***A Survey of Modern Algebra.*** (Reprint of the Third 1968 Edition). Chelsea Publishing, New York, NY, 1988. ISBN: 0 – 023 – 74310 – 7.

Let $\mathbf{f : A \to B}$ and $\mathbf{g : B \to A}$ be $\mathbf{1-1}$ mappings which exist by the assumptions. Each $\mathbf{a \in A}$ is the image of at most one parent element $\mathbf{b \in B}$; in turn, the latter (if it exists) has at most one parent element in **A,** and so on. The idea is to trace back the ancestry of each element as far as possible. For each point in **A** or **B** there are exactly three possibilities:

1. The ancestral chain may go back forever.
2. The ancestral chain may end in **A.**
3. The ancestral chain may end in **B.**

We can then split **A** and **B** into three pairwise disjoint pieces corresponding to these cases, and we shall call the pieces $\mathbf{A_1}$, $\mathbf{A_2}$, $\mathbf{A_3}$ and $\mathbf{B_1}$, $\mathbf{B_2}$, $\mathbf{B_3}$ (where the possibilities are ordered as in the list).

The map **f** defines a $\mathbf{1-1}$ correspondence between $\mathbf{A_1}$ and $\mathbf{B_1}$ (and likewise for **g**). Furthermore, **g** defines a $\mathbf{1-1}$ correspondence from $\mathbf{B_2}$ to $\mathbf{A_2}$, and **f** defines a $\mathbf{1-1}$ correspondence from $\mathbf{A_3}$ to $\mathbf{B_3}$. If we combine these $\mathbf{1-1}$ correspondences $\mathbf{A_1 \leftrightarrow B_1}$, $\mathbf{A_2 \leftrightarrow B_2}$, and $\mathbf{A_3 \leftrightarrow B_3}$, we get a $\mathbf{1-1}$ correspondence between all of **A** and all of **B.**■

Here is an immediate consequence of the Schröder – Bernstein Theorem:

Proposition 2. *If* **A** *is an infinite subset of the nonnegative integers* $\mathbf{N}$, *then* $\mathbf{|A| = |N|}$.

Proof. **(∗)** We shall define a $\mathbf{1-1}$ mapping from $\mathbf{N}$ to **A** recursively; the existence of such a map will imply $\mathbf{|A| \leq |N|}$; by hypothesis we have the reverse inequality $\mathbf{|N| \leq |A|}$, and therefore the Schröder – Bernstein Theorem implies that $\mathbf{|A| = |N|}$.

Since $\mathbf{N}$ is well – ordered, it follows that every nonempty subset of **A** has a least element. Define **f** recursively by setting $\mathbf{f(0)}$ equal to the least element of **A,** and if we are given a partial $\mathbf{1-1}$ function $\mathbf{g_n: \{0, \ldots, n-1\} \to A}$, extend the definition to the set $\mathbf{\{0, \ldots, n\}}$ by noting that the image of $\mathbf{g_n}$ is a proper subset of **A** (which is infinite) and taking $\mathbf{g_{n+1}(n)}$ to be the ***first*** element in $\mathbf{A - \mathit{Image}(g_n)}$. The increasing union of these functions will be the required function from $\mathbf{N}$ to **A.** It is $\mathbf{1-1}$ because it is $\mathbf{1-1}$ on each subset $\mathbf{\{0, \ldots, n-1\}}$; if $\mathbf{f(x) = f(y)}$, then there is some **n** such that **x** and **y** both belong to $\mathbf{\{0, \ldots, n-1\}}$, and therefore it follows that **x** and **y** must be equal.■

Definition. A set is ***countable*** if it is in $\mathbf{1-1}$ correspondence with a subset of the natural numbers, and it is ***denumerable*** if it is in $\mathbf{1-1}$ correspondence with the natural numbers. However, many writers also use countable as a synonym for denumerable, so one must be careful. Frequently one also sees the phrase ***"countably infinite"*** employed as a synonym for denumerable.

Following Cantor, it is customary to denote the cardinal number of the natural numbers by $\mathbf{\aleph_0}$ (verbalized as ***aleph – null***).

The next result generalizes a simple fact about cardinal numbers from finite sets to countable sets.

Proposition 3. *Suppose that* **A** *is a nonempty countable set and there is a surjective mapping* **f** *from* **A** *to* **B.** *Then* **B** *is also countable, and in fact* $\mathbf{|B| \leq |A|}$.

Proof. By hypothesis there is a $\mathbf{1-1}$ correspondence between **A** and a subset of the nonnegative integers **N,** and thus one can use the standard ordering of the latter to make **A** into a well – ordered set. Define a function $\mathbf{h : B \to A}$ as follows; given a typical $\mathbf{b \in B,}$ take **h(b)** to be the least element in the inverse image $\mathbf{f^{-1}[\{b\}].}$ Then by definition we have $\mathbf{f\,h(b) = b.}$ The result will follow from the previous proposition if we can show that **h** is a $\mathbf{1-1}$ mapping, and the latter holds because **h(x) = h(y)** implies $\mathbf{x = f\,h(x) = f\,h(y) = y.}$ ■

VI.3 : Countable and uncountable sets

(Halmos, §§ 23 – 23; Lipschutz, §§ 6.3 – 6.7)

A theory of transfinite cardinal numbers might not be particularly useful if all infinite sets had the same cardinality. In the first paragraphs of this unit we indicated that the cardinalities of $\mathbf{R}$ and $\mathbf{N}$ are different, and the goal of this section is to prove this result. The first step in this process is to extend some basic arithmetic operations on $\mathbf{N}$ to arbitrary transfinite cardinal numbers.

Binary operations on cardinal numbers

One can perform a limited number of arithmetic operations with cardinal numbers, but it is necessary to realize that these do not enjoy all the familiar properties of the corresponding operations on positive integers. Before doing so, it is convenient to introduce a set – theoretic construction which associates to two sets **A** and **B** a third set which is a union of disjoint isomorphic copies of **A** and **B.** Formally, the ***disjoint sum*** (or ***disjoint union***) is defined to be the set

$$\mathbf{A \amalg B = A \times \{1\} \cup B \times \{2\}}$$

and the standard ***injection*** mappings $\mathbf{i_A : A \to A \amalg B}$ and $\mathbf{i_B : B \to A \amalg B}$ are defined by

$$\mathbf{i_A(a) = (a, 1)} \quad \text{and} \quad \mathbf{i_B(b) = (b, 2)}$$

respectively. By construction, we have the following elementary consequences of the definition:

Proposition 1. *Suppose that we are given the setting and constructions described above.*

- **(1)** *The maps* $\mathbf{i_A}$ *and* $\mathbf{i_B}$ *determine* $\mathbf{1-1}$ *correspondences* $\mathbf{j_A}$ *from* **A** *to* $\mathbf{i_A(A)}$ *and* $\mathbf{j_B}$ *from* **B** to $\mathbf{i_B(B).}$
- **(2)** *The images of* **A** *and* **B** *are disjoint.*
- **(3)** *The union of the images of* **A** *and* **B** *is all of* $\mathbf{A \amalg B}$ **.**

The proof of this result is fairly simple, but we shall include it for the sake of completeness and because it is not necessarily easy to locate in the literature.

***Proof of* (1).** The sets $\mathbf{i_A\ (A)}$ and $\mathbf{i_B\ (B)}$ are equal to $\mathbf{A \times \{1\}}$ and $\mathbf{B \times \{2\}}$ respectively, and we have $\mathbf{j_A\ (a) = (a, 1)}$ and $\mathbf{j_B\ (b) = (b, 2).}$ It follows that inverse maps are given by projection onto **A** and **B** respectively**.**■

***Proof of* (2).** The first coordinate of an element in the image of $\mathbf{i_A}$ is equal to **1,** and the first coordinate of an element in the image of $\mathbf{i_B}$ is equal to **2.** Therefore points in the image of one map cannot lie in the image of the other**.**■

***Proof of* (3).** Clearly the union is contained in $\mathbf{A \amalg B}$**.** Conversely, if we are given a point in the latter, then either it has the form $\mathbf{(a,}\ 1\mathbf{) = i_A\ (a)}$ or $\mathbf{(b,}\ 2\mathbf{) = i_B\ (b).}$■

Definition. (Addition of cardinal numbers). If **A** and **B** are sets with cardinal numbers **|A|** and **|B|** respectively**,** then the *sum* $\mathbf{|A| + |B|}$ is equal to $\mathbf{|\ A \amalg B\ |.}$

Definition. (Multiplication of cardinal numbers). If **A** and **B** are sets with cardinal numbers **|A|** and **|B|** respectively**,** then the *product* $\mathbf{|A| \times |B|}$ or equivalently $\mathbf{|A| \cdot |B|}$ (or sometimes even $\mathbf{|A|\ |B|}$) is equal to $\mathbf{|\ A \times B\ |.}$

Definition. (Exponentiation of cardinal numbers). If **A** and **B** are sets with cardinal numbers **|A|** and **|B|** respectively**,** then the *power operation* $\mathbf{|A|^{|B|}}$ is $\mathbf{|A^B|,}$ where $\mathbf{A^B}$ denotes the set of functions from **B** to **A** (as in Unit **IV**).

In order to justify these definitions we need to verify two things; namely, that **[** $\boldsymbol{i}$ **]** these definitions agree with the counting results Section **V.3** if **A** and **B** are finite sets, and also **[** $\boldsymbol{ii}$ **]** that the construction is ***well – defined***; we have defined the operations by choosing specific sets **A** and **B** with given cardinal numbers, and we need to make sure that if choose another pair of sets, say **C** and **D,** then we obtain the same cardinal numbers. The first point is easy to check; if **A** and **B** are finite sets, then the formulas in Section **V.3** show that the numbers of elements in $\mathbf{A \amalg B,}$ $\mathbf{A \times B,}$ and $\mathbf{A^B}$ are respectively equal to $\mathbf{|A| + |B|,}$ $\mathbf{|A| \cdot |B|}$ and $\mathbf{|A^B|.}$ The following elementary result disposes of the second issue.

