Contents

Chapter 1

Some Elementary Informal Set Theory

Jan. 12, 2022

Set theory is due to Georg Cantor.

"Elementary" in the title above does not apply to the body of his work, since he went into considerable technical depth in this, his new theory.

It applies however to *our* coverage as we are going to restrict ourselves to elementary topics only.

Cantor made many technical mistakes in the process of developing set theory, some of considerable consequence. The next section is about the easiest to explain and most fundamental of his mistakes.

How come he made mistakes?

The reason is that his theory was not based on axioms and rigid rules of reasoning —a state of affairs for a theory that we loosely characterise as "informal".

At the opposite end of informal we have the *formal* theories that are based on axioms *and* logic and are thus "safer" to develop (they do not lead to *obvious* contradictions).

One *cannot* fault Cantor for not using logic in arguing his theorems —that process was not invented when he built his theory— but then, *a fortiori*, mathematical logic was not invented in Euclid's time either, *and yet* he did use axioms that stated how his building blocks, *points*, *lines* and *planes* interacted and behaved!

Guess what: Euclidean Geometry leads to no contradictions.

The problem with Cantor's set theory is that anything goes as to what sets *are* and *how they come about*.

He neglected to ask the most fundamental question:

"How are sets formed?"[†] He just sidestepped this and simply said that a *set* is *any collection*.

In fact he took the term "set" as just a synonym for "collection", "class", "aggregate", etc. What you just get from a Dictionary: synonyms!

Failure to ask and answer this question leads to "trouble", which is the subject matter of the next section.

One can still do "safe" set theory —devoid of "trouble", that is— within an *informal* (non axiomatic) setting, but

> we have to ask and answer how sets are built *first* and derive from our answer some *principles* that will guide (and protect!) the theory's development!

We will do so.

[†] It's amazing how much trouble could be avoided if he had asked AND investigated!

1.1. Russell's "Paradox"

Cantor's *naïve* (this adjective is not derogatory but is synonymous in the literature with *informal* and *non axiomatic*) set theory was plagued by *paradoxes*, the most famous of which (and the *least* "technical") being pointed out by Bertrand Russell and thus nicknamed "Russell's paradox".[†]

[†] From the Greek word "paradoxo" (παράδοξο) meaning against one's belief or knowledge; a contradiction.

His theory is the theory of collections (i.e., sets) of objects, as we mentioned above, terms that were neither defined *nor was it said* how they were built.[†]

This theory studies operations on sets, properties of sets, and aims to use set theory as the foundation *of all mathematics*. Naturally, mathematicians "do" set theory of *mathematical object collections* — not collections of birds and other beasts.

[†] This is not a problem *in itself*. Euclid too did not say *what* points and lines *were*; but his axioms did characterise their nature and interrelationships: For example, he started from these (among a few others) *a priori truths* (axioms): *a unique line passes through two distinct points*; also, *on any plane, a unique line l can be drawn parallel to another line k on the plane if we want l to pass through a given point A that is not on k.*

The point is:

You cannot leave out *both* what the nature of your objects is and *how* they behave/interrelate and get away with it! Euclid omitted the former but provided the latter, so all worked out.

We have learnt some elementary aspects of set theory at high school. We will learn more in this course.

1. **Variables**. Like any theory, informal or not, informal set theory —a safe variety of which we will develop here— uses *variables* just as algebra does. There is only *one type* of variable that varies over *set* and over *atomic objects* too, the latter being objects that have no set structure. For example integers. We use the names $A, B, C, \ldots$ and $a, b, c, \ldots$ for such variables, sometimes with primes (e.g., A'') or subscripts (e.g., x_{23}), or both (e.g., x'''_{22}, Y'_{42}).

2. **Notation**. *Sets given by listing.* For example, $\{1, 2\}$ is a set that contains precisely the objects 1 and 2, while

$$\{\overbrace{\ 1\ }^{\text{atom}}, \overbrace{\{5, 6\}}^{\text{set}}\}$$

is a set that contains precisely the objects 1 and $\{5, 6\}$. The braces $\{$ and $\}$ are used to show the collection/set by outright listing.

3. **Notation**. *Sets given by "defining property"*. But what if we cannot (or will not) explicitly list all the members of a set? Then we may define what objects x get in the set/collection by having them to *pass an entrance requirement*, $P(x)$:

   > **An object x gets in the set *iff* (*if and only if*) $P(x)$ is true of said object.**

   Let us parse "iff":

   (a) The *IF*: So, IF $P(x)$ is true, then x gets in the set (it passed the "admission requirement").

   (b) The *ONLY IF*: So, IF x gets in the set, then the ***only** way for this to happen* is for it to pass the "admission requirement"; that is, $P(x)$ is true.

   In other words, "iff" (as we probably learnt in high school or some previous university course such as calculus) is the same thing as "is equivalent":

   "x is in the set" is equivalent to "$P(x)$ is true".

We denote the collection/set[†] defined by the entrance condition $P(x)$ by

$$\{x : P(x)\} \tag{1}$$

but also as

$$\{x \,|\, P(x)\} \tag{1'}$$

reading it "the set of *all* x *such that* (this "such that" is the ":" or "|") $P(x)$ is true [or holds]"

[†] We have not yet reached Russell's result, so keeping an open mind and humouring Cantor we still allow ourselves to call said collection a "set".

4. "$x \in A$" is the assertion that "object x is in the set A". Of course, this assertion may be true or false or "it depends", just like the assertions of algebra $2 = 2$, $3 = 2$ and $x = y$ are so (respectively).

5. $x \notin A$ is the negation of the assertion $x \in A$.

6. **Properties**

- Sets are *named* by letters of the Latin alphabet (cf. **Variables**, above). *Naming is pervasive in mathematics* as in, e.g., "let $x = 5$" in algebra.

  So we can write "let $A = \{1, 2\}$" and let "$c = \{1, \{1, 5, 6\}\}$" to give the names A and c to the two example sets above, ostensibly because we are going to discuss these sets, and refer to them often, and it is cumbersome to keep writing things like $\{1, \{5, 6\}\}$. Names are *not permanent*;[†] they are *local* to a discussion (argument).

[†] OK, there *are* exceptions: $\emptyset$ is the permanent name for the *empty set* —the set with no elements at all— and for that set only; $\mathbb{N}$ is the permanent name of the set of all *natural numbers*.

- **Equality of sets** (repetition and permutation do not matter!) Two sets A and B are equal iff they have the same members. Thus order and multiplicity do not matter! E.g., $\{1\} = \{1, 1, 1\}$, $\{1, 2, 1\} = \{2, 1, 1, 1, 1, 2\}$.

- Here is *the fundamental equivalence pertaining to definition of sets by "defining property"*:

So, if we name the set in (1) above (p.13), S, that is, if we say "let $S = \{x : P(x)\}$", then "$x \in S$ iff $P(x)$ is true"

☡ By the way, we almost *never say* "is true" unless we want to shout out this fact. We would say instead:

$$x \in S \text{ iff } P(x) \tag{†}$$

Equipped with the knowledge of the previous bullet, we see that the symbol $\{x : P(x)\}$ defines a *unique* set/collection: Well, say A and B are so defined, that is, $A = \{x : P(x)\}$ and $B = \{x : P(x)\}$. Thus

$$x \in A \overset{A=\{x:P(x)\}}{\text{iff}} P(x) \overset{B=\{x:P(x)\}}{\text{iff}} x \in B$$

thus

$$x \in A \text{ iff } x \in B$$

and thus $A = B$. ☡

Let us pursue, as Russell did, the point made in the last bullet above. Take $P(x)$ to be specifically the assertion $x \notin x$. He then gave a name to

$$\{x : x \notin x\}$$

say, R. But then, by the last bullet above, in particular, the equivalence (†),

$$x \in R \text{ iff } x \notin x \tag{2}$$

If we now *believe*,[†] as *Cantor*, the father of set theory did not question and went ahead with it, that every $P(x)$ defines a *set*, *then R is a set.*

☡ What is wrong with that? ☡

Well, if R is a set then this object has the proper *type* to be plugged into the *variable of type "math object"*, namely, x, throughout the equivalence (2) above. But this yields the contradiction

$$R \in R \text{ iff } R \notin R \tag{3}$$

This contradiction is called the Russell's Paradox.

[†] Informal mathematics often relies on "I know so" or "I believe" or "it is 'obviously' true". Some people call "proofs" like this —i.e., "proofs" without justification(s)— "proofs by intimidation". Nowadays, with the ubiquitousness of the qualifier "fake", one could also call them "fake proofs".

This and similar paradoxes motivated mathematicians to develop formal symbolic logic and look to axiomatic set theory[†] as a means to avoid paradoxes like the above.

Other mathematicians who did not care to use mathematical logic and axiomatic theories found a way to do set theory *informally*, yet *safely*.

See, they asked *and* answered "how are sets formed?"[‡]
Read on!

[†] There are many flavours or axiomatisations of set theory, the most frequently used being the "ZF" set theory, due to Zermelo and Fraenkel.

[‡] Actually, axiomatic set theory —in particular, its axioms are— is built upon the answers this group came up with. This story is told at an advanced level in [Tou03b].

Chapter 2

Safe Set Theory

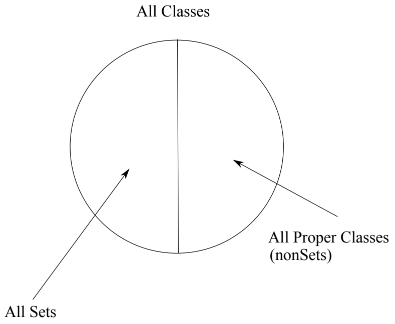
Jan. 14, 2022

So, *some* collections of sets and/or atoms are *not* —technically— sets, as the Russell Paradox taught us! How do we tell them apart?

From now on we will deal with collections that *may or may not* be sets, with a promise of learning how to create sets if we want to!

The modern literature uses the terminology “**class**” for any such collection of sets and/or atoms (and uses the term “collection” non technically and sparsely).

The above is captured by the following picture:



2.0.1 Definition. (Classes and sets)
From now on we call *all* collections **classes**.

Definitions by defining property "Let $\mathbb{A} = \{x : P(x)\}$" **always** defines a **class**, but as we saw, sometimes —e.g., if "$P(x)$" is specifically "$x \notin x$"— that class is *not* a set (Section 1.1).

x is a set/atom-type variable!

Classes that are not sets are called **proper classes**. We will normally use what is known as "blackboard bold" notation and capital latin letters to denote classes by names such as $\mathbb{A}, \mathbb{B}, \mathbb{X}$. If we determine that some class $\mathbb{A}$ *is* a set, we would rather write it as A, but we make an *exception* for the following **sets**:

Mathematicians use notation and results from set theory in their everyday practice. We call the sets that mathematicians use the "real sets" of our mathematical *intuition*, like the set of natural numbers, $\mathbb{N}$ (also denoted by ω), integers $\mathbb{Z}$, rationals $\mathbb{Q}$ and reals $\mathbb{R}$. □

2.0.2 Example. Thus if R is the Russel (proper) class, then the configuration $\{R\}$ is not allowed —it is meaningless.

Of course Cantor would not care and allow this and even this

$$\{\{\{R\}\}, R\}$$

□

☡ In forming the class $\{x : P(x)\}$ for any property $P(x)$ we say that we apply *comprehension*. It was the Frege/Cantor belief (explicitly or implicitly) that comprehension was *safe* —i.e., they believed that $\{x : P(x)\}$ always was a set. Russell proved that it was not. ☡

It is known that set theory, using as primitives the notions of *set*, *atom* (an object that is not sub-divisible; not a collection of objects), and the relation *belongs to* ($\in$), is sufficiently strong to serve as *the foundation of all mathematics*.

Mathematicians use notation and results from set theory in their everyday practice. We call the sets that mathematicians use the "real sets" of our mathematical *intuition*, like the set of natural numbers, $\mathbb{N}$ (also denoted by *omega*), integers $\mathbb{Z}$, rationals $\mathbb{Q}$ and reals $\mathbb{R}$.

2.1. The "real sets" —Introduction to Stages

So, how can we tell, or indeed *guarantee*, that a certain *class* is a *set*?

Russell proposed this "recovery" from his Paradox:

Make sure that sets are built by stages, where at stage 0 all atoms are available.

Atoms are also called *urelements* in the literature, from the German *Urelemente*, which in analogy with the word "*urtext*" —meaning *the earliest text*— would mean that they are the "earliest" mathematical objects. Witness that they are available at stage 0!

We may then collect atoms to form all sorts of "first level" *sets*. We may proceed to collect any mix of atoms and first-level sets to build new collections —second-level sets— *and so on*.

Much of what set theory does is attempting to remove the ambiguity from this "and so on". See below, **Principles** 0–2.

Thus, at the beginning we have all the level-0, or type-0, objects available to us. For example, *atoms* such as $1, 2, 13, \sqrt{2}$ are available.

At the next level we can include any number of such atoms (from none at all, to all) to **build a set**, that is, a new mathematical object.

Allowing the usual notation, i.e., listing of what is included within braces, we may cite a few examples of level-1 sets:

L1-1. $\{1\}$.

L1-2. $\{1, 1\}$.

L1-3. $\{1, \sqrt{2}\}$.

L1-4. $\{\sqrt{2}, 1\}$.

L1-5. $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$.

We already can identify a level-2 object, using what (we already know) *is* available:

L2-1. $\{\{\sqrt{2}, 1\}, 42\}$.

Note how the level of nesting of $\{\ \}$-brackets matches the level or stage of the formation of these objects!

2.1.1 Definition. (Class and set *equality*) This definition applies to any classes, hence, in particular, to any *sets* as well.

Two classes $\mathbb{A}$ and $\mathbb{B}$ are *equal* —written $\mathbb{A} = \mathbb{B}$— means

$$x \in \mathbb{A} \text{ iff } x \in \mathbb{B} \tag{1}$$

That is, an object is in $\mathbb{A}$ iff it is also in $\mathbb{B}$.

$\mathbb{A}$ is a *subclass* of $\mathbb{B}$ —written $\mathbb{A} \subseteq \mathbb{B}$— means that every element of the first class occurs also in the second, or

$$\text{If } x \in \mathbb{A}, \text{ then } x \in \mathbb{B} \tag{2}$$

If $\mathbb{A}$ is a *set*, then we say it is a *subset* of $\mathbb{B}$.

If we have $\mathbb{A} \subseteq \mathbb{B}$ but $\mathbb{A} \neq \mathbb{B}$, then we write $\mathbb{A} \underset{\neq}{\subset} \mathbb{B}$ (some of the literature uses $\mathbb{A} \subsetneq \mathbb{B}$ or even $\mathbb{A} \subset \mathbb{B}$ instead) and say that $\mathbb{A}$ is a *proper subclass* of $\mathbb{B}$.

Caution. In the terminology "*proper subclass*" the "proper" refers to the fact that $\mathbb{A}$ is not all of $\mathbb{B}$. It does *NOT* say that $\mathbb{A}$ is not a set! It *may* be a set and then we say that it is "*proper subset*" of $\mathbb{B}$. □

If n is an integer-valued variable, then what do you understand by “$2n$ is even”?

The normal understanding is that “no matter what the value of n is, $2n$ is even”, or “for all values of n, $2n$ is even”.

When we get into our logic topic in the course we will see that we *can* write “for all values of n, $2n$ is even” with less English as “$(\forall n)(2n$ is even)”. So “$(\forall n)$” says “for all (values of) n”.

Mathematicians often prefer to have statements like “$2n$ is even” with the “for all” *always implied.*[†] You can write a whole math book without writing $\forall$ even once, and without overdoing the English.

Thus in (1) and (2) above the “forall x” is implied.

2.1.2 Remark. Since “iff” between two statements S_1 and S_2 means that we have *both* directions

$$\text{If } S_1, \text{ then } S_2$$

and

$$\text{If } S_2, \text{ then } S_1$$

we have that “$\mathbb{A} = \mathbb{B}$” is the same as (equivalent to) “$\mathbb{A} \subseteq \mathbb{B}$ and $\mathbb{B} \subseteq \mathbb{A}$”. □

[†] An exception occurs in Induction that we will study later, where you *fix* an n (but keep it as a variable, not as 5 or 42) and assume the “induction hypothesis” $P(n)$. But do not worry about this now!