Proposition 2. *Suppose that we are given sets* **A, B, C, D** *and we also have* $\mathbf{1 - 1}$ *correspondences* $\mathbf{f : A \to C}$ *and* $\mathbf{g : B \to D.}$ *Then there are* $\mathbf{1 - 1}$ *correspondences from* $\mathbf{A \amalg B,}$ $\mathbf{A \times B,}$ *and* $\mathbf{A^B}$ *to* $\mathbf{C \amalg D,}$ $\mathbf{C \times D,}$ *and* $\mathbf{C^D}$ *respectively.*

Proof. Define mappings

$$\mathbf{H : A \amalg B \to C \amalg D,} \quad \mathbf{J : A \times B \to C \times D,} \quad \mathbf{K : A^B \to C^D}$$

by the following formulas:

$$\mathbf{H(a, 1) = (\ f(a), 1\ ),} \quad \mathbf{H(b, 2) = (\ g(b), 2\ )}$$

$$\mathbf{J(a, b) = (\ f(a), g(b)\ )}$$

$$\mathbf{[\ K(\varphi)\ ]\ (c) = f \varphi g^{-1}(c)}$$

Define mappings in the opposite direction(s)

$$L : C \coprod D \rightarrow A \coprod B, \quad M : C \times D \rightarrow A \times B, \quad N : C^D \rightarrow A^B$$

by substituting f^{-1}, g^{-1}, and **g** for the variables **f, g,** and g^{-1} in the corresponding definitions of **H, J** and **K** respectively. Routine calculations (left to the reader) show that the maps **L, M** and **N** are inverses to the corresponding mappings **H, J** and **K.**■

We shall see that operations on transfinite cardinal numbers do not satisfy some of the fundamental properties that hold for integers; for example, we shall see below that an equation of the form $x + y = x$ does not necessarily imply that $x = 0$. However, here is one important relationship that does generalize:

Proposition 3. *If* **A** *is a set then* $|P(A)| = 2^{|A|}$.

Proof. We need to define a $1-1$ correspondence χ from **P(A)** to the set of functions from **A** to the set $\{0, 1\}$. Given a subset **B,** its ***characteristic function*** $\chi_B : A \rightarrow \{0, 1\}$ is defined by $\chi_B(x) = 1$ if $x \in B$ and 0 otherwise. The map sending a subset to its characteristic function is $1-1$ because $B = \chi_B^{-1}[\{1\}]$, so that $\chi_B = \chi_C$ implies $B = \chi_B^{-1}[\{1\}] = \chi_B^{-1}[\{1\}] = C$. To see this is onto, let $f : A \rightarrow \{0, 1\}$ and note that by definition we have $f = \chi_B$ where $B = f^{-1}[\{1\}]$.■

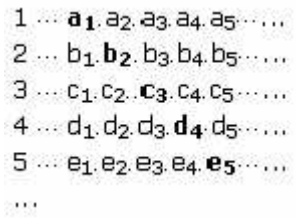
Finally, we have the following fundamentally important result due to Cantor.

Theorem 4. *If* **A** *is a set then* $|A| < |P(A)| = 2^{|A|}$.

Proof. **(*)** Define a $1-1$ mapping from **A** to **P(A)** sending an element $a \in A$ to the one point subset $\{a\}$. This shows that $|A| \leq |P(A)|$.

The proof that $|A| \neq |P(A)|$ is given by the ***Cantor diagonal process.*** Suppose that there is a $1-1$ correspondence $F : A \rightarrow \{0, 1\}^A$. The idea is to construct a new function $g \in \{0, 1\}^A$ that is not in the image of **F**. Specifically, choose **g** such that, for each $a \in A$, the value **g(a)** will be the unique element of $\{0, 1\}$ which is not equal to $[F(a)](a)$; recall that **F(a)** is also a function from **A** to $\{0, 1\}$ and as such it can be evaluated at **a.** Since the values of **g** and **F(a)** at $a \in A$ are different, these two functions are distinct, and since $a \in A$ is arbitrary it follows that **g** cannot lie in the image of **F.** However, we were assuming that **F** was onto, so this yields a contradiction. Therefore there cannot be a $1-1$ correspondence between **A** and **P(A)**.■

Comments on the method of proof. The reason for the name ***diagonal process*** is illustrated below when **A** is the set $\mathbf{N}^+$ of positive integers. One assumes the existence of a $1-1$ correspondence between $\mathbf{N}^+$ and $P(\mathbf{N}^+)$ and identifies the latter with the set of functions from $\mathbf{N}^+$ to $\{0, 1\}$ in the standard fashion. Then for each positive integer one has an associated sequence of 0's and 1's that are indexed by the positive integers, and one can represent them in a table or matrix form as illustrated below, in which each of the terms x_j (where **x** is a letter and **j** is a positive integer) is equal to either 0 or 1.



The existence of a $\mathbf{1-1}$ correspondence implies that all sequences appear on the list. However, if we change each of the bold entries (***i.e.,*** the entry in the $\mathbf{n^{th}}$ row and $\mathbf{n^{th}}$ column for each **n**) by taking $\mathbf{0}$ if the original entry is $\mathbf{1}$ and vice versa, we obtain a new sequence that is not already on the list, showing that $\mathbf{P(\mathbb{N}^{+})}$ cannot be put into correspondence with $\mathbf{\mathbb{N}^{+}}$ and thus represents a higher order of infinity**.**■

The preceding result implies that ***"there is no set of all cardinal numbers."*** Stated differently, there is no set **S** such that every set **A** is in 1 – 1 correspondence with a subset of **S.** If such a set existed, then the set **P(S)** would be in 1 – 1 correspondence with some subset $\mathbf{T \subset S,}$ and hence we would obtain the contradiction

$$\mathbf{|P(S)| = |T| \leq |S| < |P(S)|.}\blacksquare$$

This observation is often called ***Cantor's Paradox***, and was noted by Cantor in 1899; it is very close to the original set – theoretic paradox that was discovered by C. Burali – Forti (1861 – 1931) a few years earlier and will be discussed in the next section.

Some basic rules of cardinal arithmetic

Addition and multiplication of cardinal numbers satisfy many of the same basic equations and inequalities that hold for nonnegative integers. Here is a list of the most fundamental examples:

Theorem 5. *The sum and product operations on cardinal numbers have the following properties for all cardinal numbers* α **,** β *and* γ**:**

(Associative law of addition) $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$

(Commutative law of addition) $\alpha + \beta = \beta + \alpha$

(Associative law of multiplication) $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$

(Commutative law of multiplication) $\alpha \cdot \beta = \beta \cdot \alpha$

(Distributive law) $\alpha \cdot (\beta + \gamma) = (\alpha \cdot \beta) + (\alpha \cdot \gamma)$

(Equals added to unequals) $\alpha \leq \beta \Rightarrow \alpha + \gamma \leq \beta + \gamma$

(Equals multiplied by unequals) $\alpha \leq \beta \Rightarrow \alpha \cdot \gamma \leq \beta \cdot \gamma$

The verifications of all these equations and inequalities are extremely straightforward. For example, the commutative law of addition merely reflects the commutative law for set – theoretic unions, and the commutative law of multiplication reflects the existence of the canonical $\mathbf{1-1}$ correspondence from the Cartesian product $\mathbf{A \times B}$ to the analogous

product with interchanged factors $\mathbf{B \times A}$, which sends **(a, b)** to **(b, a).** All the details are worked out on page 161 of Lipschutz. These proofs do not use our formal definition for the sum of two cardinal numbers, but instead they use the following characterization:

Lemma 6. *If* **X** *and* **Y** *are disjoint sets, then* $|X \cup Y| = |X| + |Y|$. *Furthermore, if* **A** *and* **B** *are arbitrary sets, then there exist sets* **X** *and* **Y** *such that* $|X| = |A|$, $|Y| = |B|$, *and also* $X \cap Y = \varnothing$.

Proof. The second part of the lemma follows from our disjoint union construction. The first part will follow if there is a $1 - 1$ correspondence **H** from $X \sqcup Y$ to $X \cup Y$. An explicit construction of such a map is given by $H(x, 1) = x$ and $H(y, 2) = y$. Since the image of this map contains both **X** and **Y,** it follows that **H** is onto. To see it is $1 - 1$, note that the restrictions to $X \times \{1\}$ and $Y \times \{2\}$ are both $1 - 1$ so the only way the map might not be $1 - 1$ is if one has $x \in X$ and $y \in Y$ such that $H(x, 1) = H(y, 2)$. The latter would imply that **X** and **Y** are not disjoint, and since we know they are disjoint it follows that there are no such elements **x** and **y,** so that **H** must also be $1 - 1$ as required.■

Although arbitrary cardinal numbers satisfy many of the same basic equations and inequalities as nonnegative integers, it is important to recognize that some algebraic properties of the latter do not extend. In particular, the results below prove that a cardinal number equation of the form $\alpha + \beta = \alpha$ does ***not*** necessarily imply $\beta = 0$. Similarly, an equation of the form $\alpha \cdot \beta = \alpha$ does ***not*** necessarily imply that either $\beta = 1$ or $\alpha = 0$.

Identities and inequalities for cardinal numbers

The following simple result illustrates a major difference between finite and transfinite cardinals:

Proposition 7. *If* **A** *is finite, then* $|A| + \aleph_0 = \aleph_0$.

Proof. If $|A| = 0$ this is trivial. Suppose now that $|A| = 1$, and let **a** be the unique element of **A.** Let $\mathbf{N}$ be the natural numbers, and define a mapping **h** from $A \sqcup \mathbf{N}$ to $\mathbf{N}$ by setting $h(a, 1) = 0$ and $h(n, 2) = n + 1$ for $n \in \mathbf{N}$. By the Peano Axioms for the natural numbers, the restriction of **h** to $\mathbf{N} \times \{2\}$ is injective, and its image is the set of all positive integers. Since $h(a, 1) = 0$, it follows that **h** is $1 - 1$ and onto. Therefore we have $1 + \aleph_0 = \aleph_0$.

From this point on we proceed by induction on $k = |A|$. Suppose we know the result in this case; we need to prove it is also true for $|A| = k + 1$. This is an direct consequence of the following chain of equations:

$$(k + 1) + \aleph_0 = (1 + k) + \aleph_0 = 1 + (k + \aleph_0) = 1 + \aleph_0 = \aleph_0$$

This completes the proof of the inductive step and hence of the result itself.■

The following standard identities involving $\aleph_0$ were first noted by Galileo (thus is frequently known as ***Galileo's Paradox***) and Cantor respectively.

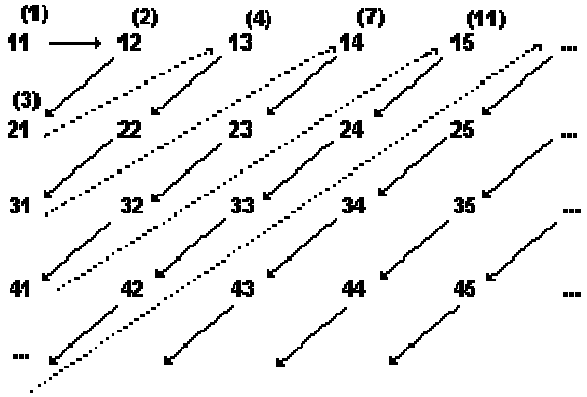
Theorem 8. (Idempotent Laws). *We have* $\aleph_0 + \aleph_0 = \aleph_0$ *and* $\aleph_0 \cdot \aleph_0 = \aleph_0$.