2.1.3 Example. In the context of the "$\mathbb{A} = \{x : P(x)\}$" notation we should remark that notation-by-listing can be simulated by notation-by-defining-property: For example, $\{a\} = \{x : x = a\}$ —here "$P(x)$" is $x = a$.

Also $\{A, B\} = \{x : x = A \text{ or } x = B\}$. Let us verify the latter: Say $x \in$ lhs.[†] Then $x = A$ or $x = B$. But then the entrance requirement of the rhs[‡] is met, so $x \in$ rhs.

Conversely, say $x \in$ rhs. Then the entrance requirement is met so we have (at least) one of $x = A$ or $x = B$ ("true" implied) Trivially, in the first case $x \in$ lhs and ditto for the second case. □

[†] Left Hand Side.

[‡] Right Hand Side.

Jan. 17, 2022

We now postulate the principles of formation of sets!

Principle 0. Sets and atoms are *the mathematical objects* of our (safe) set theory.

Sets are formed by stages. At stage 0 we have (and acknowledge) the *presence* of atoms. *They are given outright, they are not built.*

At *any* stage Σ we *are allowed to* build a *set*, collecting together other *mathematical objects* (sets or atoms) *provided* these (mathematical) objects we put into our set *were available at stages before* Σ.

Principle 1. *Every set is built at some stage.* *A set does not just happen!*

Principle 2. If Σ is a stage of set construction, then *there is* a stage Φ *after* it.

Principle 2 makes clear that we have *infinitely many* stages of set formation in our toolbox.

"Clear"?

Can you argue that informally? (**Exercise**! *Hint.* Combine Property 2 statement with a "what if": *What if there are only finitely many stages?* and go for a contradiction from the what if.)

Incidentally the property of a stage being "before" another is exactly like "$<$" on the integers:

1. For any two integers n, m the statement "$n = m$ or $n < m$ or $m < n$" is true.
2. We cannot have $n < n$, for any n (this is the "irreflexivity" of "$<$").
3. If we have $n < m$ and $m < r$, then $n < r$ (this is the "transitivity" of "$<$").

For stages,

Using "$<$" as short for "lhs comes *before* rhs", then

$1'$. For any two stages Σ and Σ' the statement "$\Sigma = \Sigma'$ or $\Sigma < \Sigma'$ or $\Sigma' < \Sigma$" is true.

$2'$. We cannot have Σ is before (or after) Σ, for any Σ.

$3'$. If we have $\Sigma < \Sigma'$ and $\Sigma' < \Sigma''$, then $\Sigma < \Sigma''$.

2.1.4 Remark. If some set is definable (“buildable”) at some stage Σ, then it is also definable at any later stage as well, as **Principle** 0 makes clear.

> The informal set-formation-by-stages Principle will guide us to build, safely, all the sets we may need in order to do mathematics.

□

2.2. What caused Russell's paradox

How would the set-building-by-stages doctrine avoid Russell's paradox?

☡ Recall that *à la Cantor* we get a paradox (contradiction) because we *insisted to believe* that all classes are sets, that is, following Cantor we "believed" Russell's "R" was a *set*. ☡

Principles 0–2 allow us to know *a priori* that R is a proper class. No contradiction is obtained!

How so?

OK, is $x \in x$ true or false? Is there *any* mathematical object x —say, A— for which it *is* true?

$$A \in A? \tag{1}$$

1. Well, for atom A, (1) is false since atoms have no set structure, that is, are not collections of objects. An atom A *cannot contain anything*, in particular it cannot contain A.

2. What if A is a set and $A \in A$? Then in order to build A, the *set* on the rhs, we have to wait until *after* its member, A —the set on the lhs— is built (Principle 0). So, we need (the left) A to be built *before* (the right) A in (1).

   Absurd!

So (1) is **false**. A being arbitrary, we have just demonstrated that

$$x \in x \text{ is false}$$

thus $x \notin x$ is true (*for all* x), therefore R of Section 1.1 is $\mathbb{U}$, the universe of *all sets and atoms*.

$$R = \mathbb{U}$$

So?

Well,

So here is why $\mathbb{U}$ —that is R— is *not* a set. Well, if it is

- $\mathbb{U} \in \mathbb{U}$ since the rhs contains *all sets* and we believe the lhs to be a *set*.
- but we just saw that the above is false if $\mathbb{U}$ is a *set*!

So $\mathbb{U}$, aka R, is a *proper* class. Thus, the fact that R is not a set is neither a surprise, nor paradoxical. It is just a proper class as we just have recognised.

BTW,

A class $\mathbb{A}$ is proper iff we have NO stage left to build it (Principles 0 and 1).

Intuitively then if we ran out of stages building $\mathbb{A}$ it means that *there are are far too many elements in* $\mathbb{A}$ —this class is "enormous", as indeed is $\mathbb{U}$.

Often the informal (and sloppy) literature on sets will blame “size” for a class failing to be a set. That is dangerous. Lack of set status must be connected with lack of stage at which to build said class as a set.

Incidentally not all “LARGE” classes contain “everything”. We will see later that if we remove ALL atoms from $\mathbb{U}$, then what remains is a proper class too. So is $\mathbb{S} = \{\{x\} : x \in \mathbb{U}\}$: The class of *all 1-element sets*.

2.3. Some useful sets

2.3.1 Example. (Pair) By Principles 0, 1, if A and B are sets or atoms, then let A be available at stage Σ and B at stage Σ'. Without loss of generality say Σ' is not later than (after) Σ.

For two stages Φ and Ψ we can write $\Phi \le \Psi$ as short for $\Phi < \Psi$ or $\Phi = \Psi$.

Pick a stage Σ'' *after* Σ (Principle 2). This will be be after both (cf. Principle 2 and the ☡-remark there, on p.30) Σ, Σ'. So,

$$\Sigma' \le \Sigma < \Sigma''$$

At stage Σ'' we can build

$$\{A, B\} \tag{1}$$

as a *set* (because both A and B are available! cf. Principle 0).

We call (1) the (unordered) *pair set*.

Pause. Why "unordered"? See 2.1.1.◀ □

We have just proved a theorem above:

2.3.2 Theorem. *If A, B are sets or atoms, then $\{A, B\}$ is a set.*

2.3.3 Exercise. Without referring to stages in your proof, prove that if A is a set or atom, then $\{A\}$ is a set. $\square$

Jan. 19, 2022

2.3.4 Remark. A very short digression into Boolean Logic — for now. It will be convenient —but not necessary; we are doing fine so far— to use *truth tables* to handle many simple situations that we will encounter where “logical connectives” such as “*not*”, “*and*”, “*or*”, “*implies*” and “*is equivalent*” enter into our arguments.

We will put on record here how to *compute* things such as “S_1 and S_2”, “S_1 implies S_2”, etc., where S_1 and S_2 stand for two arbitrary statements of mathematics. In the process we will introduce the *mathematical symbols* for “and”, “implies”, etc.

The *symbol translation table* from English to symbol is:

NOT	$\neg$
AND	$\wedge$
OR	$\vee$
IMPLIES (IF...,THEN)	$\rightarrow$
IS EQUIVALENT	$\equiv$

The truth table below has a simple reading. For *all possible* truth values —true/false, for short **t**/**f**— of the “simpler” statements S_1 and S_2 we indicate the *computed truth value* of the compound (or “more complex)” statement that we obtain when we *apply* one or the other Boolean connective of the previous table to S_1 and S_2.

S_1	S_2	$\neg S_1$	$S_1 \wedge S_2$	$S_1 \vee S_2$	$S_1 \rightarrow S_2$	$S_1 \equiv S_2$	$S_2 \rightarrow S_1$
f	**f**	**t**	**f**	**f**	**t**	**t**	**t**
f	**t**	**t**	**f**	**t**	**t**	**f**	**f**
t	**f**	**f**	**f**	**t**	**f**	**f**	**t**
t	**t**	**f**	**t**	**t**	**t**	**t**	**t**

Comment. All the computations of truth values *satisfy our intuition*, except perhaps that for "$\rightarrow$": Indeed, $\neg$ flips the truth value as it should, $\wedge$ is eminently consistent with common sense, $\vee$ is the "inclusive or" of the mathematician, and $\equiv$ is just equality on the set $\{\mathbf{f}, \mathbf{t}\}$, as it should be.

The "problem" with $\rightarrow$ is that there is no *causality* from left to right. The only "sane" entry is for $\mathbf{t} \rightarrow \mathbf{f}$. The outcome should be false for a "bad implication"[†] and so it is. But look at it this way:

- Looking at $\rightarrow$ in the "red columns" see how the given table for $\rightarrow$ is eminently consistent with that for $\equiv$. Intuitively $\equiv$ means $\rightarrow$ from left to right AND $\rightarrow$ from right to left. It IS!
- This version of $\rightarrow$ goes way back to Aristotle. It is the version used by the vast majority of practising mathematicians and is nicknamed "material implication" or "classical implication".

[†] A bad implication has a true premise but a false conclusion. A correct implication ought to preserve truth!

Practical considerations. Thus

1. if you want to demonstrate that $S_1 \vee S_2$ is true, for any component statements S_1, S_2, then show that *at least one* of the S_1 and S_2 is true.
2. If you want to demonstrate that $S_1 \wedge S_2$ is true, then show that *both* of the S_1 and S_2 are true.

   Note, incidentally, the if we *know* that $S_1 \wedge S_2$ is true, then the truth table *guarantees* that each of S_1 and S_2 *must* be true.

3. If now you want to show the implication $S_1 \to S_2$ is true, **then the only real work is to show that if we assume S_1 is true, then S_2 is true too**.

   If S_1 is known to be false, then no work is required to prove the implication because of the first two lines of the truth table!!

4. If you want to show $S_1 \equiv S_2$, then —because the last three columns show that this is *computed* with *the same result* as $\big(S_1 \to S_2\big) \wedge \big(S_2 \to S_1\big)$— it follows that you just have to *compute* and "show" that **each** of the two implications $S_1 \to S_2$ and $S_2 \to S_1$ is true.

An important variant of $\rightarrow$ and $\equiv$

Pay attention to this point since almost everybody gets it wrong! In the literature and in the interest of creating a usable shorthand many practitioners of mathematical writing use sloppy notation

$$S_1 \rightarrow S_2 \rightarrow S_3 \tag{1}$$

attempting to convey the meaning

$$(S_1 \rightarrow S_2) \wedge (S_2 \rightarrow S_3) \tag{2}$$

Alas, $\underline{\text{(2) is not the same as (1)}}$! But what about $a < b < c$ for $a < b \wedge b < c$? That is wrong too!

Back to $\rightarrow$-chains like (1) vs. chains like (2):

Take S_1 to be $\mathbf{t}$ (true), S_2 to be $\mathbf{f}$ and S_3 to be $\mathbf{t}$. Then (1) is true because in a chain using the same Boolean connective *we put brackets from right to left*: (1) is $S_1 \rightarrow (S_2 \rightarrow S_3)$ and evaluates to $\mathbf{t}$, while (2) evaluates clearly to false ($\mathbf{f}$) since $S_1 \rightarrow S_2 = \mathbf{f}$ and $S_2 \rightarrow S_3 = \mathbf{t}$.

So we need a special symbol to denote (2) "economically". We need a *conjunctional implies*! Most people use $\Longrightarrow$ for that:

$$S_1 \Longrightarrow S_2 \Longrightarrow S_3 \tag{3}$$

that means, by **definition**, (2) above.

Similarly,

$$S_1 \equiv S_2 \equiv S_3 \tag{4}$$

is **NOT** conjunctional. It is **not** two equivalences —two statements— connected by an *implied* "$\wedge$", rather it says

$$S_1 \equiv (S_2 \equiv S_3)$$

ONE formula, ONE statement.

Now if $S_1 = \mathbf{f}$, $S_2 = \mathbf{f}$ and $S_3 = \mathbf{t}$, then (4) evaluates as $\mathbf{t}$ but the conjunctional version

$$(S_1 \equiv S_2) \wedge (S_2 \equiv S_3) \tag{5}$$

evaluates as $\mathbf{f}$ since the second side of $\wedge$ is $\mathbf{f}$.

So how do we denote (5) correctly without repeating the consecutive S_2's and omitting the implied "$\wedge$"? This way:

$$S_1 \Longleftrightarrow S_2 \Longleftrightarrow S_3 \tag{4}$$

By definition, "$\Longleftrightarrow$" —just like "iff"— is conjunctional: It applies to two statements —S_i and S_{i+1}— only and implies an $\wedge$ before the adjoining next similar equivalence. □ ☡

2.3.5 Theorem. (The subclass theorem) *Let $\mathbb{A} \subseteq B$ (B a set). Then $\mathbb{A}$ is a set.*

Proof. Well, B being a set there is a stage Σ where it is built (Principle 1). By Principle 0, all members of B are *available or built* before *stage* Σ.

But by $\mathbb{A} \subseteq B$, all the members of $\mathbb{A}$ are among those of B.

Hey! By Principle 0 we can build $\mathbb{A}$ at stage Σ, so *it is a set.* □

Some corollaries are useful:

2.3.6 Corollary. (Modified comprehension I) *If for all x we have*

$$P(x) \rightarrow x \in A \tag{1}$$

for some set A, then $\mathbb{B} = \{x : P(x)\}$ is a set.

Proof. I will show that $\mathbb{B} \subseteq A$, that is,

$$x \in \mathbb{B} \rightarrow x \in A$$

Indeed (see 3 under **Practical considerations** in 2.3.4), let $x \in \mathbb{B}$. Then $P(x)$ is true, hence $x \in A$ by (1). Now invoke 2.3.5. □

2.3.7 Corollary. (Modified comprehension II) *If A is a set, then so is $\mathbb{B} = \{x : x \in A \wedge P(x)\}$ for any property $P(x)$.*

Proof. The defining property here is "$(x \in A)$[†] $\wedge P(x)$". This implies $x \in A$ —by 2 in 2.3.4— that is, we have

$$(x \in A \wedge P(x)) \rightarrow x \in A$$

Now invoke 2.3.6. □

[†] Brackets not needed; inserted for extra clarity.

2.3.8 Remark. **(*The* empty set)** The class $\mathbb{E} = \{x : x \neq x\}$ has no members at all; it is empty. Why? Because

$$x \in \mathbb{E} \equiv x \neq x$$

but the condition $x \neq x$ is *always false*, therefore *so is the statement*

$$x \in \mathbb{E} \tag{1}$$

We do not collect anything into $\mathbb{E}$. Is the class $\mathbb{E}$ a set?

Well, take $A = \{1\}$. This is a set as the atom 1 is given at stage 0, and thus we can construct the *set* A at stage 1.

Note that, by (1) and 3 in 2.3.4 we have that

$$x \in \mathbb{E} \to x \in \{1\}$$

is true (for all x). That is, $\mathbb{E} \subseteq \{1\}$.

By 2.3.5, $\mathbb{E}$ *is a set.*

But is it *unique* so we can justify the use of the definite article "the"? Yes. The specification of *an empty set is a class with no members*. So if D is another empty set, then we will *also* have $x \in D$ always *false*. But then

$$x \in \mathbb{E} \equiv x \in D \text{ (both sides of } \equiv \text{ are false)}$$

and we have $\mathbb{E} = D$ by 2.1.1.

The unique *empty set is denoted by the symbol $\emptyset$ in the literature.* **Never** use "{}" for the empty set. Incorrect notation! □

2.4. Operations on classes and sets

Jan. 21, 2022

The reader probably has seen before (perhaps in calculus) the operations on sets denoted by $\cap, \cup, -$ and others. We will look into them in this section.

2.4.1 Definition. (Intersection of two classes) We define for any classes $\mathbb{A}$ and $\mathbb{B}$

$$\mathbb{A} \cap \mathbb{B} \stackrel{Def}{=} \Big\{x : x \in \mathbb{A} \wedge x \in \mathbb{B}\Big\} \tag{1}$$

We call the operator $\cap$ *intersection* and the result $\mathbb{A} \cap \mathbb{B}$ the intersection of $\mathbb{A}$ and $\mathbb{B}$.

If $\mathbb{A} \cap \mathbb{B} = \emptyset$ —which happens precisely when the two classes have no common elements— we call the classes *disjoint*.

Taking liberties with notation (of definition by defining property) we may write instead of (1) either

$$\mathbb{A} \cap \mathbb{B} \stackrel{Def}{=} \Big\{x \in \mathbb{A} : x \in \mathbb{B}\Big\} \tag{1'}$$

or

$$\mathbb{A} \cap \mathbb{B} \stackrel{Def}{=} \Big\{x \in \mathbb{B} : x \in \mathbb{A}\Big\} \tag{1''}$$

It is meaningless to have $\cap$ operate on atoms.[†] □

We have the easy theorem below:

[†] The definition expects $\cap$ to *operate on classes*. As we know, atoms (by definition) *have no set/class structure* thus no class and no set is an atom.

2.4.2 Theorem. *If B is a set, as its notation suggests, then $\mathbb{A} \cap B$ is a set.*

Proof. I will prove $\mathbb{A} \cap B \subseteq B$ which will rest the case by 2.3.5. So, I want

$$x \in \mathbb{A} \cap B \rightarrow x \in B$$

To this end, let then $x \in \mathbb{A} \cap B$ (cf. 3 in 2.3.4). This says that $x \in \mathbb{A} \wedge x \in B$ is true, so $x \in B$ is true. □

2.4.3 Corollary. *For sets A and B, $A \cap B$ is a set.*

2.4.4 Definition. (Union of two classes) We define for any classes $\mathbb{A}$ and $\mathbb{B}$

$$\mathbb{A} \cup \mathbb{B} \overset{Def}{=} \Big\{x : x \in \mathbb{A} \vee x \in \mathbb{B}\Big\}$$

We call the operator $\cup$ *union* and the result $\mathbb{A} \cup \mathbb{B}$ the union of $\mathbb{A}$ and $\mathbb{B}$.

It is meaningless to have $\cup$ operate on atoms. □

2.4.5 Theorem. *For any sets A and B, $A \cup B$ is a set.*

Proof. By assumption say A is built at stage Σ while B is built at stage Σ'. Without loss of generality (in short, "wlg") say Σ is no later than Σ', that is, $\Sigma \le \Sigma'$.

By Principle 2, I can pick a stage $\Sigma'' > \Sigma'$, thus (transitivity of "later")

$$\Sigma'' > \Sigma' \tag{1}$$

and

$$\Sigma'' > \Sigma \tag{2}$$

Lets us pick any item $x \in A \cup B$:

I have two (not necessarily mutually exclusive) cases (by 2.4.4):

- $x \in A$. Then x was available or built[†] at a stage $< \Sigma$,

$$\text{hence, by (2), } \underline{x \text{ is available } \textit{before } \Sigma''} \tag{3}$$

- $x \in B$. Then x was available or built at a stage $< \Sigma'$,

$$\text{hence, by (1), } \underline{x \text{ is available } \textit{before } \Sigma''} \tag{4}$$

In either case, (3) or (4), the arbitrary x from $A \cup B$ is available before Σ'', so we can collect all those x-values at stage Σ'' in order to form a *set*: $A \cup B$. □

[†] As x may be an atom, we allow the *possibility* that it was available *with no building involved*, hence we said "available or built". For A and B though we are told they are *sets*, so they *were* built at some stage, by Principle 1!

2.4.6 Definition. (Difference of two classes) We define for any classes $\mathbb{A}$ and $\mathbb{B}$

$$\mathbb{A} - \mathbb{B} \overset{Def}{=} \Big\{ x : x \in \mathbb{A} \land x \notin \mathbb{B} \Big\} \tag{1}$$

We call the operator "$-$" *difference* and the result $\mathbb{A} - \mathbb{B}$ the difference of $\mathbb{A}$ and $\mathbb{B}$, in that order.

It is meaningless to have "$-$" operate on atoms. □

Notation. As was the case for $\cap$ (Definition 2.4.1) for "$-$" too we have a shorter alternative notation to (1) above:

$$\mathbb{A} - \mathbb{B} \overset{Def}{=} \Big\{ x \in \mathbb{A} : x \notin \mathbb{B} \Big\}$$

2.4.7 Theorem. *For any* set A *and class* $\mathbb{B}$, $A - \mathbb{B}$ *is a set.*

Proof. The reader is asked to verify that $A - \mathbb{B} \subseteq A$. We are done by 2.3.5. □

2.4.8 Exercise. Prove that $\{\mathbb{Z}\}$ is a set, where $\mathbb{Z}$ is the set of integers $\{\ldots,-1,0,1,\ldots\}$. □

2.4.9 Exercise. Demonstrate —using Definition 2.4.1— that for any $\mathbb{A}$ and $\mathbb{B}$ we have $\mathbb{A}\cap\mathbb{B}=\mathbb{B}\cap\mathbb{A}$. □

2.4.10 Exercise. Demonstrate —using Definition 2.4.4— that for any $\mathbb{A}$ and $\mathbb{B}$ we have $\mathbb{A}\cup\mathbb{B}=\mathbb{B}\cup\mathbb{A}$. □

2.4.11 Exercise. By picking two particular very small sets A and B show that $A-B=B-A$ is not true for all sets A and B.

Is it true of all classes? □

Let us generalise unions and intersections next. First a definition:

2.4.12 Definition. (Family of sets) A class $\mathbb{F}$ is called a *family of sets* iff it contains no atoms. The letter $\mathbb{F}$ is here used generically, and a family may be given any name, usually capital (blackboard bold if we do not know that it is a set). □

2.4.13 Example. Thus, $\emptyset$ is a family of sets; the empty family.

So are $\{\{2\},\{2,\{3\}\}\}$ and $\mathbb{V}$, the latter given by

$$\mathbb{V}\stackrel{Def}{=}\Big\{x : x \text{ is a set}\Big\}$$

BTW, as $\mathbb{V}$ contains all sets (but no atoms!) it is a proper class! Why? Well, if it is a set, then it is one of the x-values that we are collecting, thus $\mathbb{V}\in\mathbb{V}$. But we saw that this statement is false for sets!

Here are some classes that are *not* families: $\{1\}$, $\{2, \{\{2\}\}\}$ and $\mathbb{U}$, the latter being the universe of all objects —sets *and* atoms— and equals Russell's "R" as we saw in Section 2.2. These all are disqualified as they contain atoms. □

2.4.14 Definition. (Intersection and union of families) Let $\mathbb{F}$ be a family of sets. Then

(i) the symbol $\bigcap \mathbb{F}$ denotes the class that contains *all the objects* that *are common to all* $A \in \mathbb{F}$.

In symbols the definition reads:

$$\bigcap \mathbb{F} \stackrel{Def}{=} \Big\{ x : \text{for all } A, A \in \mathbb{F} \to x \in A \Big\} \tag{1}$$

(ii) the symbol $\bigcup \mathbb{F}$ denotes the class that contains *all the objects* that *are found among the various* $A \in \mathbb{F}$. That is, imagine that the members of *each* $A \in \mathbb{F}$ are "emptied" into a single —originally empty— container $\{\ldots\}$. The class we get this way is what we denote by $\bigcup \mathbb{F}$.

In symbols the definition reads (and I think it is clearer):

$$\bigcup \mathbb{F} \stackrel{Def}{=} \Big\{ x : \text{for some } A, A \in \mathbb{F} \wedge x \in A \Big\} \tag{2}$$

□

2.4.15 Example. Let $\mathbb{F} = \{\{1\}, \{1, \{2\}\}\}$. Then emptying all the contents of the members of $\mathbb{F}$ into some (originally) empty container we get

$$\{1, 1, \{2\}\} \tag{3}$$

This is $\bigcup \mathbb{F}$.

Would we get the same answer from the mathematical definition (2)? Of course:

1 *is* in some member of $\mathbb{F}$, indeed in both of the members $\{1\}$ and $\{1, \{2\}\}$, and in order to emphasise this I wrote two copies of 1 —it is emptied/contributed twice. Then $\{2\}$ is the member that only $\{1, \{2\}\}$ of $\mathbb{F}$ contributes.

We do not see any other members in the two set-members —$\{1\}$ and $\{1, \{2\}\}$— of $\mathbb{F}$. So, all done!

What is $\bigcap \mathbb{F}$? Well, only 1 is common between the two sets —$\{1\}$ and $\{1, \{2\}\}$— that are in $\mathbb{F}$. So, $\bigcap \mathbb{F} = \{1\}$. □

2.4.16 Exercise.

1. Prove that $\bigcup \Big\{A, B\Big\} = A \cup B$.
2. Prove that $\bigcap \Big\{A, B\Big\} = A \cap B$.

Hint. In each of part 1. and 2. show that lhs $\subseteq$ rhs and rhs $\subseteq$ lhs. For that analyse membership, i.e., "assume $x \in$ lhs and prove $x \in$ rhs", and conversely (cf. 2.1.1 and 2.1.2.) □

Jan. 24, 2022

2.4.17 Theorem. *If the* <u>*set*</u> F *is a family of sets, then* $\bigcup F$ *is a set.*

Proof. Let F be built at stage Σ (Princ. 1). Now,

$$x \in \bigcup F \equiv x \in \overset{\text{some}}{\overset{\downarrow}{A}} \in F$$

Thus x is available or built *before* A which is built *before* stage Σ since that is when F was built. x being arbitrary, all members of $\bigcup F$ are available/built *before* Σ, so we can build $\bigcup F$ as a set at stage Σ. □

2.4.18 Theorem. *If the class* $\mathbb{F} \neq \emptyset$ *is a family of sets, then* $\bigcap \mathbb{F}$ *is a set.*

Proof. By assumption there <u>is</u> some set in $\mathbb{F}$. Fix one such and call it D.

First note that

$$x \in \bigcap \mathbb{F} \rightarrow x \in D \tag{$*$}$$

Why? Because (1) of Definition 2.4.14 says that

$$x \in \bigcap \mathbb{F} \equiv \underline{\text{for all}}\ A \in \mathbb{F} \text{ we have } x \in A$$

Well, D *is* one of those "A" sets in $\mathbb{F}$, so if $x \in \bigcap \mathbb{F}$ then $x \in D$. We established $(*)$ and thus we established

$$\bigcap \mathbb{F} \subseteq D$$

by 2.1.1. We are done by 2.3.5. □

☡ **2.4.19 Remark.** What if $\mathbb{F} = \emptyset$? Does it affect Theorem 2.4.18? Yes, **badly**!

In Definition 2.4.14 we read

$$\bigcap \mathbb{F} \overset{Def}{=} \Big\{ x : \text{for all } A, A \in \mathbb{F} \to x \in A \Big\} \tag{$**$}$$

However, as *the hypothesis (i.e., lhs) of the implication in ($**$) is* ***false***, the implication itself is **true**. Thus the entrance condition "for all $A, A \in \mathbb{F} \to x \in A$" is true for *all* x and thus allows *ALL* objects x to get into $\bigcap \mathbb{F}$,

Thus $\bigcap \mathbb{F} = \mathbb{U}$, the universe of *all* objects which we saw (cf. Section 2.2) is a proper class —i.e., *not* a set. □ ☡

2.4.20 Exercise. What is $\bigcup F$ if $F = \emptyset$? Set or proper class? Can you "compute" which class it is exactly? □

2.4.21 Remark. (More notation)

Suppose the family of sets Q is a set of sets A_i, for $i = 1, 2, \ldots, n$ where $n \geq 3$.

$$Q = \{A_1, A_2, \ldots, A_n\}$$

Then we have a few alternative notations for $\bigcap Q$:

(a)

$$A_1 \cap A_2 \cap \ldots \cap A_n$$

or, more elegantly,

(b)

$$\bigcap_{i=1}^{n} A_i$$

or also

(c)

$$\textstyle\bigcap_{i=1}^{n} A_i$$

Similarly for $\bigcup Q$:

(i)

$$A_1 \cup A_2 \cup \ldots \cup A_n$$

or, more elegantly,

(ii)

$$\bigcup_{i=1}^{n} A_i$$

or also

(iii)

$$\textstyle\bigcup_{i=1}^{n} A_i$$

If the family has so many elements that *all the natural numbers are needed* to index the sets in the set family Q we will write

$$\bigcap_{i=0}^{\infty} A_i$$

or

$$\bigcap\nolimits_{i=0}^{\infty} A_i$$

or

$$\bigcap_{i\geq 0} A_i$$

or

$$\bigcap\nolimits_{i\geq 0} A_i$$

for $\bigcap Q$ and

$$\bigcup_{i=0}^{\infty} A_i$$

or

$$\bigcup\nolimits_{i=0}^{\infty} A_i$$

or

$$\bigcup_{i\geq 0} A_i$$

or

$$\bigcup\nolimits_{i\geq 0} A_i$$

for $\bigcup Q$ □

2.4.22 Example. Thus, for example, $A \cup B \cup C \cup D$ can be seen — just changing the notation— as $A_1 \cup A_2 \cup A_3 \cup A_4$, therefore it means, $\bigcup\{A_1, A_2, A_3, A_4\}$, or $\bigcup\{A, B, C, D\}$.

Same comment for $\cap$. □

Pause. How come for the case for $n = 2$ we *proved*[†] $A \cup B = \bigcup\{A, B\}$ (2.4.16) but *here* we say ($n \geq 3$) that something like the content of the previous remark and example are just *notation* (*definitions*)?

Well, we had *independent* definitions (and associated theorems re set status for each, 2.4.5 and 2.4.17) for $A \cup B$ and $\bigcup\{A, B\}$ so it makes sense to compare the two *independent* definitions after the fact and see if we can *prove* that *they say the same thing*.

For $n \geq 3$ we opted to *NOT* give a definition for $A_1 \cup \ldots \cup A_n$ that is independent of $\bigcup\{A_1 \cup \ldots \cup A_n\}$, rather we gave the definition of the former in terms of the latter.

No independent definitions, no theorem to compare the two!◀

[†] Well, *you* proved! Same thing :-)

2.5. The powerset

2.5.1 Definition. For any set A the symbol $\mathscr{P}(A)$ —pronounced the *powerset* of A— is defined to be the class

$$\mathscr{P}(A) \overset{Def}{=} \Big\{x : x \subseteq A\Big\}$$

Thus we collect *all* the subsets x of A to form $\mathscr{P}(A)$.

The literature most frequently uses the symbol 2^A in place for of $\mathscr{P}(A)$. □

☡ (1) The term "power*set*" is slightly premature, but it is apt. Under the conditions of the definition —A a set— 2^A is a *set* as we prove immediately below.

(2) We said "*all* the sub*sets* x of A" in the definition. This is correct. As we know from 2.3.5, if $\mathbb{X} \subseteq Y$ and Y is a set, then so is $\mathbb{X}$. ☡

Jan. 26, 2022

2.5.2 Theorem. *For any set A, its powerset $\mathscr{P}(A)$ is a set.*

Proof. Let A be built at stage Σ.

We need three steps to argue this:

Step 1. By Princ. 2, pick a stage Σ' *after* Σ: That is, $\Sigma < \Sigma'$.

Step 2. IF $x \subseteq A$, then we can build x at stage Σ (*same* as A).

Indeed, by Princ. 0, —and because all members of x are *also members of A*— they are *all* available *BEFORE* Σ (the stage at which A was built). So we can also build x at stage Σ.

Step 3. But then (Princ. 0) I can collect *ALL* the x satisfying $x \subseteq A$ and form 2^A as a *set* at stage Σ' ($> \Sigma$).

□

2.5.3 Example. Let $A = \{1, 2, 3\}$. Then

$$\mathscr{P}(A) = \Big\{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{3, 2\}, \{1, 2, 3\}\Big\}$$

Thus the powerset of A has 8 elements.