Proof. Let $\mathbf{N}$ be the nonnegative integers, and let $\mathbf{N(0)}$ and $\mathbf{N(1)}$ denote the subsets of even and odd nonnegative integers respectively. Then the mappings sending **n** to $\mathbf{2n}$ and $\mathbf{2n + 1}$ define $\mathbf{1 - 1}$ correspondences from **N** to **N(0)** and **N(1)** respectively. Since $\mathbf{N(0) \cup N(1) = N}$ and $\mathbf{N(0) \cap N(1) = \emptyset}$, it follows that

$$\aleph_0 = |\mathbf{N}| = |\mathbf{N(0)}| + |\mathbf{N(1)}| = |\mathbf{N}| + |\mathbf{N}| = \aleph_0 + \aleph_0$$

proving the first assertion in the theorem.

To prove the second assertion, we shall first define a $\mathbf{1 - 1}$ mapping from $\mathbf{N \times N}$ to $\mathbf{N}$ by defining an equivalent map from $\mathbf{N^+ \times N^+}$ to $\mathbf{N^+}$ by a diagonal construction due to Cantor (also see Halmos, page 92). The following picture illustrates the idea behind the definition of the function; the explicit formula is $\mathbf{f(m, n) = \frac{1}{2}(m + n - 1)(m + n - 2) + m}$.



(Source: http://www.cut-the-knot.org/do_you_know/numbers.shtml)

A verification that **f** is $\mathbf{1 - 1}$ is sketched in the exercises. We also have an easily defined $\mathbf{1 - 1}$ mapping in the opposite direction sending **n** to **(n, 0).** We can now use the Schröder – Bernstein Theorem to prove the equality $\mathbf{|N| = |N \times N|}$, or equivalently that $\aleph_0 \cdot \aleph_0 = \aleph_0$.■

Corollary 9. *For each positive integer* **n** *we have* $\mathbf{n} \cdot \aleph_0 = \aleph_0$ *and* $(\aleph_0)^n = \aleph_0$.

Proof. The main result proves the result for $\mathbf{n = 2}$, and it is trivial if $\mathbf{n = 1}$.

The proof that the special case $\mathbf{n = 2}$ implies the general case can be done abstractly as follows: Suppose that we are given any associative binary operation and an element **a** such that $\mathbf{a^2 = a}$. Under this condition we claim that $\mathbf{a^n = a}$ for all $\mathbf{n > 1}$. The case $\mathbf{n = 2}$ is given, so assume that the result is true for some $\mathbf{k > 1}$. Then we have

$$\mathbf{a^{k+1} = a^k a = a a = a}$$

completing the inductive step of the derivation. We have written the binary operation multiplicatively, but of course we also could have written it additively, and thus the whole argument works for both addition and multiplication of cardinal numbers.■

We now have the following standard consequences.

Proposition 10. *Let* **C** *be a countable family of sets, each of which is countable. Then the countable union of the countable sets* $\$(C) = \cup_{B \in C} B$ *is also countable.*

Proof. Let **A** be the set of all ordered pairs **(x, B)** such that $x \in B$ and $B \in C$. If we define $g : A \to \$(C)$ by projection onto the first coordinate, then **g** is onto. By Proposition 3, it will suffice to prove that **A** is countable. Let $f : C \to N$ be a $1 - 1$ mapping, and for each $B \in C$ define a 1 – 1 mapping $g_B : B \to N$. All these maps exist because **C** is countable and each subset **B** in **C** is countable. Next we define a mapping $h: A \to N \times N$ by $h(x, B) = (g_B(x), f(B))$. We claim that **h** is $1 - 1$. Suppose that we have $h(x, B) = h(y, D)$. By definition we then have $f(B) = f(D)$, and since **f** is $1 - 1$ it follows that $B = D$. Once again using the definitions we see that $g_B(x) = g_B(y)$, and since g_B is $1 - 1$ it follows that $x = y$. This completes the proof that **h** is $1 - 1$, which implies the key assertion that **A** is countable; as noted earlier in the discussion, this completes the proof.■

Proposition 11. *If* $\mathbf{Z}$ *and* $\mathbf{Q}$ *are the integers and rational numbers respectively, then we have* $|Z| = |Q| = \aleph_0$.

The result for the integers was anticipated in Galileo's writings on infinite sets, but the result regarding the rational numbers was something of a surprise to mathematicians when it was discovered by Cantor in the 1870s.

Proof. The standard inclusions $N \subset Z \subset Q$ imply a chain of corresponding inequalities $\aleph_0 = |N| \leq |Z| \leq |Q|$. Define a surjective mapping $N \amalg N \to Z$ sending **(n, 1)** to **n** and **(n, 2)** to **– n.** By Theorem 8 it follows that

$$|Z| \leq |N \amalg N| = \aleph_0 + \aleph_0 = \aleph_0,$$

so the result for $|Z|$ follows from the Schröder – Bernstein Theorem.

Next define a surjective mapping $Z \times (Z - \{0\}) \to |Q|$ sending **(a, b)** to **a/b.** We then have $|Q| \leq |Z \times (Z - \{0\})| \leq \aleph_0 \cdot \aleph_0 = \aleph_0$. Once again the Schröder – Bernstein Theorem implies that $|Q| = \aleph_0$.■

The next natural question concerns the cardinality of the set of the real numbers, and the result is again due to Cantor.

Theorem 12. *If* $\mathbf{R}$ *denotes the real numbers, then its cardinality satisfies* $|R| = 2^{\aleph_0}$ *and therefore we have* $|R| > \aleph_0$.

Proof. Usually this is derived using decimal expansions of real numbers, but we shall give a proof that does not involve decimals (although the idea is similar). The idea is to

construct $\mathbf{1-1}$ maps from $\mathbf{R}$ to $\mathbf{P(N)}$ and vice versa and then to apply the Schröder – Bernstein Theorem.

Let $\mathbf{D: R \to P(Q)}$ be the Dedekind cut map sending a real number **r** to the set of all rational numbers less than **r**. Since there is always a rational number between any two distinct real numbers, it follows that this map is $\mathbf{1-1}$. Since $\mathbf{|Q| = \aleph_0}$, it follows that there is a $\mathbf{1-1}$ correspondence from $\mathbf{P(Q)}$ to $\mathbf{P(N)}$, and the composite of **D** with this map gives the desired $\mathbf{1-1}$ map from $\mathbf{R}$ to $\mathbf{P(N)}$.

Let $\mathbf{P_\infty(N)}$ denote the set of all ***infinite*** subsets of $\mathbf{N}$, and define a function from $\mathbf{P_\infty(N)}$ to $\mathbf{R}$ as follows: Given an infinite subset **B**, let χ_B be its characteristic function and consider the infinite series

$$\Sigma_B = \Sigma_k \chi_B(k) \cdot 2^{-k}.$$

This series always converges by the Comparison Test because its terms are nonnegative and less than or equal to those of the geometric series $\Sigma_k\ 2^{-k}$, which we know is convergent. Furthermore, different ***infinite*** subsets will yield different values (look at the first value of **k** that is in one subset but not in the other; if, say, **k** lies in **A** but not in **B** then we have $\Sigma_A > \Sigma_B$. Note that all these sums lie in the interval $\mathbf{[0, 1]}$ because $\Sigma_k\ 2^{-k} = 1$.

If **A** is a ***finite*** subset, consider the finite sum

$$\Sigma_B = 2 + \Sigma_k \chi_B(k) \cdot 2^{-k}.$$

Once again it follows that different finite subsets determine different real (in fact, rational) numbers. Furthermore, since the value associated to a finite set lies in the interval $\mathbf{[2, 3]}$ it is clear that a finite set and an infinite set cannot go to the same real number. Therefore we have constructed a $\mathbf{1-1}$ function from $\mathbf{P(N)}$ to $\mathbf{R}$.■

Since we have constructed $\mathbf{1-1}$ mappings in both directions, we can apply the Schröder – Bernstein Theorem to complete the proof.

Finally, we prove another fundamental, well – known result about the cardinality of $\mathbf{R^n}$:

<u>Theorem 13.</u> *Given an arbitrary set* **A**, *let* $\mathbf{A^n}$ *denote the* **n** *– fold product of* **A** *with itself. If* **R** *denotes the real numbers, then for all positive integers* **n** we have $\mathbf{|R^n| = |R|}$.

One slightly nonintuitive consequence of this theorem is the existence of a $\mathbf{1-1}$ correspondence between the points of the number line and the points on the coordinate plane. Of course, these objects with all their standard mathematical structures are quite different, but the theorem says that they cannot be shown to be distinct simply by means of transfinite cardinal numbers.

Using the axiom(s) introduced in the next section one can show that $\mathbf{n \cdot |A| = |A|}$ and $\mathbf{|A^n| = |A|}$ as above for every infinite set **A** and positive integer **n**, but here we shall

outline a direct and relatively standard argument which does not depend upon the additional axiom(s).

Proof. There are two parts to the proof. The first is to verify the result when $n = 2$ and the second is to show that the case $n = 2$ implies the general case. The argument to prove the latter is essentially the same as in the Corollary to the Idempotent Laws for the cardinal number $\aleph_0$ (specifically, see Corollary 9).

We now concentrate on the case $n = 2$. The argument is based upon the existence of a $1 - 1$ correspondence

$$\{0, 1\}^{N} \rightarrow \{0, 1\}^{N(0)} \times \{0, 1\}^{N(1)}$$

sending a function $N \rightarrow \{0, 1\}$ to the ordered pair given by its restrictions to the even and odd natural numbers; clearly a function is completely determined by these restrictions, and conversely given functions on the even and odd natural numbers there is a unique way of assembling them into a function defined on all the natural numbers. This observation yields the cardinal number identity

$$2^{\aleph_0} = 2^{\aleph_0} \times 2^{\aleph_0}$$

and the validity of the theorem for $n = 2$ follows from this and the previously established identity $|R| = 2^{\aleph_0}$.■

Corollary 14. *We also have* $2^{\aleph_0} = 2^{\aleph_0} + 2^{\aleph_0}$ *and* $2^{\aleph_0} = \aleph_0 \times 2^{\aleph_0}$.