We will later see that if A has n elements, for any $n \geq 0$, then 2^A has 2^n elements. This observation is at the root of the notation "2^A". □

2.5.4 Remark. For any set A it is trivial (verify!) that we have $\emptyset \subseteq A$ and $A \subseteq A$. Thus, for any A, $\{\emptyset, A\} \subseteq 2^A$. □

Chapter 3

The Ordered Pair and Cartesian Products

To introduce the concepts of cartesian product —so that, in principle, plane analytic geometry can be developed within set theory— we need an object "(A, B)" that is *like* the set pair (2.3.1) in that it contains *two* objects, A and B ($A = B$ is a possibility), but in (A, B) order *and* length (here it is 2) matter!

That is,

> *We want* $(A, B) = (A', B')$ implies $A = A'$ *and* $B = B'$*. Moreover,* (A, A) *is not* $\{A\}$*! It is still an ordered* pair *but so happens that the first and second* component*, as we call the members of the ordered* pair*, are equal in this example.*

☡ So, are we going to accept a new type of object in set theory? *Not at all*! We will build (A, B) so that it is a set! ☡

3.0.1 Definition. (Ordered pair) *By definition*, (A, B) is the *abbreviation* (short name) given below:

$$(A, B) \overset{Def}{=} \Big\{A, \{A, B\}\Big\} \tag{1}$$

We call "(A, B)" an *ordered pair*, and A its first *component*, while B is its second component. □

3.0.2 Remark.

1. Note that $A \neq \{A, B\}$ because we would otherwise get $A \in A$, which is false for *sets or atoms* A. Thus (A, B) does contain exactly two members, or *has length 2*:

   $$A \text{ and } \{A, B\}.$$

   Pause. We have *not* said in 3.0.1 that A and B are sets or atoms. So what right do we have in the paragraph above to so declare?◀

2. What about the desired property that

   $$(A, B) = (X, Y) \rightarrow A = X \wedge B = Y \tag{2}$$

   Well, **assume the lhs** of "$\rightarrow$" in (2) and prove the rhs, "$A = X \wedge B = Y$".

   From our truth table we know that we do the latter by proving *each* of $A = X$ and $B = Y$ true (separately).

   The lhs that we *assumed* translates to

   $$\Big\{A, \{A, B\}\Big\} = \Big\{X, \{X, Y\}\Big\} \tag{3}$$

   By the remark #1 above there are *two* distinct members in each of the two sets that we equate in (3).

   So since (3) is true (by assumption) we have (by definition of set equality) one of:

   (a) $A = \{X, Y\}$ and $\{A, B\} = X$, that is, **1st listed element in lhs of "=" equals the 2nd listed in rhs; and 2nd listed element in lhs of "=" equals the 1st listed in rhs**.

   (b) $A = X$ and $\{A, B\} = \{X, Y\}$.

Now case (a) above *cannot hold*, for it leads to $A = \{\{A, B\}, Y\}$. This in turn leads to

$$\{A, B\} \in A$$

and thus the set $\{A, B\}$ is built *before* ONE of its members, A, which contradicts Principle 0.

Let's then work with case (b).

We have

$$\{A, B\} = \{A, Y\} \tag{4}$$

Well, all the members on the lhs must also be on the rhs. I note that A is.

- What if B is also equal to A? Then we have $\{B\} = \{A, Y\}$ and thus $Y \in \{B\}$ (why?). Hence $Y = B$. We showed so far $A = X$ (listed in case (b)) and $B = Y$ (proved here); great!
- Here B is *not* equal to A. But B must be in the rhs of (4), so the only way is $B = Y$. *All Done*! □

Worth noting as a theorem what we proved above:

3.0.3 Theorem. *If $(A, B) = (X, Y)$, then $A = X$ and $B = Y$.*

But is (A, B) a set? (atom it is not, of course!) Yes!

3.0.4 Theorem. *(A, B) is a set.*

Proof. Now $(A, B) = \Big\{A, \{A, B\}\Big\}$. By 2.3.1, $\{A, B\}$ is set. Applying 2.3.1 once more, $\Big\{A, \{A, B\}\Big\}$ is a set. □

3.0.5 Example. So, $(1, 2) = \{1, \{1, 2\}\}$, $(1, 1) = \{1, \{1\}\}$, and $(\{a\}, \{b\}) = \{\{a\}, \{\{a\}, \{b\}\}\}$. □

3.0.6 Remark. We can extend the ordered pair to ordered *triple*, ordered *quadruple*, and beyond!

We take this approach in these notes:

$$(A, B, C) \overset{Def}{=} \Big((A, B), C\Big) \tag{1}$$

$$(A, B, C, D) \overset{Def}{=} \Big((A, B, C), D\Big) \tag{2}$$

$$(A, B, C, D, E) \overset{Def}{=} \Big((A, B, C, D), E\Big) \tag{3}$$

etc. So suppose we defined what an n-tuple is, for *some fixed unspecified* n, and denote it by $(A_1, A_2, \ldots, A_n)$ for convenience. Then

$$(A_1, A_2, \ldots, A_n, A_{n+1}) \overset{Def}{=} \Big((A_1, A_2, \ldots, A_n), A_{n+1}\Big) \tag{$*$}$$

This is an "*inductive*" or "*recursive*" definition, defining a concept ($n + 1$-tuple) in terms of *a smaller instance of itself*, namely, in terms of the concept for an n-tuple, and in terms of the case $n = 2$ that we dealt with by *direct* definition (*not* in terms of the concept itself!) in 3.0.1.

$(*)$ is a general (for each length n that is) formation rule that allows us to build a tuple longer by ONE, compared to a tuple *we have already built*.

Suffice it to say this "case of $n+1$ in terms of case of n" provides just *shorthand notation* to take the mystery out of the red "etc." above. We **condense**/*codify* infinitely many definitions (1), (2), (3), . . . into just **two**:

- 3.0.1

  and

- $(*)$

The reader has probably seen such recursive definitions before (likely in calculus and/or high school).

The most frequent example that occurs is to define, for any natural number n and any real number $a > 0$, what a^n means. One goes like this:

$$\begin{aligned} a^0 &= 1 \\ a^{n+1} &= a \cdot a^n \end{aligned}$$

The above condenses infinitely many definitions such as

$$\begin{aligned} a^0 &= 1 \\ a^1 &= a \cdot a^0 = a \\ a^2 &= a \cdot a^1 = a \cdot a \\ a^3 &= a \cdot a^2 = a \cdot a \cdot a \\ a^4 &= a \cdot a^3 = a \cdot a \cdot a \cdot a \\ &\vdots \end{aligned}$$

into just two!

We will study *inductive definitions* and *induction* soon!

Before we exit this remark note that $(A, B, C) = (A', B', C')$ implies $A = A', B = B', C = C'$ because it says (3.0.6 (1))

$$((A, B), C) = ((A', B'), C')$$

and thus (3.0.3) implies

$$C = C' \text{ and } (A, B) = (A', B')$$

That is, (A, B, C) is an **ordered** triple (3-tuple).

We can also prove that $(A_1, A_2, \ldots, A_n, A_{n+1})$ is an **ordered** $n+1$-tuple, i.e.,

$$(A_1, A_2, \ldots, A_{n+1}) = (A'_1, A'_2, \ldots, A'_{n+1}) \rightarrow A_1 = A'_1 \wedge \ldots \wedge A_{n+1} = A'_{n+1}$$

if we have followed the "etc." all the way to the case of $(A_1, A_2, \ldots, A_n)$.

We will do the "etc."-argument *elegantly* once we learn induction!

□

Jan. 28, 2022

3.0.7 Definition. (Finite sequences) An n-tuple for $n \geq 1$ is called a finite sequence of length n, where we extend the concept to a *one element sequence* —**by definition**— to be

$$(A) \overset{Def}{=} A$$

□

☡ Note that now we can redefine all sequences of lengths $n \geq 1$ —pushing the starting point of the "etc.-construction" in 3.0.6 to $n = 1$ (from $n = 2$).

Using again $(*)$ above, but this time with starting condition that of 3.0.7, for $n = 2$ we rediscover (A_1, A_2):

$$\text{the "new" 2-tuple pair: } (A_1, A_2) \overset{\text{by } (*)}{=} \Big((A_1), A_2 \Big) \overset{\text{by } 3.0.7}{=} \Big(A_1, A_2 \Big)$$

The big red brackets are applications of the ordered pair defined in 3.0.1, just as it was in the general definition $(*)$. ☡

3.1. *The Cartesian product*

We next define classes of *ordered* pairs.

3.1.1 Definition. (Cartesian product of classes) Let $\mathbb{A}$ and $\mathbb{B}$ be classes. Then we define

$$\mathbb{A} \times \mathbb{B} \overset{Def}{=} \Big\{ (x, y) : x \in \mathbb{A} \wedge y \in \mathbb{B} \Big\}$$

The definition requires both sides of × to be classes. It makes no sense if one or both are atoms.

□

3.1.2 Theorem. *If A and B are sets, then so is $A \times B$.*

Proof. By 3.1.1 and 3.0.1

$$A \times B = \Big\{ \{x, \{x, y\}\} : x \in A \wedge y \in B \Big\} \tag{1}$$

Plan: I want to "find" a *set* "X" so that the inclusion $A \times B \subseteq X$ is true. Then I can then apply the subclass theorem (2.3.5).

Thus I am starting my search with "let $\{x, \{x, y\}\} \in A \times B$" and I am analysing this statement attempting to find a super**set**, X, of $A \times B$, that is, *find* an X such that $\{x, \{x, y\}\} \in X$.

So, for each $\big\{x, \{x, y\}\big\} \in A \times B$ we have $\underline{x \in A \text{ and } \{x, y\} \subseteq A \cup B}$, or $x \in A$ and $\{x, y\} \in 2^{A \cup B}$. Thus $\big\{x, \{x, y\}\big\} \subseteq A \cup 2^{A \cup B}$ and hence (changing notation) $(x, y) \in 2^{A \cup 2^{A \cup B}}$.

I found a "X" that works: $2^{A \cup 2^{A \cup B}}$

We have established that

$$A \times B \subseteq 2^{A \cup 2^{A \cup B}}$$

thus $A \times B$ is a set by 2.3.5, 2.4.5 and 2.5.2. □

3.1.3 Definition. Mindful of the Remark 3.0.6 where we defined $\Big(A, B, C\Big)$ as short for $\Big((A, B), C\Big)$, $\Big(A, B, C, D\Big)$ as short for $\Big((A, B, C), D\Big)$, etc., we define here $A_1 \times \ldots \times A_n$ for any $n \geq 3$ to mean

$$\{(x_1, x_2, \ldots x_n) : x_i \in A_i, \text{ for } i = 1, \ldots, n\}$$

and do it as follows:

$$\begin{array}{ll}
A \times B \times C & \stackrel{Def}{=} (A \times B) \times C \quad \text{justified as follows:} \\
 & = \{((x, y), z) : (x, y) \in A \times B \wedge z \in C\} \\
 & = \{((x, y), z) : x \in A \wedge y \in B \wedge z \in C\} \\
 & =^{\text{Def of } (x,y,z)} \{(x, y, z) : x \in A \wedge y \in B \wedge z \in C\} \\
A \times B \times C \times D & \stackrel{Def}{=} (A \times B \times C) \times D \\
\vdots & \\
A_1 \times A_2 \times \ldots \times A_n \times A_{n+1} & \stackrel{Def}{=} (A_1 \times A_2 \times \ldots \times A_n) \times A_{n+1} \\
\vdots &
\end{array}$$

We may write $\bigtimes_{i=1}^{n} A_i$ for $A_1 \times A_2 \times \ldots \times A_n$

If $A_1 = \ldots = A_n = B$ we may write B^n for $A_1 \times A_2 \times \ldots \times A_n$. □

3.1.4 Remark. Thus, what we learnt in 3.1.3 is, in other words,

$$\bigtimes_{i=1}^{n} A_i \stackrel{Def}{=} \Big\{(x_1, \ldots, x_n) : x_i \in A_i, \text{ for } i = 1, 2, \ldots, n\Big\}$$

and

$$B^n \stackrel{Def}{=} \Big\{(x_1, \ldots, x_n) : x_i \in B\Big\}$$

□

3.1.5 Theorem. *If A_i, for $i = 1, 2, \ldots, n$ is a set, then so is $\bigtimes_{i=1}^{n} A_i$.*

Proof. $A \times B$ is a set by 3.1.2. By 3.1.3, **and in this order**, we verify that so is $A \times B \times C$ and $A \times B \times C \times D$ and ... and $A_1 \times A_2 \times \ldots \times A_n$ and ... □

☡ If we had inductive definitions available already, then Definition 3.1.3 would simply read

$$A_1 \times A_2 \overset{Def}{=} \Big\{(x_1, x_2) : x_1 \in A_1 \wedge x_2 \in A_2\Big\}$$

and, for $n \geq 2$,

$$A_1 \times A_2 \times \ldots \times A_n \times A_{n+1} \overset{Def}{=} (A_1 \times A_2 \times \ldots \times A_n) \times A_{n+1}$$

Correspondingly, the proof of 3.1.5 would be far more elegant, via induction. ☡

Chapter 4

Relations and functions

Jan. 31, 2022

The topic of relations and functions is central in all *mathematics* and *computing*.

In *mathematics*, whether it is calculus, algebra or anything else, one deals with relations (notably *equivalence relations*, *order*) and all sorts of functions while in *computing* one computes relations and functions, that is, writing programs that given an input to a relation they compute the response (true or false) or given an input to a function they compute a response which is some object (number, graph, tree, matrix, other) or *nothing*, *in case there is no response* for said input (for example, there is no response to input "x, y" if what we are computing is $\frac{x}{y}$ but $y = 0$).

We are taking an "extensional" point of view in this course —as is customary in set theory, algebra, calculus— of relations and functions, that is, *we view them as classes of (input, output) ordered pairs*.

It is also possible to take an *intentional* point of view, *especially in computer science* and some specific areas of mathematics, viewing relations and functions as *methods* to compute outputs from given inputs.

4.1. Relations

4.1.1 Definition. (Binary relation) A binary relation is a class $\mathbb{R}$[†] of ordered pairs.

The statements $(x, y) \in \mathbb{R}$, $x\mathbb{R}y$ and $\mathbb{R}(x, y)$ are *equivalent*. $x\mathbb{R}y$ is the preferred "infix" notation —imitating notation such as $A \subset B$, $x < y$, $x = y$ and has notational advantages. □

4.1.2 Remark. $\mathbb{R}$ contains just pairs (x, y), that is, just sets $\{x, \{x, y\}\}$, that is, it is a family of sets.

Since $(x_1, x_2, \ldots, x_n) = \Big((x_1, x_2, \ldots, x_{n-1}), x_n\Big)$, it follows that binary relations (classes of ordered pairs) *is all we need to study*.

BTW, a class of ordered n-tuples, $(x_1, x_2, \ldots, x_n)$, is called *an n-ary relation*. As I said above we do not need to pay special attention to them. □

[†] I write "$\mathbb{R}$" or "R" for a relation, generically, but $\mathbb{P}$, $\mathbb{Q}$, $\mathbb{S}$ are available to use as well. I will avoid specific names such as $<$, $\subseteq$ in a general discussion. These two are apt to bring in in examples.

4.1.3 Example. Examples of relations:

(i) $\emptyset$

(ii) $\{(1,1)\}$

(iii) $\{(1,1),(1,2)\}$

(iv) $\mathbb{N}^2$, that is $\{(x,y) : x \in \mathbb{N} \wedge y \in \mathbb{N}\}$. This is a set by the fact that $\mathbb{N}$ is (Why?) and thus so is $\mathbb{N} \times \mathbb{N}$ by 3.1.2.

(v) $<$ on $\mathbb{N}$, that is $\{(x,y) : x < y \wedge x \in \mathbb{N} \wedge y \in \mathbb{N}\}$. This is a set since $< \subseteq \mathbb{N}^2$.

(vi) $\in$, that is,

$$\{(x,y) : x \in y \wedge x \in \mathbb{U} \wedge y \in \mathbb{V}\} \tag{$*$}$$

This is a proper class (nonSet). Why? Well, if $\in$ *is* a set, then it is built at some stage Σ.

Now examine the arbitrary (x,y) in $\in$. This is $\{x,\{x,y\}\}$ so it is built before Σ, but then so is its member x (available before Σ). Thus we can collect *all* such x into a *set* at stage Σ. But this "set" contains *all* $x \in \mathbb{U}$ due to the middle conjunct in the entrance condition in $(*)$.[†] That is, this "set" is $\mathbb{U}$. This is absurd!

□

[†] Hmm. Doesn't the first conjunct "$x \in y$" reduce the number of x-values? No: *For every* x out there take $y = \{x\}$ thus the conjunct $x \in y$ is fulfilled for all x-values, as I showed how to find a y that works.

So, a binary relation $\mathbb{R}$ is a table of pairs:

input: x	output: y
a	b
a'	b'
$\vdots$	$\vdots$
u	v
$\vdots$	$\vdots$

1. Thus one way to view R is as a device that for inputs x, valued $a, a', \ldots, u, \ldots$ one gets the outputs y, valued $b, b', \ldots, v, \ldots$ respectively. It is all right that a given input may yield multiple outputs (e.g., case (iii) in the previous example).

2. Another point of view is to see *both* x and y as inputs and the outputs are **true** or **false** (**t** or **f**). *Such is the way we usually view the relations $<$ and $=$ on the natural numbers.*

   For example, (a, b) is in the table (that is, aRb is true) hence if both a and b are ordered input values, then the relation outputs **t**.

Most of the time we will take the point of view in 1 above. This point of view compels us to define *domain* and *range* of a relation $\mathbb{R}$, that is, the class of all inputs that *cause an output* and the class of all *caused outputs* respectively.

4.1.4 Definition. (Domain and range) For any relation $\mathbb{R}$ we define *domain*, in symbols "dom" by

$$\mathrm{dom}(\mathbb{R}) \overset{Def}{=} \{x : (\exists y)x\mathbb{R}y\}$$

where we have introduced the notation "$(\exists y)$" as short for "there exists some y such that", or "for some y".

Range, in symbols "ran", is defined also in the obvious way:

$$\mathrm{ran}(\mathbb{R}) \overset{Def}{=} \{x : (\exists y)y\mathbb{R}x\} \qquad \square$$

We settle the following, before other things:

4.1.5 Theorem. *For a set relation R, both* $\mathrm{dom}(R)$ *and* $\mathrm{ran}(R)$ *are sets.*

Proof. For domain we collect all the x such that xRy, for some y, that is, all the x such that

$$\{x, \{x, y\}\} \in R \tag{1}$$

for some y.

So, R has the form

$$\Big\{\{x, \{x, y\}\}, \{x', \{x', y'\}\}, \{x'', \{x'', y''\}\}, \ldots\Big\}$$

$$\Big\{x, \{x, y\}, x', \{x', y'\}, x'', \{x'', y''\}, \ldots\Big\}$$

$\mathrm{dom}(R)$ is thus the collection of all the $x, x', x'', \ldots$ (4.1.4).

Thus

$$\mathrm{dom}(R) \subseteq \bigcup R \tag{†}$$

Now, R is a set-family of sets, thus $\bigcup R$ is a set. But then by (†) and the subclass theorem, $\mathrm{dom}(R)$ is a set. This settles the domain case.

Let A be the set of all atoms in $\bigcup R$ and define

$$S \overset{Def}{=} \Big(\bigcup R\Big) - A$$

So, S is a *set family*, and it contains *all* the $\{x, y\}$ parts of *all* $\{x, \{x, y\}\} \in R$. Thus,

$$S = \Big\{\{x, y\}, \{x', y'\}, \{x'', y''\}, \ldots;\ \text{plus those } x, x', x'', \ldots \text{ that are } \textit{sets}\Big\}$$

Then $\bigcup S$ contains all the y. That is, $\text{ran}(R) \subseteq \bigcup S$, and that settles the range case. □

4.1.6 Definition. In practice we often have an *a priori decision* about what are *in principle* "legal" inputs for a relation $\mathbb{R}$, and where its outputs go.

Thus we have two classes, $\mathbb{A}$ and $\mathbb{B}$ for the class of legal inputs and possible outputs respectively. Clearly we have $\mathbb{R} \subseteq \mathbb{A} \times \mathbb{B}$.

We call $\mathbb{A}$ and $\mathbb{B}$ *left field* and *right field* respectively, and instead of $\mathbb{R} \subseteq \mathbb{A} \times \mathbb{B}$ we often write

$$\mathbb{R} : \mathbb{A} \to \mathbb{B}$$

and also

$$\mathbb{A} \xrightarrow{\mathbb{R}} \mathbb{B}$$

pronounced "$\mathbb{R}$ is a relation *from* $\mathbb{A}$ *to* $\mathbb{B}$".

The term *field* —without left/right qualifiers— for $\mathbb{R} : \mathbb{A} \to \mathbb{B}$ refers to $\mathbb{A} \cup \mathbb{B}$.

If $\mathbb{A} = \mathbb{B}$ then we have

$$\mathbb{R} : \mathbb{A} \to \mathbb{A}$$

but rather than pronouncing this as "$\mathbb{R}$ is a relation *from* $\mathbb{A}$ *to* $\mathbb{A}$" we *prefer*[†] to say "$\mathbb{R}$ is on $\mathbb{A}$". □

4.1.7 Example. The a priori legal inputs in *Number Theory* and in *Computability* are all the natural numbers. □

[†] Both ways of saying it are correct.

4.1.8 Remark. Trivially, for any $\mathbb{R} : \mathbb{A} \to \mathbb{B}$, we have $\mathrm{dom}(\mathbb{R}) \subseteq \mathbb{A}$ and $\mathrm{ran}(\mathbb{R}) \subseteq \mathbb{B}$ (give a quick proof of each of these inclusions).

Also, for any relation $\mathbb{P}$ with no *a priori* specified left/right fields, $\mathbb{P}$ is a relation from $\mathrm{dom}(\mathbb{P}) \to \mathrm{ran}(\mathbb{P})$.

Naturally, we say that $\mathrm{dom}(\mathbb{P}) \cup \mathrm{ran}(\mathbb{P})$ is the field of $\mathbb{P}$ in this case.

□

4.1.9 Example. As an example, consider the *divisibility relation* on all integers (their set denoted by $\mathbb{Z}$) denoted by "$|$":

$$x|y \text{ means } x \text{ divides } y \text{ with } 0 \text{ remainder}$$

thus, for $x = 0$ and all y, the division is *illegal*, therefore

> *The input $x = 0$ to the relation "$|$"* ***produces no output****, in other words, "**for input $x = 0$ the relation is undefined**."*

We walk away with two things from this example:

1. It **does** make sense for some relations to *a priori* choose left and right fields, here
$$| : \mathbb{Z} \to \mathbb{Z}$$
You would not have divisibility on *real numbers*!

2. $\mathrm{dom}(\,|\,)$ is the set of all inputs that produce some output. Thus, it is NOT the case for all relations that their domain is the same as the left field *chosen*! Note the case in this example! And forget the term "codomain" that you may find in flawed publications on discrete MATH! □

4.1.10 Example. Next consider the relation $<$ with left/right fields restricted to $\mathbb{N}$. Then $\operatorname{dom}(<) = \mathbb{N}$, but $\operatorname{ran}(<) \underset{\neq}{\subseteq} \mathbb{N}$. Indeed, $0 \in \mathbb{N} - \operatorname{ran}(<)$. □

Feb. 2, 2022

Let us extract some terminology from the above examples:

4.1.11 Definition. Given

$$\mathbb{R} : \mathbb{A} \to \mathbb{B}$$

If $\operatorname{dom}(\mathbb{R}) = \mathbb{A}$, then we call $\mathbb{R}$ *total* or *totally defined*. If $\operatorname{dom}(\mathbb{R}) \underset{\neq}{\subseteq} \mathbb{A}$, then we say that $\mathbb{R}$ is *nontotal*.

If $\operatorname{ran}(\mathbb{R}) = \mathbb{B}$, then we call $\mathbb{R}$ *onto*. If $\operatorname{ran}(\mathbb{R}) \underset{\neq}{\subseteq} \mathbb{B}$, then we say that $\mathbb{R}$ is *not onto*. □

So, the relation $\mid$ above is nontotal, and $<$ is not onto.

In what follows we move away from the full generality of classes (possibly proper) and *restrict attention to relations that are sets*.

4.1.12 Example. Let $A = \{1, 2\}$.

- The relation $\{(1, 1)\}$ on A is neither total nor onto.
- The relation $\{(1, 1), (1, 2)\}$ on A is onto but not total.
- The relation $\{(1, 1), (2, 1)\}$ on A is total but not onto.
- The relation $\{(1, 1), (2, 2)\}$ on A is total *and* onto.
- The relation $\{(1, 2), (2, 1)\}$ on A is total *and* onto. □

4.1.13 Definition. The relation Δ_A on the set A is given by

$$\Delta_A \overset{Def}{=} \{(x,x) : x \in A\}$$

We call it the *diagonal* ("Δ" for "diagonal") or *identity* relation on A.

Consistent with the second terminology, we may also use the symbol $\mathbf{1}_A$ for this relation. □

4.1.14 Definition. A relation R (not *a priori* restricted to have *predetermined* left or right fields) is

1. *Transitive*: Iff $xRy \land yRz$ implies xRz.
2. *Symmetric*: Iff xRy implies yRx.
3. *Antisymmetric*: Iff $xRy \land yRx$ implies $x = y$.
4. *Irreflexive*: Iff xRy implies $x \neq y$. Also said this way: *For NO x can we have xRx.*
5. Now assume R is *on a set* A. Then we call it reflexive iff $\Delta_A \subseteq R$.

□

4.1.15 Example.

(i) *Transitive* examples: $\emptyset$ (*vacuously*), $\{(1,1)\}$, $\{(1,2),(2,3),(1,3)\}$, $<$, $\leq$, $=$, $\mathbb{N}^2$.

(ii) *Symmetric* examples: $\emptyset$ (*vacuously*), $\{(1,1)\}$, $\{(1,2),(2,1)\}$, $=$, $\mathbb{N}^2$.

(iii) *Antisymmetric* examples: $\emptyset$ (*vacuously*), $\{(1,1)\}$, $=$, $\leq$, $\subseteq$.

(iv) *Irreflexive* examples: $\emptyset$ (*vacuously*), $\{(1,2)\}$, $\subsetneqq$, the relations "$<$" and "$\neq$" on $\mathbb{N}$.

(v) *Reflexive* examples: $\mathbf{1}_A$ on A, $\{(1,1)\}$ on $\{1\}$, $\{(1,2),(2,1),(1,1),(2,2)\}$ on $\{1,2\}$, $=$ on $\mathbb{N}$, $\leq$ on $\mathbb{N}$. □

We can compose relations:

4.1.16 Definition. (Relational composition) Let R and S be (set) relations. Then, their **composition**, *in that order*, denoted by $R \circ S$ is defined for all x and y by:

$$xR \circ Sy \overset{Def}{\equiv} (\exists z)\Big(xRz \wedge zSy\Big)$$

It is *customary* (lazy and incorrect, though) to *abuse* notation and write "$xRzSy$" for "$xRz \wedge zSy$" just as one writes $x < y < z$ for $x < y \wedge y < z$.

The definition unchanged applies to any *class* relations $\mathbb{R}$ and $\mathbb{S}$ as well. □

4.1.17 Example. Here is whence the emphasis "*in that order*" above. Say, $R = \{(1,2)\}$ and $S = \{(2,1)\}$. Thus, $R \circ S = \{(1,1)\}$ while $S \circ R = \{(2,2)\}$. Thus, $R \circ S \neq S \circ R$ *in general*. □

4.1.18 Example. For any R, we *diagrammatically* indicate xRy by

$$x \xrightarrow{R} y$$

Thus, the situation where we have that $xR \circ Sy$ means, for some z, $xRzSy$ is depicted as:

$$\begin{tikzcd} & z \arrow[dr, "S"] & \\ x \arrow[ur, "R"] \arrow[rr, "R \circ S"'] & & y \end{tikzcd}$$

□

4.1.19 Theorem. *The composition of two (set) relations R and S in that order is also a set.*

Proof. Trivially, $R \circ S \subseteq \mathrm{dom}(R) \times \mathrm{ran}(S)$ because

$$xR \circ Sy$$

means

$$\overset{\overset{\in \mathrm{dom}(R)}{\downarrow}}{x} \quad R \quad z \quad S \quad \overset{\overset{\in \mathrm{ran}(S)}{\downarrow}}{y} \quad \text{, for some } z$$

Moreover, we proved in 4.1.5 that $\mathrm{dom}(R)$ and $\mathrm{ran}(S)$ are sets. Thus so is $\mathrm{dom}(R) \times \mathrm{ran}(S)$ (3.1.2). □

4.1.20 Corollary. *If we have $R : A \to B$ and $S : B \to C$, then $R \circ S : A \to C$.*

Proof. From $R \circ S \subseteq \mathrm{dom}(R) \times \mathrm{ran}(S)$ above and $\mathrm{dom}(R) \subseteq A$ and $\mathrm{ran}(S) \subseteq C$. □

The result of the corollary is depicted diagrammatically as

$$\begin{tikzcd}
 & B \arrow[dr, "S"] & \\
A \arrow[ur, "R"] \arrow[rr, "R \circ S"'] & & C
\end{tikzcd}$$

4.1.21 Theorem. (Associativity of composition) *For any relations* $\mathbb{R}, \mathbb{S}$ *and* $\mathbb{T}$*, we have*

$$(\mathbb{R} \circ \mathbb{S}) \circ \mathbb{T} = \mathbb{R} \circ (\mathbb{S} \circ \mathbb{T})$$

We state and prove this central result for any class relations.

Proof. We have two directions:

$\rightarrow$: Fix x and y and let $x(\mathbb{R} \circ \mathbb{S}) \circ \mathbb{T}y$.

Then, for some z, we have $x(\mathbb{R} \circ \mathbb{S})z\mathbb{T}y$ and hence for some w, the above becomes

$$x\mathbb{R}w\mathbb{S}z\mathbb{T}y \tag{1}$$

But $w\mathbb{S}z\mathbb{T}y$ means $w\mathbb{S} \circ \mathbb{T}y$, hence we rewrite (1) as

$$x\mathbb{R}w(\mathbb{S} \circ \mathbb{T})y$$

Finally, the above says $x\mathbb{R} \circ (\mathbb{S} \circ \mathbb{T})y$.

$\leftarrow$: Fix x and y and let $x\mathbb{R} \circ (\mathbb{S} \circ \mathbb{T})y$.

Then, for some z, we have $x\mathbb{R}z(\mathbb{S} \circ \mathbb{T})y$ and hence for some u, the above becomes

$$x\mathbb{R}z\mathbb{S}u\mathbb{T}y \tag{2}$$

But $x\mathbb{R}z\mathbb{S}u$ means $x\mathbb{R} \circ \mathbb{S}u$, hence we rewrite (2) as

$$x(\mathbb{R} \circ \mathbb{S})u\mathbb{T}y$$

Finally, the above says $x(\mathbb{R} \circ \mathbb{S}) \circ \mathbb{T}y$. □

The following is almost unnecessary, but offered for emphasis:

4.1.22 Corollary. *If R, S and T are (set) relations, all on some set A,[†] then "$R \circ S \circ T$" has a meaning independent of how brackets are inserted.*

The corollary allows us to just omit brackets in a chain of compositions, even longer than the above. It also leads to the definition of relational exponentiation, below:

4.1.23 Definition. (Powers of a binary relation) Let R be a (set) relation. We define R^n, for $n > 0$, as

$$\underbrace{R \circ R \circ \cdots \circ R}_{n\ R} \tag{1}$$

Note that the resulting relation in (1) is independent of how brackets are inserted (4.1.22).