Proof. These are consequences of the following chain of inequalities:

$$2^{\aleph_0} \leq 2^{\aleph_0} + 2^{\aleph_0} \leq \aleph_0 \times 2^{\aleph_0} \leq 2^{\aleph_0} \times 2^{\aleph_0} = 2^{\aleph_0}\text{■}$$

Remark. The following generalizations of the usual laws of exponents also hold for cardinal numbers:

Theorem 15. (Transfinite Laws of Exponents). *If* α, β *and* γ *are (finite or transfinite) cardinal numbers, then the following equations hold:*

$$\gamma^{\alpha+\beta} = \gamma^{\alpha} \cdot \gamma^{\beta}$$

$$(\gamma^{\alpha})^{\beta} = \gamma^{\alpha\beta}$$

$$(\beta\gamma)^{\alpha} = \beta^{\alpha} \cdot \gamma^{\alpha}$$

The last two equations follow from the $1 - 1$ correspondences for function sets that were discussed in Section IV.5 (see Theorem IV.5.7), and the proof of the first follows from the analogous $1 - 1$ correspondence between $C^{A \amalg B}$ and $C^A \times C^B$, a special case of which was discussed in the proof of Theorem 13 in this section.■

Applications to transcendental numbers

Cantor was led to develop set theory in his study of some basic questions about trigonometric series, and a few years after beginning this work he found a striking

application to a longstanding problem of independent interest. We begin with the definitions needed to formulate the problem.

Definition. Let **x** be a real number. Then **x** is said to be ***algebraic*** if there is a nontrivial polynomial with rational coefficients (equivalently, integral coefficients; *cf*. next paragraph) for which **x** is a root. A real number is said to be ***transcendental*** if it is not a root of any such polynomial.

Since every polynomial over the rational numbers can be written as an integral polynomial divided by a nonzero integer, it follows that a number is a root of a nontrivial polynomial over the rational numbers if and only if it is a root of a nontrivial polynomial over the integers..

Lemma 16. *If* **x** *and* **y** *are real numbers such that* **x** *is rational and* **y** *is transcendental, then their sum* **x + y** *is transcendental.*

Proof. Suppose that **x + y** is algebraic. Then there is a nontrivial polynomial **p** with rational coefficients which has **x + y** as a root. Dividing through by the (nonzero) coefficient of the highest degree term of **p** if necessary, we can assume that **p** is a monic polynomial. Express this monic polynomial as $\mathbf{t^n + q(t)}$, where **q** has lower degree. Our hypotheses then imply that $\mathbf{(x + y)^n + q(x + y) = 0.}$ By the Binomial Theorem we may rewrite this as $\mathbf{y^n + r(y) = 0,}$ where **r(t)** is another polynomial of lower degree with rational coefficients. This implies that **y** is algebraic, contradicting our original assumption, and hence the only possibility is that **x + y** must be transcendental**.**■

Corollary 17. *If there is at least one transcendental real number, then the cardinality of the set* **T** *of transcendental real numbers satisfies* $\mathbf{\aleph_0 \leq |T|.}$

Proof. Suppose that **y** is transcendental. Then one can define a mapping from the rational numbers **Q** to **T** sending $\mathbf{x \in Q}$ to $\mathbf{x + y \in T.}$ This mapping must be $\mathbf{1 - 1}$ because **x + y = z + y** implies **x = z.**■

In the next unit we shall prove a more general result about infinite cardinal numbers, but the preceding corollary gives us what we need for the time being.

In order to compare the algebraic and transcendental real numbers, we need to know the cardinality of the former, and it is given by the following result:

Theorem 18. *The set of all algebraic real numbers is countably infinite.*

***Proof.* (*)** Since the set of algebraic real numbers contains the integers, it will suffice to show that the set of algebraic numbers is countable. For each positive integer **n** let $\mathbf{A_n}$ be the set of all real numbers **r** such that **r** is a root of a polynomial of degree **n** with rational coefficients. Since a countable union of countable sets is countable, it will suffice to show that each set $\mathbf{A_n}$ is countable.

Let $\mathbf{P_n}$ denote the set of all polynomials of degree **n,** and for each $\mathbf{p \in P_n}$ let **W(p)** denote the set of real roots for **p**. Basic results on roots of polynomials show that each set **W(p)** is finite. If we can show that $\mathbf{|W(p)| = \aleph_0,}$ it will follow that $\mathbf{A_n}$ is a countable

union of the finite sets **W(p),** where **p** runs through the elements of $\mathbf{P_n}$, and hence $\mathbf{A_n}$ is countable.

Now a polynomial in $\mathbf{P_n}$ has the form

$$\mathbf{p(t) = a_n t^n + \ldots + a_1 t + a_0}$$

where $\mathbf{a_n \neq 0}$, and hence it is completely determined by the coefficients of the powers of the indeterminate, say **t,** ranging from 0 to the degree, which in this case is **n.** This means there is a canonical $\mathbf{1-1}$ correspondence between $\mathbf{P_n}$ and $\mathbf{(Q - \{0\}) \times Q^n}$ (where as usual $\mathbf{Q}$ denotes the rational numbers) which is given by taking the coefficients of $\mathbf{t^k}$ as **k** runs from **n** to 0. Now we know that $\mathbf{|Q| = \aleph_0}$ by Proposition 11, and we also know that $\mathbf{|Q - \{0\}| = |Q|}$ by Propositions 7 and 11, so that we have $\mathbf{|W(p)| = (\aleph_0)^{n+1}}$. However, by Corollary 9 we also know that $\mathbf{(\aleph_0)^k = \aleph_0}$ for all values of k, and this means that $\mathbf{|W(p)| = \aleph_0}$ must be true. As noted before, this completes the proof of the theorem.■

Historical remarks on transcendental numbers. It is not clear when mathematicians first considered the concept of a transcendental number, but various historical facts strongly suggest that this took place near the middle of the 17th century in connection with the results and viewpoints of R. Descartes (***cf.*** page 343 of Burton). A few years later, J. Gregory (1638 – 1675) tried to show that both π and **e** were transcendental; however, his work had a small but irreparable error. Leibniz also concluded that π was transcendental but did not make a significant effort to prove this. Several 18th century mathematicians such as C. Goldbach (1690 – 1764), D. Bernoulli (1700 – 1782), J. H. Lambert (1728 – 1777), and A. – M. Legendre (1752 – 1833) had considered the possible existence of transcendental numbers, and there was a general agreement that numbers π and e should be transcendental although it was not clear how one might actually prove these statements. One important piece of evidence was the understanding at the time that some of the standard functions in calculus like **sin x** and $\mathbf{e^x}$ were not algebraic functions (***i.e.,*** there is no nontrivial polynomial in two variables such that **P(x, f(x))** is identically zero). We shall discuss this point in greater detail below. The existence transcendental numbers was first shown rigorously by J. Liouville (1809 – 1882) in the 1840s. Probably the best known example arising from his work is the so – called ***Liouville constant***:

$$c = \sum_{j=1}^{\infty} 10^{-j!} = 0.11000100000000000000001000\ldots.$$

The following online sites provide further information about Liouville's methods and results:

http://planetmath.org/encyclopedia/ExampleOfTranscendentalNumber.html

http://en.wikipedia.org/wiki/Liouville_number

During the next few decades, proofs that **e** and π were transcendental finally appeared; these results were due to C. Hermite (1822 – 1901) and F. Lindemann (1852 – 1939) respectively. Many other easily constructed numbers have been shown to be transcendental numbers since the original results of Liouville, but there are still many

open questions that are very easy to state but seem unlikely to be answered in the near future. The current state of affairs is summarized in the following online site:

http://mathworld.wolfram.com/TranscendentalNumber.html

The purpose of the preceding discussion is to put Cantor's result on transcendental numbers into perspective. At the time, the existence of such numbers had only recently been established, and the proofs required delicate manipulations of equations and inequalities. In contrast, Cantor's existence proof did not require any significant computations, but it also did not produce any explicit examples (although one can combine Cantor's diagonal process argument with Liouville's construction to describe an uncountable family of transcendental numbers). We should note that currently known results are still not adequate to answer many very easily stated questions; for example, whether πe or $\pi + e$ is transcendental (however, we do know that at least one of these numbers is transcendental).

Theorem 19. (Strong existence theorem for real transcendental numbers – Cantor). *The set of transcendental real numbers is nonempty, and its cardinality is equal to* $2^{\aleph_0}$.

Proof. As in the preceding discussion, the set of real numbers $\mathbf{R}$ splits into a union of the disjoint subsets **A** of algebraic real numbers and **T** of transcendental real numbers. Thus we have $|\mathbf{R}| = |A| + |T| = \aleph_0 + |T|$. If **T** were empty we would have $|\mathbf{R}| = \aleph_0$, and we know this is false by the results of Section 4. Therefore **T** must be nonempty, and by the lemma above it follows that there is a $1-1$ mapping from **A** into **T;** let T_0 denote the complement of its image, so that $|T| = |A| + |T_0| = \aleph_0 + |T_0|$. Therefore we have

$$|T| = \aleph_0 + |T_0| = \aleph_0 + \aleph_0 + |T_0| = \aleph_0 + |T| = |R| = 2^{\aleph_0}.$$

We now indicate how one can use Cantor's result to answer one of the questions at the beginning of these notes in the very strong informal sense:

> *Almost every real number is transcendental. In particular, if one "chooses a real number at random," it will almost certainly be transcendental.*

Giving a mathematically precise definition of random choice is far beyond the scope of this course, but here is a discussion that can be made mathematically rigorous. Let us agree to restrict attention to real numbers in the closed unit interval $[0, 1]$. Given a reasonable subset **A** of the latter (these will include all countable subsets), one would like to estimate the probability that an element of the interval chosen at random will belong to **A.** If, say, we divide the interval into **n** nonoverlapping pieces of equal length, then the likelihood of choosing an element from one of the pieces should be just $1/n$. More generally, if we are given a subinterval of length **L** then the likelihood of choosing a point from the subinterval should be **L.**

How does this apply in our situation? Suppose that **B** denotes the algebraic numbers in the closed unit interval, so that **B** is countable by our previous results. Choose a $1-1$ correspondence with the natural numbers, and let $m > 0$ be an integer. For each **n,** let J_n be a subinterval of length $2^{-(m+n)}$ containing the n^{th} point in **B.** The likelihood that a chosen element will lie in **B** should be no greater than the likelihood that it will lie in the union of the intervals J_n and hence it should be no greater than the sums of the lengths of these intervals. We can use a geometric series argument to see that the latter sum is

equal to 2^{1-m}. Now **m** is arbitrary, so this means that the likelihood of randomly choosing an element from **B** is no greater than 2^{1-m} for every positive integer **m,** and hence (since it is nonnegative) this likelihood must be equal to zero. Informally, this means that if we pick a number from the unit interval at random, it is almost certain to be a transcendental number.■

Footnote on transcendental functions. In the discussion above we have asserted that certain basic functions such as trigonometric functions and exponential functions are transcendental. Since it is difficult to find statements or proofs of these facts written out explicitly, we shall explain how the proof for the usual exponential function follows from standard results on solutions to ordinary differential equations which are covered in lower division undergraduate courses and we shall give an online reference that considers the remaining elementary transcendental functions.

The first step is fairly simple.