If moreover we have defined R to be *on a set A*, then we also define the 0-th power: R^0 stands for Δ_A or $\mathbf{1}_A$. □

[†] Recall that "R is on a set A" means $R \subseteq A^2$, which is the same as $R : A \to A$.

Feb. 4, 2022

4.1.24 Remark. Say aR^nb.

Then, viewing R^n as $R \circ R^{n-1}$ I have

$$\begin{aligned}
aR^nb \iff & aRa_1R^{n-1}b && \text{for some } a_1\\
\iff & aRa_1Ra_2R^{n-2}b && \text{similarly, for some } a_2\\
\iff & aRa_1Ra_2Ra_3R^{n-3}b && \text{similarly, for some } a_3\\
\vdots &&&\\
\iff & \overbrace{aRa_1Ra_2Ra_3Ra_4\ldots a_{n-1}Rb}^{n\ R} && \text{similarly, for some } a_{n-1}
\end{aligned}$$

Summarising:

Thus aR^nb means that for some $a_1, a_2, \ldots, a_{n-1}$ we have

$$\underbrace{a}_{\in\mathrm{dom}(R)}\ Ra_1Ra_2Ra_3Ra_4\ldots a_{n-1}R\ \underbrace{b}_{\in\mathrm{ran}(R)} \tag{1}$$

□

4.1.25 Exercise. Let R be a relation on A. Then for all $n \geq 0$, R^n is a set.

Hint. See (1) above

□

4.1.26 Example. Let $R = \{(1,2),(2,3)\}$. What is R^2?

Well, when can we have xR^2y? Precisely if/when we can find x, y, z that satisfy $xRzRy$. The values $x = 1$, $y = 3$ and $z = 2$ are the *only ones* that satisfy $xRzRy$.

Thus $1R^23$, or $(1,3) \in R^2$. We conclude $R^2 = \{(1,3)\}$ by the "only ones" above. □

4.1.27 Exercise. Show that if for a relation R we know that $R^2 \subseteq R$, then R is transitive and conversely. □

4.1.1. Transitive closure

4.1.28 Definition. (Transitive closure of R) A *transitive closure* of a relation R —if it exists— is the $\subseteq$-*smallest* transitive T that *contains R as a subset*.

More precisely,

1. T is transitive, and $R \subseteq T$.
2. If S is also transitive and $R \subseteq S$, then $T \subseteq S$. This makes the term "$\subseteq$-*smallest*" precise. □

Note that we *hedged twice* in the definition, because at this point we do not know yet:

- If every relation has a transitive closure; hence the "if it exists".
- We do not know *if it is unique* either, hence the emphasised indefinite article "A".

4.1.29 Remark. *Uniqueness* can be settled immediately *from the definition above*: Suppose T and T' fulfil Definition 4.1.28, that is, suppose *both are* transitive closures of some R. Thus,

1. $R \subseteq T$

   and

2. $R \subseteq T'$

since both are closures.

But now think of T as a closure and T' as the "S" of 4.1.28 (it includes R all right!)

Hence $T \subseteq T'$.

Now reverse the role-playing and think of T' as a closure, while T plays the role of "S". We get $T' \subseteq T$. Hence, $T = T'$. □

4.1.30 Definition. The unique transitive closure, *if it exists*, is denoted by R^+. □

4.1.31 Exercise. If R is transitive, then R^+ exists. In fact, $R^+ = R$. □

The above exercise is hardly exciting, but learning that R^+ exists for *every* R and also learning how to "compute" R^+ *is* exciting. We do this next.

4.1.32 Lemma. *Given a (set) relation R. Then $\bigcup_{n=1}^{\infty} R^n$ is a transitive (set) relation.*

Proof. We have two things to do.

1. $\bigcup_{n=1}^{\infty} R^n$ is a set.
2. $\bigcup_{n=1}^{\infty} R^n$ is a transitive relation.

Proof of **1.** By (1) in 4.1.24, $aR^{n+1}b$ implies $a \in \mathrm{dom}(R)$ and $b \in \mathrm{ran}(R)$.

Thus

$$R^{n+1} \subseteq \mathrm{dom}(R) \times \mathrm{ran}(R)$$

for $n \geq 0$.

So[†]

$$\begin{aligned} X \in \mathbb{F} = \{R^i : i = 1, 2, 3, \ldots\} &\Longrightarrow X \subseteq \mathrm{dom}(R) \times \mathrm{ran}(R) \\ &\Longrightarrow X \in 2^{\mathrm{dom}(R)\times \mathrm{ran}(R)} \end{aligned}$$

Thus, $\mathbb{F} \subseteq 2^{\mathrm{dom}(R)\times \mathrm{ran}(R)}$ is a *set* family of sets R^n, for $n \geq 1$ (apply 2.3.5) and we can use the notation from 2.4.21

$$\bigcup_{i=1}^{\infty} R^i = \bigcup \mathbb{F}$$

a set, as we know (2.4.17)

[†] Recall that "$\Longrightarrow$" says "implies" just like →, but the former is conjunctional!

Proof of **2.** Now, $\bigcup_{i=1}^{\infty} R^i$ is a set (by part 1) but also, of course, a *binary relation* since trivially it is a set of *ordered pairs*.

Next, we prove it is *transitive*.

Let

$$x \bigcup_{i=1}^{\infty} R^i \, y \bigcup_{i=1}^{\infty} R^i \, z$$

Thus for some n and m we have

$$x \, R^n \, y \,^{\dagger} R^m \, z$$

this says the same thing as

$$x \, \overbrace{R \circ R \circ \cdots R}^{n} \, y \, \overbrace{R \circ R \circ \cdots R}^{m} \, z$$

or

$$x \, \overbrace{R \circ R \circ \cdots R}^{n} \circ \overbrace{R \circ R \circ \cdots R}^{m} \, z$$

or

$$x \, \overbrace{R \circ R \circ \cdots R}^{n+m} \, z$$

hence, since $(x, z) \in R^{n+m}$ from above, we have

$$(x, z) \in \bigcup \Big\{ \ldots, R^{n+m}, \ldots \Big\}, \text{ that is, (2.4.21), } x \bigcup_{i=1}^{\infty} R^i \, z$$

□

[†] $x \bigcup_{i=1}^{\infty} R^i \, y$ means $(x, y) \in \bigcup_{i=1}^{\infty} R^i$, therefore $(x, y) \in R^n$ for some n by definition of $\bigcup_{n=1}^{\infty}$.

4.1.33 Remark. **Read me!** Why all this work for Part 1 of the proof above? Why not just *use* 2.4.21 right away? Because 2.4.21 offers *only notation* provided we *know* that

$$\mathbb{F} = \{A_0, A_1, A_2, A_3, \ldots\} \tag{3}$$

is a set! Cf. "Suppose the family of sets Q is a set of sets", the opening statement in the passage 2.4.21 on *notation*.

Here we do *not know* (yet) if *every* family of sets like (3) is indeed a set —but in *this* case it turns out that we *do not care* because *every* member of $\mathbb{F} = \{R^i : i = 1, 2, 3, \ldots\}$ is included (as a subset) in $\mathrm{dom}(R) \times \mathrm{ran}(R)$ (a set), which allows us to sidestep the issue!

Whether *every* family of *sets* like $\mathbb{F}$ in (3) is a set will be answered affirmatively in 4.1.42. For now note that we cannot recklessly say that after *any* sequence of construction by stages there is a stage *after all those stages*. Why? Well, take *all* the objects in set theory. Each is given outright (atom; stage 0) or is constructed at some stage (set). If we could *prove* there is a stage *after all these stages* then we could also *prove* that $\mathbb{U}$ is a set, a claim we refuted with *two* methods so far! □

Feb. 7, 2022

Since $R \subseteq \bigcup_{i=1}^{\infty} R^i$ due to $R = R^1$, all that remains to show is that $\bigcup_{i=1}^{\infty} R^i$ is a transitive closure of R is to show the Lemma below.

4.1.34 Lemma. *If $R \subseteq S$ and S is transitive, then $\bigcup_{i=1}^{\infty} R^i \subseteq S$.*

Proof. I will just show that for all $n \geq 1$, $R^n \subseteq S$.

(1) OK, $R \subseteq S$ is our *assumption*, thus $R^1 \subseteq S$ is true.

(2) For $R^2 \subseteq S$ let xR^2y, thus (for some z), $xRzRy$ hence $xSzSy$. But S is transitive, so xSy. Done.

(3) For $R^3 \subseteq S$ let xR^3y, thus (for some z), xR^2zRy hence $\overbrace{xSz}^{By\ (2)} Sy$. But S is transitive, so the last expression gives xSy. Done.

$(n+1)$ **You see the pattern**: Pretend we proved up to *some fixed unspecified* n:

$$R^n \subseteq S \qquad (\ddagger)$$

Thus, for the $n+1$ case, for the same n we just fixed,

$$xR^{n+1}y \Leftrightarrow xR^n \circ Ry \Leftrightarrow xR^n zRy \text{ (some } z\text{)} \overset{by\ (\ddagger)}{\Rightarrow} xSzSy \Rightarrow xSy^{\dagger}$$

□

† S is transitive.

We have proved:

4.1.35 Theorem. (*The* transitive closure exists) *For any relation R, its transitive closure R^+ exists and is unique. We have that $R^+ = \bigcup_{i=1}^{\infty} R^i$.*

An interesting corollary that will lend a computational flavour to 4.1.35 is the following.

4.1.36 Corollary. *If R is on the set $A = \{a_1, a_2, \ldots, a_n\}$ where, for $i = 1, \ldots, n$, the a_i are* distinct, *then $R^+ = \bigcup_{i=1}^{n} R^i$.*

Proof. By 4.1.35, all we have to do is prove

$$\bigcup_{i=1}^{\infty} R^i \subseteq \bigcup_{i=1}^{n} R^i \tag{1}$$

since the $\supseteq$ part is obvious.

So let $x \bigcup_{i=1}^{\infty} R^i y$. This means that

$$xR^q y, \text{ for some } q \geq 1 \tag{2}$$

Thus, I have two cases for (2):

Case 1. $q \leq n$. Then $x \bigcup_{i=1}^{n} R^i y$ since $R^q \subseteq \bigcup_{i=1}^{n} R^i$, R^q being one of the "R^i", $1 \leq i \leq n$.

Case 2. $q > n$. In this case I will show that there is *also* a $k \le n$ such that $xR^k y$, which sends me back to the "easy **Case 1**".

Well, if there is one $q > n$ that satisfies (2) there are probably more. Let us pretend that our q is *the smallest* $> n$ that gives us (2).

Wait! Why is there a *smallest* q such that

$$xR^q y \text{ and } q > n\,? \tag{3}$$

Because among those "q" that fit (3)[†] imagine we fix attention to one such.

Now, if it is not the smallest such, then go down to the *next smaller* one that still satisfies (3), call it q'.

Now go down to the next smaller, $q'' > n$, if q' is not smallest.

Continue like this. Can I do this forever? That is, can we have the following?

$$n < \ldots < q^{(k)\ddagger} < \ldots < q''' < \ldots < q'' < q' < q$$

If yes, then I will have an infinite "descending" chain of distinct natural numbers between q and n.

Absurd!

[†] There is at least one, else we would **not** be in **Case 2**.

[‡] By "$q^{(n)}$" I mean q with k primes.

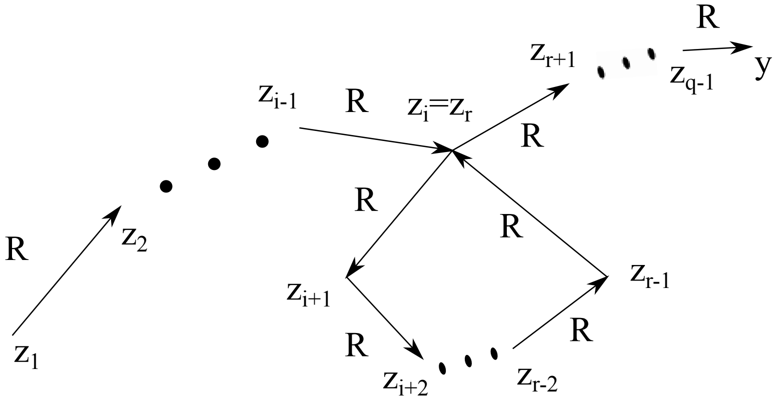
Back to the proof. So let *the q we are working with* be the *smallest* that satisfies (3). Then we have the configuration (see Remark 4.1.24 (1))

$$xRz_1Rz_2Rz_3\ldots\boxed{z_iRz_{i+1}\ldots}z_rRz_{r+1}\ldots z_{q-1}Ry \tag{4}$$

Now the sequence

$$z_1, z_2, z_3\ldots z_i, z_{i+1}, \ldots z_r, z_{r+1}, \ldots, z_{q-1}, y$$

in (4) above contains $q > n$ members and can also be depicted as



As they all come from A, **not all are distinct**. So let $z_i = z_r$ (the z_r could be as late in the sequence as y, i.e., equal to y).

Now omit the boxed part in (4) —that is, *the loop in the figure above*. We obtain

$$xRz_1Rz_2Rz_3\ldots \underset{\displaystyle z_i}{\underset{\|}{z_r}}Rz_{r+1}\ldots z_{q-1}Ry \tag{5}$$

which contains at least **one** "R" less than the sequence (4) does —the entry "z_iRz_{i+1}" (and everything else in the "…" part in the box) was removed. That is, (5) states

$$xR^{q'}y$$

with $q' < q$. Since the q in (3) was *smallest* $> n$, we *must have* $q' \leq n$ (Why?) which sends us to **Case 1** and we are done. □

4.1.2. Equivalence relations

Feb. 9, 2022

Equivalence relations must be *on some set A*, since we require reflexivity. They play a significant role in many branches of *mathematics* and even in *computer* science. For example, the minimisation process of finite automata (a topic that we will not cover) relies on the concept of equivalence relations.

4.1.37 Definition. A relation R on A is an equivalence relation, provided it is all of

1. Reflexive
2. Symmetric
3. Transitive □

☡ An equivalence relation on A has the effect, *intuitively*, of "grouping" elements that we view as *interchangeable in their roles*, or "equivalent", into so-called (see Definition 4.1.40 below) "*equivalence classes*" —kind of mathematical clubs! ☡

4.1.38 Example. The following are equivalence relations

- $\{(1,1)\}$ on $A=\{1\}$.
- $=$ (or $\mathbf{1}_A$ or Δ_A) on A.
- Let $A=\{1,2,3\}$. Then $R=\{(1,2),(1,3),(2,3),(2,1),(3,1),(3,2),(1,1),$ $(2,2),(3,3)\}$ is an equivalence relation on A.
- $\mathbb{N}^2$ is an equivalence relation on $\mathbb{N}$. □

Here is a longish, more sophisticated example, that is central in *number theory*. We will have another instalment of it after a few definitions and results.

4.1.39 Example. (Congruences) Fix an $m \geq 2$. We define the relation $\equiv_m$on $\mathbb{Z}$ by

$$x \equiv_m y \text{ iff } m \,|\, (x - y)$$

Recall that "$|$" is the "divides with zero remainder" relation. We verify the required properties for $\equiv_m$ to be an equivalence relation.

A notation that is very widespread in the literature is to split the symbol "$\equiv_m$" into two and write

$$x \equiv y \pmod{m} \text{ instead of } x \equiv_m y$$

"$x \equiv y \pmod{m}$" and $x \equiv_m y$ are read "x is *congruent* to y *modulo* m (or just 'mod m')". Thus "$\equiv_m$" is the congruence (mod m) short symbol, while "$\equiv \ldots \pmod{m}$" is the long two-piece symbol. *We will be using the short symbol*.

1. *Reflexivity*: Indeed, $m \mid (x - x)$, hence $x \equiv_m x$.

2. *Symmetry*: Clearly, if $m \mid (x - y)$, then $m \mid (y - x)$. I translate: If $x \equiv_m y$, then $y \equiv_m x$.