Lemma 20. *Let* **f(x)** *be a continuous function on some interval. Then f is transcendental if and only if for every positive integer* **m** *the* $(m+1)^2$ *functions* $x^p f(x)^q$ *are linearly independent over the real numbers, where* $0 \le p, q \le m$.

Proof. The $(m+1)^2$ functions $x^p f(x)^q$ are linearly ***dependent*** over the reals if and only if there is some set of coefficients $c_{p,q}$ which are not all zero such that $\sum c_{p,q} x^p f(x)^q = 0$. Thus if they are linearly dependent for some **m,** then there is a nontrivial polynomial $G(x, y) = \sum c_{p,q} x^p y^q$ such that $G(x, f(x)) = 0$. Conversely, if we are given such a polynomial **G** and **m** is the highest power of **x** or **y** that appears, then it follows that the $(m+1)^2$ functions $x^p f(x)^q$ are linearly dependent over the real numbers. Sdfa sadf By the lemma, proving that the exponential function e x is transcendental amounts to showing that the functions $x^p e^{qx}$ are linearly independent functions for $0 \le p, q \le m$, where **m** is an arbitrary positive integer. One relatively quick way to see this is to notice that the functions in question all satisfy an N^{th} order *homogeneous linear (ordinary) differential equation with constant coefficients*

$$D^N y + a_{N-1} D^{N-1} y + \ldots + a_1 D y + a_0 y = 0$$

where $N = (m+1)^{(m+1)}$ and $D^k y$ denotes the k^{th} derivative of **y.** Specifically, this is the equation for which the associated characteristic polynomial

$$p(t) = a_N t^N + a_{N-1} t^{N-1} + \ldots + a_1 t + a_0$$

is given by the following product:

$$p(t) = t^{m+1} (t-1)^{m+1} \ldots (t-m)^{m+1}$$

The linear independence of these solutions is a standard fact in the theory of ordinary differential equations, and in particular, the proof is described in Section 9.2 of the following representative textbook on the subject:

> W. F. Trench, ***Elementary Differential Equations***. Brooks/Cole (Thomson Learning), Pacific Grove CA, 2000. ISBN: 0–534–36841–7.

More specific references for the proof are essentially the entire content of pages 453 – 454 as well as Exercise 40 on page 457.

This linear independence result was essentially known in the 18th century to L. Euler (1707 – 1783), with some refinements of the concepts due to G. Monge (1746 – 1818) and A. – L. Cauchy (1789 – 1857).

The online document

http://math.ucr.edu/~res/math144/transcendentals.pdf

establishes similar results for the other so – called ***elementary transcendental functions*** that are studied in precalculus and calculus, and it provides some additional general perspective on determining when a function is algebraic or transcendental. Since this document uses material on extension fields from advanced undergraduate and beginning graduate courses, it is included mainly for reference purposes; although the main results are extremely well – known, it is extremely difficult to find a reference in which the various functions are actually proven to be transcendental. Asd

Cardinal number problems for further consideration

Here are some natural questions that arise in connection with the results of this section. Some involve generalizations of these results, and others are simple questions about the arithmetic and ordering properties of cardinal numbers.

1. Is the partial ordering of cardinal numbers a linear ordering?
2. Is $\aleph_0$ the smallest transfinite cardinal number?
3. If **A** is an infinite set, does it follow that the idempotent identities $|A| \cdot |A| = |A|$ and $|A| + |A| = |A|$ always hold?
4. If there is a surjection from **A** to **B,** does it follow that $|B| \leq |A|$?
5. Given a cardinal number α**,** is there a unique minimal cardinal number β such that $\beta > \alpha$?

Most of these seem likely, and the final question is closely related to Cantor's terminology for transfinite cardinal numbers. For example, if the answers to this question and the first one are yes, then one can define $\aleph_1$ to be the unique minimal cardinal number strictly greater than $\aleph_0$**,** then take $\aleph_2$ to be the unique minimal cardinal number strictly greater than $\aleph_1$**,** and so on.

However, despite strong intuitive feelings that the preceding questions have affirmative answers, we are not yet equipped to answer such questions, and the material in the next two units is needed to provide answers. Before introducing this material, we shall devote the next section to a discussion of some ways in which Cantor's theory of sets was a radical departure from previous views of infinite objects in mathematics.

VI.5 : The impact of set theory on mathematics

Given the routine use of set theory throughout modern mathematics, it is easy to overlook the precedent shattering nature of Cantor's legacy. The rest of this section provides some historical perspective.

It is not known exactly when questions about the concept of infinity first arose, but the well – known paradoxes due to Zeno of Elea (***c.*** 490 – 430 B. C. E) indicate that ancient Greek philosophers and mathematicians recognized that difficulties arise when one attempts to discuss the infinite. The writings of Aristotle (384 – 322 B. C. E.) provided an effective way of confronting such questions by arguing that there were two kinds of infinity.

1. ***Actual infinity***, or ***completed infinity***, which Aristotle believed could not exist, is endlessness fully realized at some point in time.
2. ***Potential infinity***, which Aristotle maintained was manifest in nature — for example, in the unending cycle of the seasons or the indefinite divisibility of measurements — is infinitude spread over unlimited time and space.

This fundamental distinction between potential and actual infinity persisted in European mathematics for more than 2000 years.

However, the adoption of this distinction did not mean that speculation about infinity was absent from all of mathematics during that time. Speculations about infinity appeared in classical Indian mathematics, particularly in the writings of Bhaskara (also known as Bhaskara **II** or Bhaskaracharya, 1114 – 1185). By the end of the Middle Ages, various scientific, philosophical and theological questions about infinity received considerable attention in Europe as well as India and China. Many of the mathematical advances concerned summations of infinite series. With hindsight, it is apparent that the summation formulas for many series obtained during these centuries showed that the concept of completed infinity could be mathematically meaningful, at least in some contexts. Certain basic paradoxes and puzzles arose and provided further evidence that actual infinity was not an issue to be dismissed easily. Specific problems arise from many standard **1 – 1** correspondences between infinite sets and certain proper subsets; for example, between the nonnegative integers and the even nonnegative integers. These constructions seemed to contradict a commonsense idea that appears in Euclid: ***The whole is always greater than any of its (proper) parts.*** The writings of Galileo (G. Galilei, 1564 – 1642) on such problems were the first to suggest a more enlightened attitude toward the infinite; in particular, he proposed that ***"****infinity should obey a different arithmetic than finite numbers.****"*** We have seen that one version of Galileo's idea plays an important role in Cantor's work. However, during the nearly three centuries between Galileo and Cantor, mathematicians managed to avoid confronting questions about infinity for the most part. By confining their attention to Aristotle's potential infinity, mathematicians were able to address problems and develop crucial concepts including infinite series, limit, and infinitesimals [*sic*], and thus to develop calculus without having to grant that infinity itself was a mathematical object. In fact, early in the 19th century the highly eminent mathematician C. F. Gauss (1777 – 1855) expressed his "horror of the actual infinite" in the following terms:

> I protest most vehemently against the use of infinite magnitude as something completed, which is never permissible in mathematics. The infinite is merely a figure of speech, the true meaning being a limit which certain ratios approach as closely as we wish, while others may be permitted to increase beyond all bounds.

Even Cantor admitted that considering infinite sets as single entities — not as merely going on forever but as completed objects — was a concept to which he had been **"**logically forced, almost against my will**."** This erasing of the distinction between potential and actual infinities was **"**in opposition to traditions that had become valued**."**

Cantor's ideas generated considerable opposition and controversy for several reasons. For many mathematicians, the sets themselves were less disturbing than the uses to which Cantor put them; some mathematicians were particularly uneasy with Cantor's proof showing that **"**almost every**"** real number is transcendental; *i.e*., they are not roots of polynomial equations with rational coefficients. As noted in the discussion of Cantor's result, a considerable amount of intricate calculation is needed to prove that there are transcendental numbers and to verify the "obvious facts" that familiar numbers like **e** and π are transcendental. Cantor's existence proof required no significant computations at all, and in some respects it looks as if one is getting something for nothing. Of course, one reason the argument is so simple is that it does not provide any way of deciding whether a given number is algebraic or transcendental.

Cantor's result on transcendental numbers was the first important example of what has come to be called a pure-existence proof. Giving not the slightest hint of how to construct even a single transcendental number, it established the existence of a host of such numbers by proving that it would be contradictory for them not to exist. Once again the basic issue is infinity. A proof by *reductio ad absurdum* that establishes the existence of an object in a finite set is perfectly acceptable to any mathematician; one can always in principle produce the object by checking through all the members of the set.

But the same is not true for, say, the transcendental numbers, which belong to the infinite set of real numbers. For this reason many mathematicians rejected Cantor's proof completely, objecting that a contradiction was no substitute for a tangible example.

However, other mathematicians were unwilling to accept Cantor's entire approach, which challenged established mathematical principles like the previously mentioned avoidance of actual or completed infinity. For example, H. Poincaré (1854 – 1912) expressed his disapproval in a statement that Cantor's set theory would be considered by future generations as "a disease from which one has recovered." Much stronger criticism was voiced by L. Kronecker (1823 – 1891), who strongly maintained that the appropriate objects for mathematical study were those that could be realized in a fairly concrete fashion (for example, his views excluded transcendental numbers entirely). Such a perspective leaves little place for the explicit treatment of "actual infinity" that permeates Cantor's work. On the other hand, not all leading mathematicians were opposed to Cantor's ideas. Some highly eminent mathematicians such as G. Mittag – Leffler (1846 – 1927), K. Weierstrass (1815 – 1897), and long – time friend R. Dedekind supported Cantor's ideas and defended them against his critics. Aside from the revolutionary nature of Cantor's ideas, another reason for reservations about them was that some key concepts were initially expressed in a somewhat imprecise fashion, and yet another was that some basic questions about manipulating infinite sets turned out to be far more challenging than they seemed at first; these will be discussed further in Section 4. Unfortunately, the strain of the controversy over Cantor's work ultimately inflicted an extremely heavy toll on him.

Of course, our use of Cantor's ideas today and our presentation of his existence proof for transcendental numbers both indicate that his methods and results were increasingly accepted as mathematically valid. In particular, during the years immediately following Cantor's work, some mathematicians solved some other fundamental problems using pure, nonconstructive existence proofs; the most striking result of this sort called the ***Hilbert Basis Theorem*** was obtained by D. Hilbert (1862 – 1943) in 1889. A statement of this result requires concepts well beyond the scope of this course, but for the sake of completeness here is an online reference to one fundamental but (relatively) elementary class of special cases:

http://en.wikipedia.org/wiki/Hilbert's_basis_theorem

Hilbert was one of the most influential mathematicians of his time, and his acceptance of Cantor's work reflected the incorporation of set theory into the mainstream of mathematics. The following frequently quoted statement states his position very strongly but concisely: ***No one shall expel us from the paradise that Cantor has created.***

Hilbert addressed concerns about increasing abstraction by stressing the vast amount that could be done if one adopts such an approach in contrast to the relatively limited amount that could be done if one does not. To most mathematicians in the early 20th century, Hilbert's formalist viewpoint offered an attractive viewpoint, and a largely dominant majority of present day mathematicians also take a modified formalist view towards the subject. These modifications are necessary because of the fundamental incompleteness results due to K. Gödel that will be discussed in the next unit.