3. *Transitivity*: Let $m \mid (x - y)$ and $m \mid (y - z)$. The first says that, for some k, $x - y = km$. Similarly the second says, for some n, $y - z = nm$. Thus, adding these two equations I get $x - z = (k + n)m$, that is, $m \mid (x - z)$. I translate: If $x \equiv_m y$ and $y \equiv_m z$, then also $x \equiv_m z$. □

4.1.40 Definition. (Equivalence classes) Given an equivalence relation R on A. The *equivalence class* of an element $x \in A$ is $\{y \in A : xRy\}$. We use the symbol $[x]_R$, or just $[x]$ if R is understood, for the equivalence class.

Since A is a set and $[x] \subseteq A$, each equivalence class is a set by 2.3.5.

The symbol A/R denotes the *quotient* class of A with respect to R, that is,

$$A/R \overset{Def}{=} \{[x]_R : x \in A\}$$

$\square$

4.1.41 Remark. Suppose an equivalence relation R on A is given.

By reflexivity, xRx, for any x. Thus $x \in [x]_R$, hence all equivalence classes are *nonempty*.

Be careful to distinguish the brackets $\{\ldots\}$ from these $[\ldots]$.

It is NOT a priori obvious that $x \in [x]_R$ until you look at the definition 4.1.40! $[x]_R \neq \{x\}$ in general!

□

This is a good time to introduce "**Principle 3**"[†] of set formation.

4.1.42 Remark. (Principle 3) Suppose that the *class* family of sets $\mathbb{F}$ *is indexed* by some (or all) members of a *set* A. Then $\mathbb{F}$ is a set.

Being *indexed* by (some) members of a set A means that, to every $X \in \mathbb{F}$, we have attached as "*label(s)*" (often depicted as a subscript/superscript) some member(s) of A.

We can label a set of $\mathbb{F}$ with many labels, but we *may NOT use the same label twice* to label two (or the same) sets of $\mathbb{F}$ and *may NOT leave any set of $\mathbb{F}$ unlabelled*.

[†] This is the last Principle, I promise!

Thus, if $\mathbb{F} = \{A, B, C\}$, then $\{A_1, B_{13,19,0}, C_{42}\}$ is a valid labelling with members from $\mathbb{N}$.[‡]

$\{A_{1,13}, B_{13}, C_{19}\}$ is not correctly labelled (same label twice), the labelling of $\{A_{1,42}, B_{13}, C\}$ is also invalid (C was not labelled):

[‡] B has three labels attached to it.

Two observations:

1. We used *at least as many labels as members in the class* $\mathbb{F}$.

   Thus $\mathbb{F}$ has *no "more" members* than the label *set*, and thus is a set itself.

Some people call Principle 3 the **size limitation doctrine**.[†]

2. Why can't I use the Principles 0–2 to argue that $\mathbb{F}$, labelled by any *set* L, is a set?

   Well, because these Principles are notorious in not telling me if a stage *exists* after *infinitely many stages of construction* that I might have if, say, I were to build one set A_i for each natural number (here $L = \mathbb{N}$):

$$A_0, A_1, \ldots, A_n, \ldots$$

□

[†] Researchers on the foundations of set theory felt that paradoxes occurred in connection with enormous classes.

Feb. 11, 2022

We can now state the obvious:

4.1.43 Theorem. *A/R is a set for any set A and equivalence relation R on A.*

Proof. A provides labels for all members —namely, $[x]$, where $x \in A$— of A/R. Now invoke Principle 3. □

Now that we have had *an excuse to introduce Principle 3 early*, and applied it to the easy example above let us do the following exercise:

4.1.44 Exercise. Show that it *was not necessary* to apply the *new* Principle to prove 4.1.43.

Specifically show that the Theorem follows by Principles 0–2 implicitly via 2.3.5.

Hint. You will need, of course, to find a *superset* of A/R, that is, a class X that *demonstrably* is a set, and satisfies $A/R \subseteq X$. □

4.1.45 Lemma. *Let P be an equivalence relation on A. Then $[x] = [y]$ iff xPy —where we have omitted the subscript $_P$ from the $[\ldots]$-notation.*

Proof. $(\rightarrow)$ part. Assume $[x] = [y]$.

By reflexivity, $y \in [y]$ (4.1.41).

The assumption then yields $y \in [x]$ and therefore xPy by 4.1.40.

($\leftarrow$) part. Assume xPy.

Let $z \in [x]$. Then xPz.

By *assumption* yPx (by symmetry), thus, *transitivity* yields yPz.

That is, $z \in [y]$, proving

$$[x] \subseteq [y] \tag{1}$$

By *swapping letters* x, y we have *ALSO* proved above that

$$yPx \textit{ implies } [y] \subseteq [x] \tag{2}$$

Now (by symmetry) our *original* assumption, $\underline{\text{namely } xPy}$, implies yPx, hence also $[y] \subseteq [x]$ by (2).

All in all ((1)+(2)), $[x] = [y]$. □

4.1.46 Lemma. *Let R be an equivalence relation on A. Then*

(i) $[x] \neq \emptyset$, *for all* $x \in A$.

(ii) $[x] \cap [y] \neq \emptyset$ *implies* $[x] = [y]$, *for all* x, y *in* A.

(iii) $\bigcup_{x \in A}[x] = A$.

Proof.

(i) 4.1.41.

(ii) Let $z \in [x] \cap [y]$. Then xRz and yRz, therefore xRz and zRy (the latter by *symmetry*) hence xRy (transitivity).

Thus, $[x] = [y]$ by Lemma 4.1.45.

(iii) The $\subseteq$-part is obvious from $[x] \subseteq A$.

The $\supseteq$-part follows from $\bigcup_{x \in A}\{x\} = A$ and $\{x\} \subseteq [x]$. □

The properties (i)–(iii) are characteristic of the notion of a *partition of a set*.

4.1.47 Definition. (Partitions) Let F be a family of subsets of A. It is a *partition of* A iff all of the following hold:

(i) For all $X \in F$ we have that $X \neq \emptyset$.

(ii) If $\{X, Y\} \subseteq F$ and $X \cap Y \neq \emptyset$, then $X = Y$.

(iii) $\bigcup F = A$. □

There is a natural affinity between equivalence relations and partitions on a set A. In fact,

4.1.48 Theorem. *Given a partition F on a set A. This leads to the definition of an equivalence relation P whose equivalence classes are precisely the sets of the partition, that is $F = A/P$.*

Proof. First we define P:

$$xPy \overset{Def}{\text{iff}} (\exists X \in F)\{x,y\} \subseteq X \tag{1}$$

Observe that

(i) P is *reflexive*: Take any $x \in A$. By 4.1.47(iii), there is an $X \in F$ such that $x \in X$, hence $\{x,x\} \subseteq X$. Thus xPx.

(ii) P is, trivially, *symmetric* since there is no order in $\{x,y\}$.

(iii) P is *transitive*: Indeed, let $xPyPz$. Then $\{x,y\} \subseteq X$ and $\{y,z\} \subseteq Y$ for some X, Y in F.

Thus, $y \in X \cap Y$ hence $X = Y$ by 4.1.47(ii). Hence $\{x,z\} \subseteq X$, therefore xPz.

So P is an equivalence relation. Let us compare its equivalence classes with the various $X \in F$.

Now $[x]_P$ (dropping the subscript $_P$ in the remaining proof) is

$$\{y : xPy\} \tag{2}$$

Let us compare $[x]$ with the *unique* $X \in F$ that ALSO contains x —why unique? By 4.1.47(ii). Thus,

$$y \in [x] \overset{(2)}{\Longleftrightarrow} xPy \overset{(1)}{\Longleftrightarrow} x \in X \wedge y \in X \overset{x\in X \text{ is } \mathbf{t}}{\Longleftrightarrow} y \in X$$

Thus $[x] = X$. □

4.1.49 Example. (Another look at congruences) Euclid's theorem for the division of integers states:

If $a \in \mathbb{Z}$ and $2 \leq m \in \mathbb{Z}$, then *there are* unique q and r such that

$$a = mq + r \text{ and } 0 \leq r < m \tag{1}$$

There are many proofs, but here is one: Fix a and $m \geq 2$. The set

$$T = \{x : 0 \leq x = a - mz, \text{ for some z}\}$$

is *not empty*. For example,

- if $a > 0$, then take $z = 0$ to obtain $x = a > 0$ in T.

- If $a = 0$, then take $z = 0$ to obtain $x = 0 \in T$.

- Finally, if $a < 0$, then take $z = -|a|$ [†] to obtain $x = -|a| + m|a| = |a|(m - 1) > 0$ in T (since $m \geq 2$ we have $m - 1 \geq 1$).

Let then r be the *smallest* $x \geq 0$ in T.

[†] Absolute value.

The *corresponding* "z" to the *smallest* $x = r$ let us call q. So we have

$$a = mq + r, \text{ where } 0 \leq r \tag{2}$$

Can $r \geq m$? If so, then write $r = k + m$, where $k = r - m \geq 0$ and thus $k < r$. I got

$$a = m(q + 1) + k$$

As $k < r$, I have contradicted the minimality of r in (2) in the box above.

This proves that $r < m$.

We have proved *existence of at least one pair* q and r that works for (1) on p.117.

How about *uniqueness*?

Well, the worst thing that can happen is to have two representations 1). Here is another one:

$$a = mq' + r' \text{ and } 0 \le r' < m \tag{2}$$

As both r and r' are $< m$, their "distance" (absolute difference) is also $< m$.[†]

Now, from (1) and (2) we get

$$m|q - q'| = |r - r'| \tag{3}$$

This cannot be unless $q = q'$ (in which case $r = r'$, therefore uniqueness is proved).

Wait: Why it "cannot be" if $q \ne q'$?

Because then $|q - q'| \ge 1$ thus the lhs of "=" in (3) is $\ge m$ but the rhs is $< m$.

[†] From $0 \le r' < m$ I get $-m < r' \le 0$. Using (1) (p.117), I get $-m < r - r' < m$. That is, $|r - r'| < m$.

We now take a deep breath!

Feb. 14, 2022

Now, back to congruences! The above was just a preamble!

Fix an $m > 1$ and consider the congruences $x \equiv_m y$. What are the equivalence classes?

Better question is what representative members are convenient to use for each such class? Given that $a \equiv_m r$ by (1) (p.117), and using Lemma 4.1.45 we have $[a]_m = [r]_m$.

r is a far better representative than a for the class $[a]_m$ as it is "normalised".

Thus, we have just m equivalence classes $[0], [1], \ldots, [m-1]$.

Wait! Are they *distinct*? Yes! Since $[i] = [j]$ is the same as $i \equiv_m j$ (4.1.45) and, since $0 < |i-j| < m$, m *cannot* divide $i-j$ with 0 remainder, we cannot have $[i] = [j]$ if $i \neq j$. □

Feb. 14, 2022

4.1.50 Example. (A practical example) Say, I chose $m = 5$. Where does $a = -110987$ belong?

I.e., in which class out of $[0]_5$, $[1]_5$, $[2]_5$, $[3]_5$, $[4]_5$?

Well, let's do primary-school-learnt long division of $-a > 0$ divided by 5 and find quotient q and remainder r. We find, in this case, $q = 22197$ and $r = 2$. These satisfy

$$-a = 22197 \times 5 + 2$$

Thus,

$$a = -22197 \times 5 - 2 \tag{1}$$

(1) can be rephrased as

$$a \equiv_5 -2 \tag{2}$$

But easily we check that $-2 \equiv_5 3$ (since $3-(-2)=5$). Thus,

$$a \in [-2]_5 = [3]_5 \qquad \square$$

4.1.51 Exercise. Can you now <u>*easily*</u> write the same a above as

$$a = Q \times 5 + R, \text{ with } 0 \le R < 5?$$

Show all your work. □

4.1.3. Partial orders

This subsection introduces *one of the most important kind of binary relations in set theory and mathematics in general*: The *partial order* relations.

4.1.52 Definition. (Converse or inverse relation of $\mathbb{P}$) For any relation $\mathbb{P}$, the symbol $\mathbb{P}^{-1}$ is called the *converse* or *inverse* relation of $\mathbb{P}$ and is defined by

$$\mathbb{P}^{-1} = \{(x, y) : y\mathbb{P}x\} \tag{1}$$

$\underline{x\mathbb{P}^{-1}y \text{ iff } y\mathbb{P}x}$ is an equivalence that says exactly what (1) does. □

4.1.53 Example. If I take $\mathbb{P}$ to be "$<$" on $\mathbb{N}$, then $> = <^{-1}$ since

$$x > y \text{ iff } y < x$$

□

More notation!

4.1.54 Definition. (Important: “$(a)\mathbb{P}$” notation) For any relation $\mathbb{P}$ we write “$(a)\mathbb{P}$” to indicate the *class* —possibly proper— of *all outputs* of $\mathbb{P}$ *for input a*. That is,

$$(a)\mathbb{P} \overset{Def}{=} \{y : a \, \mathbb{P} \, y\}$$

If $(a)\mathbb{P} = \emptyset$, then $\mathbb{P}$ is *undefined* at a —that is, $\underline{a \notin \mathrm{dom}(\mathbb{P})}$.

The underlined statement is often denoted simply by “$\underline{(a)\mathbb{P} \uparrow}$” and is naturally read as “$\mathbb{P}$ is *undefined* at a”.

If $(a)\mathbb{P} \neq \emptyset$, then $\mathbb{P}$ is *defined* at a —a does produce outputs!— that is, $\underline{a \in \mathrm{dom}(\mathbb{P})}$.

The underlined statement is often denoted simply by “$\underline{(a)\mathbb{P} \downarrow}$” and is naturally read as “$\mathbb{P}$ is *defined* at a”. □

4.1.55 Exercise. Give an example of a specific relation $\mathbb{P}$ and one specific object (set or atom) a such that $(a)\mathbb{P}$ is *a proper class*. □

4.1.56 Example. We note that for any $\mathbb{P}$ and a,

$$(a)\mathbb{P}^{-1} = \{y : a\mathbb{P}^{-1}y\} = \{y : y\mathbb{P}a\}$$

Thus,

$$(a)\mathbb{P}^{-1}\uparrow \text{ iff } \{y : y\mathbb{P}a\} = \emptyset \text{ iff } a \notin \mathrm{ran}(\mathbb{P})$$

and

$$(a)\mathbb{P}^{-1}\downarrow \text{ iff } \{y : y\mathbb{P}a\} \neq \emptyset \text{ iff } a \in \mathrm{ran}(\mathbb{P})$$

□

4.1.57 Definition. (Partial order) A relation $\mathbb{P}$ is called a *partial order* or just an *order*, iff it is *all of*

(1) *irreflexive* (i.e., $x\mathbb{P}y \to x \neq y$ for all x, y), and

(2) *transitive*.

It is emphasised that in the interest of generality —for much of this subsection (until we say otherwise)— $\mathbb{P}$ need not be a set.

Some people call this a *strict order* as it imitates the "<" on, say, the natural numbers. □

4.1.58 Remark. (1) We will *usually* use the symbol "$<$" in *the abstract setting* to denote *any* unspecified *order* $\mathbb{P}$, and it will be pronounced "less than".

(2) If the order $<$ is a subclass of $\mathbb{A} \times \mathbb{A}$ —i.e., it is $<\colon \mathbb{A} \to \mathbb{A}$— then we say that *$<$ is an order on* $\mathbb{A}$.

(3) Clearly, for any order $<$ and any class $\mathbb{B}$, $< \cap (\mathbb{B} \times \mathbb{B})$ *is* an order on $\mathbb{B}$. □

4.1.59 Exercise. How clearly? (re (3) above.) Give a simple, short proof. □

4.1.60 Example. The concrete "less than", $<$, on $\mathbb{N}$ is an order, but $\leq$ is not (it is *not* irreflexive).

The "greater than" relation, $>$, on $\mathbb{N}$ is also an order, but $\geq$ is not.