VI. 6 : Transfinite induction and recursion

(Halmos, §§ 12 – 13, 17 – 20; Lipschutz, §§ 8.1 – 8.9, 8.12 – 8.13)

This section has two objectives. The first is to formulate concepts of

(1) proof by transfinite induction,
(2) definition by transfinite recursion,

which apply to well – ordered sets that are larger than the nonnegative integers. The second aim is to summarize the basic properties of ordinal numbers that are used most often in mathematics.

The proofs of many crucial results on well – ordered sets are considerably less elementary than most of the material in these notes. In particular, at several steps one needs slightly stronger versions of some axioms and definitions than we have stated in these notes. Precise statements appear in the book by Golrei cited at the beginning of

the first unit of these notes; in cases where we have stated simplified versions of axioms, we have done so for the sake of clarity and because the simpler versions are adequate for nearly everything one wishes to do in other branches of mathematics. Finally, for most mathematical purposes the theory of well – ordered sets are mainly significant as means to some other end, and such objects play less of direct role in other branches of mathematics than the other material discussed in these notes. For these reasons, we shall not attempt to give all the details of the more complicated proofs here, but instead we shall describe some of the arguments and give references to the book by Goldrei. None of the subsequent material in these notes will depend upon the results that are stated without complete proofs.

Given the relative difficulty of some material in this section, the following suggestions might be helpful. The most important thing to do is to concentrate on understanding the definitions and statements of the main results. This should provide enough information to read the remaining sections in these notes. When these points are understood, a natural second step is to understand the outlines and main ideas of the proofs well enough to be able to summarize or explain them. For the purposes of this course, the final level of mastery is to have a full understanding of all the steps in the proofs.

Traditionally the elements of a well – ordered set are denoted by expressions involving nonnegative integers and Greek letters, and we shall follow this convention here.

Notational conventions. Suppose that **X** is a well – ordered set. The least element of **X** will be denoted by **0** or by $\mathbf{0}_X$ when it is necessary to stress the dependence upon **X.** If $\alpha \in \mathbf{X}$, the ***initial segment*** associated to α is the set of all β such that $\beta < \alpha$, and it is denoted by $[0, \alpha)$ or less ambiguously by $[0, \alpha)_X$. Likewise, we define the ***closed interval*** $[0, \alpha]$ to be the set of all β such that $\beta \leq \alpha$. Given a well – ordered set **X,** its ***immediate successor*** $\mathbf{X + 1}$ is the set $\mathbf{X} \cup \{ \mathbf{X} \}$ with the original well – ordering on **X** and the added element **X** strictly greater than every $\alpha \in \mathbf{X}$. Recall that we have constructed set theory so that no set will be a member of itself, and thus it follows that **X** is distinct from each $\alpha \in \mathbf{X}$.

Transfinite induction and recursion

Transfinite induction is an adaptation of proof by mathematical induction to include (large) well-ordered sets.

Before discussing this principle it will be useful to make the following elementary observation.

Proposition 1. *Let* **X** *be a well – ordered set, and let* $\alpha \in \mathbf{X}$. *Then exactly one of the following is true.*

> **(1)** *There is a* $\beta \in \mathbf{X}$ *such that* α *is the first element in* **X** *that is strictly larger than* β, *and* α *is not the least upper bound of all elements of* **X** *that are strictly less than* α.

(2) *For each* β *such that* $\beta < \alpha$ *there is some* $\gamma \in \mathbf{X}$ *such that* $\beta < \gamma < \alpha$**,** *and* α *is the least upper bound of all elements of* **X** *that are strictly less than* α**.**

Proof. If the first holds, then β is the least upper bound of all elements of **X** that are strictly larger than α**.** Suppose now that the second holds. Clearly **a** is an upper bound for the set in question. To see that it is the least upper bound, note that if $\beta < \alpha$ then β cannot be an upper bound because there is always some γ such that $\beta < \gamma < \alpha$.■

Notation. Elements of the first type are called ***(immediate) successor elements*** (and one often writes $\alpha = \beta + 1$ or $\alpha = \beta^+$ in this case), and elements of the second type are called ***limit elements.***

We now proceed to the main results.

Theorem 2. (Principle of transfinite induction.) *Let* **X** *be a well – ordered set, and suppose that for each* $\alpha \in \mathbf{X}$ *we are given a statement* $\mathbf{S}(\alpha)$ *such that the following conditions hold:*

(1) *If* $\mathbf{0}_X$ *denotes the unique minimum element of* **X,** *then* $\mathbf{S}(\mathbf{0}_X)$ *is true.*

(2) *For each* β *in* **X,** *if* $\mathbf{S}(\gamma)$ *is true for all* $\gamma < \beta$**,** *then* $\mathbf{S}(\beta)$ *is true,*

Then $\mathbf{S}(\alpha)$ *is true for every* $\alpha \in \mathbf{X}$**.**

Proof. The argument is similar to the one for finite induction. Suppose that at least one of the statements is false. Then there is a unique minimum α_0 such that $\mathbf{S}(\alpha_0)$ is false. Since $\mathbf{S}(\mathbf{0}_X)$ is true we know that $\alpha_0 \neq \mathbf{0}_X$ and thus the set of all β such that $\beta < \alpha$ must be nonempty. For each such β the statement $\mathbf{S}(\beta)$ is true, and therefore by the second condition we know that $\mathbf{S}(\alpha_0)$ is also true. Now this contradicts our choice of α_0, and the problem arises from our assumption that at least one of the statements $\mathbf{S}(\alpha)$ is false. Thus all of the statements must be true.■

In practice, the verification of the second condition often splits into two cases: One for successor elements (those which have an immediate predecessor), where the usual inductive approach can be applied to show that $\mathbf{P}(\gamma)$ implies $\mathbf{P}(\gamma + 1)$**,** and the case for limit elements, which have no predecessor, and thus cannot be handled by such an argument.

Typically, the case for limit ordinals is approached by noting that a limit element β is the least upper bound of all elements $\gamma < \beta$ and using this fact to prove $\mathbf{P}(\beta)$ assuming that $\mathbf{P}(\gamma)$ holds true for all $\gamma < \beta$**.**

Transfinite recursion is a closely related to transfinite induction, but the latter is a method of ***proof*** and the former is a method of ***definition*** or of ***construction.*** The basic idea is fairly simple. We start with a well – ordered set Λ and specify the object for the zero (least element), then assuming we know how to define the object indexed by γ for

every $\gamma < \alpha$, we use this partial function to find $f(\alpha)$. In a little more detail, one defines a family of objects indexed by the well – ordered set X — say B_α, for every $\alpha \in X$, or perhaps every α less than some bound ξ — by specifying three things:

- What B_0 is.
- How to determine $B_{\alpha+1}$ from B_α (or possibly from the entire sequence up to B_α).
- For a limit element α, how to determine B_α from the sequence of previously determined B_γ for $\gamma < \alpha$.

Formally there is not much formal difference between the second and third items, but in practice they are so often distinct that it is useful to present them separately.

Here is the formal statement.

Theorem 3. (Transfinite Recursive Definition Theorem.) *Suppose that* X *is a well – ordered set and* B *is a set which does not necessarily have any additional structure. Assume also that for* $\alpha \in X$ *we have a function* $H : B^{[0,\alpha)} \to B$, *and let* $z_0 \in B$. *Then there is a unique function* $f : X \to B$ *such that* $f(0) = z_0$ *and for all positive* n *we have*

$$f(\alpha) = H\left(f\big|[0,\alpha)\right).$$

Proof. The approach is parallel to the proof of the (Finite) Recursive Definition Theorem that was proven in Section V.2. One establishes existence by defining a sequence of functions $g_\alpha : [0,\alpha] \to B$ which agree on the overlapping subsets; one then constructs a function g whose graph is the union of the graphs of the partial functions. The uniqueness proof will then reduce to proving uniqueness for the restrictions to each subset $[0,\alpha]$.

The function $g_0 : \{0\} \to B$ is defined by $g_0(0) = z_0$. Suppose we are given the functions $g_\beta : [0,\beta] \to B$ for $\beta < \alpha$, where one has the compatibility relations $g_\beta = g_\beta \mid [0,\gamma]$ for $\gamma < \beta$. Since $[0,\alpha) = \cup_{\beta<\alpha}[0,\beta]$ it follows that we can define a function k_α on the left hand side whose restriction to each subset $[0,\beta]$ is g_β. We can extend this to a function g_α the closed interval $[0,\alpha]$ by setting $g_\alpha(\delta)$ equal to $H(k_\alpha)$. Let f be the function whose union is the graphs of the functions g_α for all $\alpha \in X$. By construction this function has the properties specified in the theorem.

To conclude the proof, we need to show uniqueness. Suppose that f' is an arbitrary function satisfying the given properties, and let f be constructed as in the previous paragraphs. Suppose that $f \neq f'$. By hypothesis both agree at zero, so there exists a unique minimal element $\alpha > 0$ at which their values disagree. In particular, the functions agree on the initial segment $[0,\alpha)$, and thus by the displayed condition we have

$$f(\alpha) = H\left(f\big|[0,\alpha)\right) = H\left(f'\big|[0,\alpha)\right) = f'(\alpha),$$

where the first equation is true by construction, the second is true by the minimality hypothesis on α, and the third is true by the assumption on f'. This contradicts our assumption that the two functions had different values at α, and it follows that there cannot be a point where the values of the two functions are unequal. ■

Comparison of well – ordered sets

The following basic fact about well – ordered sets is extremely important for many purposes, and it illustrates the concept of definition by transfinite recursion**.**

Theorem 4. *Let* **X** *and* **Y** *be well – ordered sets. Then there exists a nondecreasing map* $\mathbf{f : X \to Y + 1 = Y \cup \{ Y \}}$ *such that the following hold:*

- **(1)** *If* $\mathbf{X_0 = f^{-1}[Y]}$, *then* $\mathbf{f \mid X_0}$ *is strictly increasing.*
- **(2)** *If* $\mathbf{\alpha \in X_0}$, *then* **f** *defines a* $\mathbf{1 - 1}$ *order – preserving correspondence between the initial segments* $\mathbf{[0_X, \alpha)}$ ***and*** $\mathbf{[0_X, f(\alpha))}$.
- **(3)** *If* $\mathbf{f(\alpha) = Y \in Y + 1 = Y \cup \{ Y \}}$ *then* $\mathbf{f(\beta) = Y}$ *and* $\mathbf{f[[0_X, \alpha)] \supset Y}$.

Proof. We construct the map **f** by transfinite recursion, beginning with $\mathbf{f(0_X) = 0_Y}$. Suppose that $\mathbf{\alpha > 0_X}$ and one has $\mathbf{g_\alpha = f \mid [0, \alpha)}$ is defined with the given properties on $\mathbf{[0, \alpha)}$. By construction, if $\mathbf{g_\alpha(\beta) \in Y}$ then $\mathbf{g_\alpha [[0, \beta)] \subset Y}$. There are now two cases.

Case A. $\mathbf{g_\alpha(\beta) \neq Y}$ for all $\mathbf{\beta \in [0, \alpha)}$. **CLAIM:** Either there is an upper bound for the image of $\mathbf{g_\alpha}$ or else $\mathbf{g_\alpha ([0, \alpha)) = Y}$ for some $\mathbf{\beta < \alpha}$. If the second alternative is false, then $\mathbf{g_\alpha}$ is not onto, so let $\mathbf{\gamma}$ be an element not in the image. We claim that no \delta satisfying $\mathbf{\delta > \gamma}$ can be in the image either. If it were, then the second property would imply that $\mathbf{\gamma}$ would also be in the image. Therefore $\mathbf{\gamma}$ must be an upper bound for the image of $\mathbf{g_\alpha}$. Extend the definition of $\mathbf{g_\alpha}$ to include $\mathbf{\alpha}$ by taking $\mathbf{g_\alpha(\alpha)}$ to be the least element of **X** that is not in the set $\mathbf{g_\alpha ([0, \alpha))}$.

Case B. $\mathbf{g_\alpha(\beta) = Y}$ for some $\mathbf{\beta < \alpha}$. In this case we extend the definition of $\mathbf{g_\alpha}$ to include $\mathbf{\alpha}$ by setting $\mathbf{g_\alpha(\alpha) = Y}$.

Thus we have constructed a map $\mathbf{g_\alpha}$ on $\mathbf{[0, \alpha]}$ and it is an elementary exercise to show it has the desired properties.■

The preceding result has the following important consequence; text references are page 73 of Halmos and Theorem 8.10 on page 207 of Lipschutz.

Theorem 5. *Let* **X** *and* **Y** *be well – ordered sets. Then either there is a* $\mathbf{1 - 1}$ *order – preserving map from* **X** *to* **Y** *or there is a* $\mathbf{1 - 1}$ *order – preserving map from* **Y** *to* **X.** *In each case one can choose the mapping so that its image is an initial segment or the whole set.*

Proof. Let **f** be as in the previous result. There are two possibilities.

Case A. Suppose that $\mathbf{f[X] \subset Y}$. — In this situation there are two subcases. If the image is equal to **Y,** then f is a $\mathbf{1 - 1}$ order – preserving correspondence between **X** and **Y,** so both options are realized in this case. Suppose now that the image is a proper subset. Then **f** defines a $\mathbf{1 - 1}$ order – preserving map from **X** to **Y.** We claim that the image is in fact an initial segment. Let $\mathbf{\gamma}$ be the least element of **Y** not in the image, and

suppose that $f(\beta) < \gamma$. By the previous result, we know that $f[[0, \beta)] \subset Y$, and therefore it follows that the image of **f** is equal to $[0, \gamma)$.

Case B. Suppose that $Y \in f[X]$. — Let γ be the least element in $f^{-1}[Y]$. Then **f** defines a $1 - 1$ order – preserving correspondence from $[0, \alpha)$ to **Y,** and the inverse defines a similar map from **Y** to the initial segment $[0, \alpha)$ of **X.**■

Types of well – ordered sets

Definition. If $(X, \leq_X)$ and $(Y, \leq_Y)$ are well – ordered sets, then we shall say that they have the *same well – order type* if there is an order – preserving $1 - 1$ correspondence from **X** to **Y.** We frequently denote this relationship by $|X, \leq_X| = |Y, \leq_Y|$.

It is probably not surprising that this relation is reflexive, symmetric and transitive, so we shall do so right away.

Proposition 6. *For every well – ordered set* $(X, \leq_X)$ *we have* $|X, \leq_X| = |X, \leq_X|$. *Furthermore if* $(X, \leq_X)$ *and* $(Y, \leq_Y)$ *are well – ordered sets such that* $|X, \leq_X| = |Y, \leq_Y|$, *then* $|Y, \leq_Y| = |X, \leq_X|$. *Finally, if* $(X, \leq_X)$, $(Y, \leq_Y)$ *and* $(Z, \leq_Z)$ *are well – ordered sets such that* $|X, \leq_X| = |Y, \leq_Y|$ *and* $|Y, \leq_Y| = |Z, \leq_Z|$, *then* $|X, \leq_X| = |Z, \leq_Z|$.

Proof. For each well – ordered set $(X, \leq_X)$, the identity map $\mathbf{id}_X$ is an order – preserving 1 – 1 correspondence from **X** to itself, so the relationship is reflexive. Similarly, if we have $|X, \leq_X| = |Y, \leq_Y|$ and **f** is the associated $1 - 1$ correspondence from **X** to **Y,** then its inverse is an order – preserving $1 - 1$ correspondence from **Y** to **X.** If in addition we have $|Y, \leq_Y| = |Z, \leq_Z|$ and **g** is the associated $1 - 1$ correspondence from **X** to **Y,** then the composite **g f** is an order – preserving $1 - 1$ correspondence from **X** to **Z.**■

Definition. If $(X, \leq_X)$ and $(Y, \leq_Y)$ are well – ordered sets, then we shall say the *well – order type of* $(X, \leq_X)$ *is smaller than or equal to the order type of* $(Y, \leq_Y)$ if there is an order – preserving $1 - 1$ map from **X** to **Y** whose image is an initial segment of **Y.** We frequently denote this relationship by $|X, \leq_X| \leq |Y, \leq_Y|$.

We shall show that the relationship in the preceding paragraph behaves like a linear ordering. Most of the properties are easy to check, but proving the relationship is antisymmetric requires the following input (*cf*. Lipschutz, Theorem 8.9, page 207):

Proposition 7. *Let* **X** *be a well – ordered set. Then there is no* $1 - 1$ *strictly increasing mapping from* **X** *to itself whose image is an initial segment* $[0, \alpha)$ *for some* $\alpha \in X$.

Proof. Suppose that there is such a map, and denote it by f. Since f is not onto, it cannot be the identity. On the other hand, by hypothesis we also have $f(0_X) = 0_X$. Therefore there must be a first β such that $f(\beta) \neq \beta$. Since $f(\gamma) = \gamma$ for $\gamma < \beta$ and β is the first element which is not in $[0, \beta)$, it follows that $f(\beta) \geq \beta$. The assumption that $f(\beta) \neq \beta$ implies that strict inequality holds. Since $f(\beta)$ lies in the image of **f,** which is equal to $[0, \alpha)$ it follows that $f(\beta) < \alpha$, and thus also that $\beta \in [0, \alpha)$ so that β lies in the image of **f.** Suppose that $f(\gamma) = \beta$. What can we say about γ? First of all, it cannot be less than

β, for $\gamma < \beta$ implies $f(\gamma) = \gamma < \beta$. However it also cannot be greater than or equal to β, for then we must have $\beta < f(\beta) \leq f(\gamma)$. This is a contradiction, and it traces back to our assumption about the image of **f.** It follows that every strictly increasing mapping from **X** to itself must be onto.■

Theorem 8. *The relationship* $\leq$ *on well – ordering types has the following properties:*

> **(1)** *For every well – ordered set* $(X, \leq_X)$ *we have* $|X, \leq_X| \leq |X, \leq_X|$. *Furthermore if* $(X, \leq_X)$ *and* $(Y, \leq_Y)$ *are well – ordered sets such that* $|X, \leq_X| \leq |Y, \leq_Y|$ *and* $|Y, \leq_Y| \leq |Z, \leq_Z|$, *then* $|X, \leq_X| \leq |Z, \leq_Z|$.
>
> **(2)** *If* $(X, \leq_X)$ *and* $(Y, \leq_Y)$ *are well – ordered sets such that* $|X, \leq_X| \leq |Y, \leq_Y|$ *and* $|Y, \leq_Y| \leq |X, \leq_X|$, *then we have* $|Y, \leq_Y| = |X, \leq_X|$.
>
> **(3)** *If* $(X, \leq_X)$ *and* $(Y, \leq_Y)$ *are well – ordered sets, then we have either* $|X, \leq_X| \leq |Y, \leq_Y|$ *or* $|Y, \leq_Y| \leq |X, \leq_X|$.

Proof. The proofs of the first assertions are similar to the corresponding arguments for order types. For the reflexive property we can use the identity mapping on **X,** and for the transitivity property, we are given strictly increasing mappings **f** and **g,** and the required map from **X** to **Z** is the composite **g f.** The dichotomy property in the third assertion is an immediate consequence of Theorem 5 from the previous subsection. Thus it only remains to prove the antisymmetric property which is stated in the second assertion.

Suppose that $|X, \leq_X| \leq |Y, \leq_Y|$ and $|Y, \leq_Y| \leq |X, \leq_X|$, and suppose that $f : X \to Y$ and $g : Y \to X$ are the strictly increasing mappings onto the whole set or an initial segment. By the preceding result, the composite **g f** is the identity mapping. If we can prove that **g** is onto, then the conclusion will follow because then **g** will be a $1 - 1$ onto order – preserving map and hence we have $|Y, \leq_Y| = |X, \leq_X|$. To verify that the mapping **g** is onto, let $x \in X$ be arbitrary and note that $g\,f = id_X$ yields $x = g(f(x))$.■

Ordinal numbers

Grammarians distinguish between two types of numbers in a language; namely, the ***cardinal numbers*** like ***one, two, three, …*** which we use to count objects, and the ***ordinal numbers*** like ***first, second, third, …*** which we use to order objects or concepts. Both notions of numbers are also present in set theory, and in fact Cantor introduced transfinite ordinal numbers before he introduced transfinite cardinal numbers.

In set theory, the relationship between ordinal and cardinal numbers is not quite the same as it is in ordinary language, but the fundamental pairing of cardinals with counting and ordinals with ordering carries over. We have seen that a cardinal number in mathematics in some sense corresponds to an equivalence class of sets in $1 - 1$ correspondence with each other. One way of describing an ordinal number in mathematics is that in some sense it corresponds to an equivalence class of well – ordered sets. More precisely, given two well – ordered sets $(A, <_A)$ and $(B, <_B)$, then we shall say that they have the same ***ordinal type*** (or represent the same ***ordinal number***) if there is a $1 - 1$ order preserving correspondence between them***; i.e.,*** there is a $1 - 1$ correspondence $f : A \to B$ that is *strictly increasing*: $x <_A y$ implies $f(x)$

$<_B$ **f(y)** for all **x** and **y** in **A**. It follows immediately that the inverse map $\mathbf{f^{-1} : B \to A}$ will also be strictly increasing in this case.

The simplest examples of well – ordered sets are given by subsets of the natural numbers; specifically, for each nonnegative integer **n** we can take the well – ordered set with **n** elements given by $\{0, \ldots, n - 1\}$ or we can take the entire set of natural numbers. Not surprisingly, the example with **n** elements is denoted by **n,** and following Cantor the well – ordered set given by the natural numbers is generally denoted by ω**.** However, there are also many other examples that one can construct from these. Perhaps the simplest one is the successor $\omega + 1$**,** which as before is given by the union

$$\omega \cup \{ \omega \}$$

with the original ordering on the elements of ω and the extra element ω as a unique maximal element. Of course, one can repeat this process and obtain a new successor set $\omega + 2 = (\omega + 1) + 1$**,** and this can be taken further to define a sequence of well – ordered sets $\omega + n$ for every positive integer **n.** In fact, there are standard, general arithmetic operations for constructing new well – ordered sets out of old ones. The discussions on pages 75 – 77 and 81 – 85 of Halmos and Sections 8.10 – 8.12 on pages 209 – 213 of Lipschutz provide both simple and complicated examples of how these constructions can be combined.

Aside from the successor construction taking a well – ordered set **X** to its successor set **X + 1,** we shall not need the arithmetic operations on well – ordered sets in these notes. However, the previously cited discussions in Halmos and Lipschutz imply the existence of many inequivalent well – ordered sets that are countably infinite, and of course it would be helpful to have some comprehensive means for keeping track of such objects.

The ***ordinal numbers*** will be a special class of well – ordered sets with the following crucial property: ***Every well – ordered set has the well – ordering type of a unique ordinal number.***

Originally Cantor attempted to define ordinal numbers using the previously mentioned approach with well – ordering types of well – ordered sets. However, the following definition due to J. von Neumann improves on Cantor's approach in several respects and has become the standard mathematical description for ordinal numbers (*e.g*., it is the formulation appearing page 75 of Halmos; in contrast, the formulation on page 208 of Lipschutz is essentially Cantor's definition)**:**

Definition. A set **S** is an ***ordinal*** if and only if **S** is well – ordered with respect to set membership and every element of **S** is also a subset of **S;** in other words, $\mathbf{x \in S}$ implies $\mathbf{x \subset S}$**.** The class of all ordinals (the ***ordinal numbers***) will often be denoted by Ω**;** the standard form of the Axiom of Specification (which is slightly different from the one in these notes) implies that Ω is a class**.** In Proposition 11 below shall prove that Ω cannot be a set (this is the *Burali – Forti Paradox* that we have previously mentioned).

The motivation for this definition arises from a ***standard model for the Peano axioms*** in which each nonnegative integer **n** corresponds to an explicit set with exactly **n** elements:

0 is represented by the empty set $S_0 = \emptyset$.

1 is represented by the one element set $S_1 = \{\emptyset\}$.

2 is represented by the two element set $S_2 = \{\emptyset, \{\emptyset\}\} = S_1 \cup \{S_1\}$

3 is represented by the two element set $S_3 = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\} = S_2 \cup \{S_2\}$

...

n is represented by the n element set $S_n = S_{n-1} \cup \{S_{n-1}\}$

Each of the sets S_n satisfies the definition of an ordinal, and the same is true of the union $S_\omega = \cup_n S_n$. Additional motivation for the definition is that if **S** is an ordinal, then the successor set $S + 1 = S \cup \{S\}$ is also an ordinal.

Proposition 9. *If* **S** *is an ordinal and* $x \in S$, *then* **x** *is also an ordinal.*

Proof. By the basic condition on ordinals, **x** is a subset of **S,** and therefore $y \in x$ implies $y \in S$. We need to show that **x** is well – ordered with respect to set membership and every element of **x** is also a subset of **x.** If **A** is a nonempty subset of **x,** then the definition of ordinal number implies that $A \subset x \subset S$, and therefore the set **A** has a least element with respect to set membership because **S** is well – ordered. Now suppose that $y \in x$; to show that $y \subset x$ we need to show that if $z \in y$ then $z \in x$. We claim that $z \in S$; if so, then all three of **x, y** and **z** lie in **S,** and since the ordinal **S** is linearly ordered by set membership we must have $z \in x$.

To prove that $z \in S$, note that $y \in S$ by the preceding paragraph, and since **S** is an ordinal it follows that **y** is a subset of **S,** so that $z \in S$ as required.■

Fundamental properties of ordinal numbers

The first result in this subsection might look as if it should be trivial, and it would be if we knew that the class of ordinals Ω was a set. However, at this point we do not know whether this is true (and in fact Proposition 11 below will show Ω is not a set).

Theorem 10. *If* Ω *denotes the ordinal numbers with the relation given by set membership, then every nonempty subset in* Ω *has a least element.*

Proof. Let **X** be a nonempty set of ordinals, and let $\alpha \in X$. Take **Y** to be the set of all $\beta \in Y$ such that $\beta \in \alpha$. If **Y** is empty then α is the least element of **X** because **X** is linearly ordered by set membership. If **Y** is nonempty, then **Y** is contained in α (using linear ordering again) and as such it has a least element. Thus we have found a least element in both cases.■

We have already noted that there is no "set of all ordinal numbers" just as there is no "set of all cardinal numbers." In fact, the paradox about ordinals was noticed by C.

Burali – Forti (1861 – 1931) a few years before Cantor discovered the analogous paradox about cardinal numbers.

Proposition 11. (Burali – Forti Paradox.) *The class* Ω *of ordinal numbers is not a set.*

Proof. Suppose that Ω is a set**.** We claim that it is an ordinal; since we have shown that it is well – ordered with respect to set – theoretic membership, it follows that the latter describes a well – ordering on Ω**.** To prove the second condition for an ordinal, suppose that $S \in \Omega$**;** we need to show that $S \subset \Omega$ or equivalently that $x \in S$ implies $x \in \Omega$**.** But this follows because every element of an ordinal is an ordinal. **Wreq erq** Weqr qwer Since Ω is an ordinal, it follows that $\Omega + 1$ is also an ordinal, and hence $\Omega +$ 1 is an element of Ω**.** By construction we have $\Omega \in \Omega + 1$, and since Ω is an ordinal it follows that $\Omega \in \Omega$**,** which contradicts the Axiom of Foundation. The problem arises from our assumption that Ω is a set, and therefore the latter must be false**.**■

The following basic fact has already been mentioned.

Theorem 12. (Classification of Well – Ordered sets). *Let* **X** *be a well ordered set. Then there is a unique* $\alpha \in \Omega$ *for which there is a* $1 - 1$ *order – preserving correspondence from* **X** *to* α*.*

Sketch of Proof. (See Goldrei, Theorem 8.2, pages 206 – 207, and Theorem 8.5, pages 212 – 213, for further details.) We start with existence. The idea is to construct a mapping from **X** to the ordinals by transfinite recursion such that for all $\beta \in X$**,** the function **f** maps **[0, β)** in **X** to **[0, f(β))** in Ω**.** Eventually this process terminates when one runs out of elements in **X.** Since Ω is not a set, there are elements of it that do not lie in the image of **f,** and if α is the first element not in the image of **f**, then the latter defines a $1 - 1$ order – preserving correspondence from **X** to α**. Reqw**

Uniqueness follows from our previous result that a well – ordered set cannot be in $1 - 1$ order – preserving correspondence with a proper subset of itself**.**■

The following existence result for least upper bounds is important for many purposes.

Theorem 13. *Let* **X** *be a nonempty set of ordinals. Then* **X** *has an upper bound (in the class of ordinals).*

Corollary 14. *In the above situation, the set* **X** *has a least upper bound.*

The corollary follows because the ordinals are well – ordered**.**■

Sketch of proof of upper bound theorem. (See Goldrei, Theorem 9.4, page 209 , or Halmos, the first four lines of page 80, for further details.) Let **$(X)** be the union of all ordinals in **X.** To complete the proof, it is necessary to show that **$(X)** is an ordinal and that it is an upper bound for all the ordinals in **X.** The second part uses the fact that two ordinals α and β satisfy the $\alpha \in \beta$ if and only if α is a proper subset of β**.** This fact is established in (solved) Exercise 8.6 on page 208 of Goldrei**.**■

Theorem 15. (Hartogs' Theorem.) *Given a set* **A,** *there is an ordinal* β *such that there is no* $\mathbf{1-1}$ *mapping from* β *into* **A.**

This result strongly suggests that for every set **A** there is an ordinal λ for which we have the cardinal number inequality $|A| \leq \lambda$. This will follow from the results of the next section, but the proof is considerably less trivial than it might seem at first; the problem involves proving the existence the $\mathbf{1-1}$ function from **A** to β whose existence may seem intuitively clear.

Notes. We have followed Goldrei in calling this result Hartogs' Theorem, but we must include a ***strong warning*** that usually the phrase "Hartogs' Theorem" refers to a major result in the theory of functions of several complex variables. Surprisingly, there is no biography of the German mathematician F. Hartogs (1874 – 1943) on the extensive MacTutor mathematical biography site mentioned in the first unit of these notes, but there is a biography (in German) at the following online site:

http://www.b.shuttle.de/b/pns/faecher/mathematik/Verfolgte/FHartogs.html

An expanded English translation of this biography (without many of the pictures) appears in the course directory:

http://math.ucr.edu/~res/math144/hartogsbio.pdf

Proof of Hartogs' Theorem on Ordinals. We shall only sketch the argument; the details appear in the proof of Theorem 8.19 of Goldrei on pages 224 – 225 of the latter.

The first crucial observation is that there is a set **U** of well – ordered sets such that if **W** is a well – ordered set supporting a $\mathbf{1-1}$ mapping into **A,** then **W** is in $\mathbf{1-1}$ order – preserving correspondence with some well – ordered set in **U.** To see this, note that every such **W** is in $\mathbf{1-1}$ order – preserving correspondence with a subset of **A** and thus the collection of such subsets with well – orderings is in $\mathbf{1-1}$ correspondence with a subset of the set $\mathbf{P(A) \times P(A \times A)}$.

Each well – ordered set corresponds to a unique ordinal number, so let **V** be the set of all ordinal numbers which correspond to the well – ordered sets in **U.** By Theorem 13 above we know that **V** has an upper bound, and of course there are also ordinals which are strictly larger than this upper bound. Every such ordinal fulfills the condition in the conclusion of the theorem, for each such ordinal is greater than all the ordinals that admit $\mathbf{1-1}$ mappings into **A.**■