In general, it is trivial to verify that "$\mathbb{P}$ is an order iff $\mathbb{P}^{-1}$" is an order. *Exercise*! □

4.1.61 Example. $\emptyset$ is an order.

Moreover for any $\mathbb{A}$, $\emptyset \subseteq \mathbb{A} \times \mathbb{A}$,

hence $\emptyset$ is also an order *on* $\mathbb{A}$ for the arbitrary $\mathbb{A}$. □

4.1.62 Example. The relation $\in$ is *irreflexive* by the well known $A \notin A$, for all A.

It is *not* transitive though.

For example, if a is a set (or atom), then $a \in \{a\} \in \{\{a\}\}$ but $a \notin \{\{a\}\}$.

So $\in$ is <u>not</u> an order. □

4.1.63 Example. Let $M = \Big\{\emptyset, \{\emptyset\}, \Big\{\emptyset, \{\emptyset\}\Big\}, \Big\{\emptyset, \{\emptyset\}, \Big\{\emptyset, \{\emptyset\}\Big\}\Big\}\Big\}$.

The relation

$$\varepsilon = \in \cap (M \times M)$$

is transitive (and irreflexive), hence it is an order (*on* M). *Verify*!

□

4.1.64 Example. $\subset$ is an order, $\subseteq$—failing irreflexivity— is not. □

4.1.65 Example. Consider the order $\subset$ again. In this case we have **none** of $\{\emptyset\} \subset \{\{\emptyset\}\}$, $\{\{\emptyset\}\} \subset \{\emptyset\}$ or $\{\{\emptyset\}\} = \{\emptyset\}$.

That is, $\{\emptyset\}$ and $\{\{\emptyset\}\}$ are *non comparable* items.

This justifies the qualification *partial* for orders in general (Definition 4.1.71).

On the other hand, the "natural" $<$ on $\mathbb{N}$ is such that one of $x = y$, $x < y$, $y < x$ always holds for any x, y.

That is, all (unordered) pairs x, y of $\mathbb{N}$ *are* comparable under $<$.

While *all* orders are "partial", some are *total* ($<$ above) and others are *nontotal* ($\subset$ above).

"Partial" is *not* the negation of "total".

□

Feb. 16, 2022

4.1.66 Definition. Let $<$ be a partial order on $\mathbb{A}$. We define

$$\leq \overset{Def}{=} \boldsymbol{\Delta}_{\mathbb{A}} \cup <$$

We pronounce $\leq$ "less than or equal".

$\boldsymbol{\Delta}_{\mathbb{A}} \cup >$ is denoted by $\geq$ and is pronounced "greater than or equal".

Let us call $\leq$ a *reflexive order* or *non strict order*. □

☡ (1) In plain English, given $<$ on $\mathbb{A}$, we *have defined* $x \leq y$ to mean

$$x < y \lor \overbrace{x = y}^{\text{equality is } \Delta_{\mathbb{A}}}$$

for all x, y in $\mathbb{A}$.

(2) The definition of $\leq$ *depends* on $\mathbb{A}$ due to the presence of $\boldsymbol{\Delta}_{\mathbb{A}}$.

There is no such dependency on a “reference” class in the case of $<$. ☡

4.1.67 Lemma. *For any $<: \mathbb{A} \to \mathbb{A}$, the associated relation $\leq$ on $\mathbb{A}$ is* reflexive, antisymmetric *and* transitive.

Proof. (1) *Reflexivity* is trivial.

(2) For *antisymmetry*, let $x \leq y$ and $y \leq x$. If $x = y$ then we are done.

So assume the remaining case $x \neq y$ (i.e., $(x, y) \notin \mathbf{\Delta}_{\mathbb{A}}$). Then the hypothesis becomes $x < y$ and $y < x$, therefore $x < x$ by transitivity, contradicting the irreflexivity of $<$. *This case does NOT apply*!

(3) As for *transitivity* let $x \leq y$ and $y \leq z$.

(a) If $x = z$, then $x \leq z$ (see the ☡-remark after 4.1.66) and we are done.

(b) The remaining case is $x \neq z$.

- If $x = y$ or $y = z$ (we cannot have both (Why?)), then we are done again.

- So it remains to consider $x < y$ and $y < z$.

  By transitivity of $<$ we get $x < z$, hence $x \leq z$, since $< \subseteq \leq$ (or, by logic: "$x \leq z$" is "$x < z \vee x = z$"; but we got $x < z$). □

4.1.68 Lemma. *Let $\mathbb{P}$ on $\mathbb{A}$ be reflexive, antisymmetric and transitive. Then $\mathbb{P} - \mathbf{\Delta}_{\mathbb{A}}$ is a (strict) order on $\mathbb{A}$.*

Proof. Since

$$\mathbb{P} - \mathbf{\Delta}_{\mathbb{A}} \subseteq \mathbb{P} \tag{1}$$

it is clear that $\mathbb{P} - \mathbf{\Delta}_{\mathbb{A}}$ is *on* $\mathbb{A}$.

It is also clear that it is *irreflexive*. We only need verify that it is *transitive*.

So let

$$(x, y) \text{ and } (y, z) \text{ be in } \mathbb{P} - \mathbf{\Delta}_{\mathbb{A}} \tag{2}$$

By (1) and (2)

$$(x, y) \text{ and } (y, z) \text{ are in } \mathbb{P} \tag{3}$$

hence

$$(x, z) \in \mathbb{P}$$

by *transitivity* of $\mathbb{P}$.

Can $(x, z) \in \mathbf{\Delta}_{\mathbb{A}}$, i.e., can $x = z$?

No, for antisymmetry of $\mathbb{P}$ and (3) would imply $x = y$, i.e., $(x, y) \in \mathbf{\Delta}_{\mathbb{A}}$ *contrary* to (2).

So, $(x, z) \in \mathbb{P} - \mathbf{\Delta}_{\mathbb{A}}$. □

4.1.69 Corollary. *Let $\leq$ on $\mathbb{A}$ be reflexive, antisymmetric and transitive. Then $<$ defined by*

$$x < y \overset{Def}{\equiv} x \leq y \wedge x \neq y$$

is a (strict) order on $\mathbb{A}$.

Proof. The corollary just rephrases 4.1.68. □

4.1.70 Remark. Often in the literature, but decreasingly so, it is the "reflexive order" $\le\colon \mathbb{A} \to \mathbb{A}$ that is defined as a "partial order" by the requirements that it is *reflexive*, *antisymmetric* and *transitive*.

Then $<$ is obtained as in Lemma 4.1.68, namely, as "$\le - \mathbf{\Delta}_{\mathbb{A}}$".

Lemmas 4.1.67 and 4.1.68 show that the two approaches are interchangeable, but the "modern" approach of Definition 4.1.57 avoids the nuisance of having to tie the notion of order to some particular "field" $\mathbb{A}$ (4.1.6).

For us "$\le$" is the *derived* notion defined in 4.1.66. □

4.1.71 Definition. (PO Class) If $<$ is an order *on* a class $\mathbb{A}$, we call the *informal* pair $(\mathbb{A}, <)$ a *partially ordered class*, or *PO class*.

If $<$ is an order on a *set* A, we call the pair $(A, <)$ a *partially ordered set* or *PO set*. Often, if the order $<$ is understood as being on $\mathbb{A}$ or A, one says that "$\mathbb{A}$ is a PO class" or "A is a PO set" respectively. □

☡☡ Formally, $(\mathbb{A}, <)$ is *not* an ordered pair since $\mathbb{A}$ may be a *proper* class and we do not allow class *members* —e.g., in $\{\mathbb{A}, \{\mathbb{A}, <\}\}$— to be proper classes. We may think then of "$(\mathbb{A}, <)$" as *informal* notation that simply "ties" $\mathbb{A}$ and $<$ together into a "toolbox" $(\ldots)$. ☡☡

4.1.72 Definition. (Linear order) A relation $<$ on $\mathbb{A}$ is a *total* or *linear* order *on* $\mathbb{A}$ iff it is all of

(1) An order, and

(2) For any x, y in $\mathbb{A}$ one of $x = y, \quad x < y, \quad y < x$ is true —this is the so-called "*trichotomy*" property.

If $\mathbb{A}$ is a class, then the informal pair $(\mathbb{A}, <)$ is a *linearly ordered class* —for short, a *LO class*.

If $\mathbb{A}$ is a set, then the pair $(\mathbb{A}, <)$ is a *linearly ordered set* —for short, a *LO set*.

One often calls just $\mathbb{A}$ a LO class or LO set (as the case warrants) when $<$ is understood from the context. □

4.1.73 Example. The standard $<: \mathbb{N} \to \mathbb{N}$ is a total order, hence $(\mathbb{N}, <)$ is a LO set.

4.1.74 Definition. (Minimal and minimum elements) Let $<$ be an order and $\mathbb{A}$ some class.

We are *not* postulating that $<$ is *on* $\mathbb{A}$.

An element $b \in \mathbb{A}$ is *a* $<$*-**minimal** element in* $\mathbb{A}$, or *a* $<$*-minimal element of* $\mathbb{A}$, iff

$$\neg(\exists x \in \mathbb{A})x < b$$

In words, there is nothing below b in $\mathbb{A}$.

$m \in \mathbb{A}$ is *a* $<$*-**minimum** element in* $\mathbb{A}$ iff $(\forall x \in \mathbb{A})m \leq x$.

We also use the terminology *minimal* or *minimum* *with respect to* $<$, instead of $<$-minimal or $<$-minimum.

If $a \in \mathbb{A}$ is $>$-minimal in $\mathbb{A}$, that is $\neg(\exists x \in \mathbb{A})x > a$, we call a a $<$-*maximal* element in $\mathbb{A}$. Similarly, a $>$-minimum element is called a $<$-*maximum*.

If the order $<$ is understood, then the qualification "$<$-" is omitted. □

4.1.75 Exercise. In particular, if b ($\in \mathbb{A}$) is *not* in the *field*

$$\text{dom}(<) \cup \text{ran}(<)$$

(cf. 4.1.6) of $<$, then b is *both* $<$-minimal and $<$-maximal *in* $\mathbb{A}$. □

4.1.76 Remark. Note how the notation learnt from 4.1.54 can *simplify* the expression

$$\neg(\exists x \in \mathbb{A})x < a \tag{1}$$

Since $x < a$ iff $a > x$, (1) says that *no x is in* ***both*** $\mathbb{A}$ and $(a)>$.[†]

That is, a is $<$-minimal in $\mathbb{A}$ iff

$$\mathbb{A} \cap (a)> = \emptyset \tag{2}$$

□

[†] $(a)> = \{x : a > x\} = \{x : x < a\}$ (4.1.56).

4.1.77 Example. 0 is *minimal*, also *minimum*, in $\mathbb{N}$ with respect to the natural ordering.

In $\mathscr{P}(\mathbb{N})$, $\emptyset$ is both $\subset$-minimal and $\subset$-minimum.

On the other hand, all of $\{0\}, \{1\}, \{2\}$ are $\subset$-minimal in $\mathscr{P}(\mathbb{N}) - \{\emptyset\}$ but *none* are $\subset$-*minimum* in that set.

Observe from this last example that minimal elements in a class are *not* unique. □

4.1.78 Remark. (Hasse diagrams) Read me! There is a neat pictorial way to depict orders on finite sets known as "*Hasse diagrams*". To do so one creates a so-called "*graph*" of the finite PO set $(A, <)$ where $A = \{a_1, a_2, \ldots, a_n\}$.

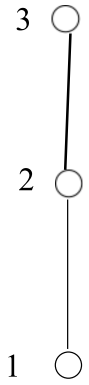
How? The graph consists of n *nodes* —which are drawn as points— each labeled by one a_i. The graph also contains 0 or more *arrows* that connect nodes. These arrows are called *edges*.

When we depict an arbitrary R on a finite set like A we draw *one* arrow (edge) from a_i to a_j iff the two *relate*: $a_i R a_j$.

In Hasse diagrams for PO sets $(A, <)$ we are more selective: We say that b *covers* a iff $a < b$, but there is no c such that $a < c < b$. In a Hasse diagram we will

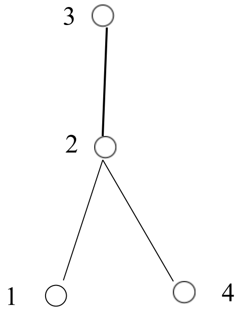
1. draw an edge from a_i to a_j iff a_j covers a_i.
2. by convention we will draw b higher than a on the page if b covers a.
3. given the convention above, using "arrow-heads" is superfluous: our edges are plain line segments.

So, let us have $A = \{1, 2, 3\}$ and $<= \{(1, 2), (1, 3), (2, 3)\}$.



The above has a minimum (1) and a maximum (3) and is clearly a linear order.

A slightly more complex one is this $(A, <)$, where $A = \{1, 2, 3, 4\}$ and $<= \{(1, 2), (4, 2), (2, 3), (1, 3), (4, 3)\}$.



This one has a maximum (3), two minimal elements (1 and 4) but no minimum, and is not a linear order: 1 and 4 are not comparable. □

Feb. 18, 2022

4.1.79 Lemma. *Given an order $<$ and a class $\mathbb{A}$.*
(1) *If m is a* minimum *in $\mathbb{A}$, then it is also* minimal.
(2) *If m is a* minimum *in $\mathbb{A}$, then it is unique.*

Proof. (1) Let m be mini*mum* in $\mathbb{A}$. Then

$$m \le x, \text{ that is, } m = x \lor m < x \tag{i}$$

for all $x \in \mathbb{A}$. Now, prove that there is no $x \in \mathbb{A}$ such that $x < m$.

OK, let us go *by contradiction*:

So let instead, for some a,

$$\mathbb{A} \ni a < m \tag{ii}$$

that is, suppose m is NOT minimal. I also have $m \le a$ by (i), that is,

$$m = a \lor m < a \tag{iii}$$

Now, by irreflexivity, (ii) rules out $a = m$. So, (iii) *nets* $m < a$.

(ii) and (iii) and transitivity yield $a < a$; contradiction ($<$ is irreflexive). Done.

(2) Let m and n both be minima in $\mathbb{A}$. Then $m \leq n$ (with m posing as minimum) and $n \leq m$ (now n is so posing), hence $m = n$ by antisymmetry (Lemma 4.1.67). □

4.1.80 Example. Let m be $<$-minimal in $\mathbb{A}$.

Let us attempt to “show” that it is also $<$-minimum (this is, of course, doomed to *fail* due to 4.1.77 and 4.1.79(2) —but the “faulty proof” below is instructive):

By 4.1.74 we have that there is no x in $\mathbb{A}$ such that $x < m$.

Another way to say this is:

For all $x \in \mathbb{A}$, “$x < m$” is false, that is, I *have* $\neg x < m$. (1)

But from “our previous math” (high school? university? Netflix?)

“$\neg x < m$ is equivalent to $m \le x$”

Thus (1) says $(\forall x \in \mathbb{A}) m \le x$, in other words, m is the minimum in $\mathbb{A}$.

Do you accept this "fact"?

$$\neg x < m \textbf{ is equivalent to } m \leq x \tag{2}$$

Don't!

If (2) were correct, then we would in particular have that

$$\text{if } x < m \text{ is false, then at least one of } x = m \text{ or } x > m \text{ is true} \tag{3}$$

No hope to prove such thing IN GENERAL! See 4.1.65. □ ☡

4.1.81 Lemma. *If $<$ is a* linear *order on $\mathbb{A}$, then every minimal element is also minimum.*

Proof. The "false proof" of the previous example is valid under the present circumstances: (3) on p.151 works in the presence of *trichotomy*. □

The following type of relation has fundamental importance for set theory, and mathematics in general.

4.1.82 Definition.

1. An order $<$ satisfies the *minimal condition*, for short *it has MC*, iff *every* nonempty $\mathbb{A}$ has $<$-minimal elements.

2. If a *total* order $<: \mathbb{B} \to \mathbb{B}$ has MC, then it is called a *well-ordering*[†] *on* (or *of*) the class $\mathbb{B}$.

3. If $(\mathbb{B}, <)$ is a LO class (or set) with MC, then it is a *well-ordered class* (or set), or *WO class* (or WO set).

□

[†] The term "well-ordering" is ungrammatical, but it is *the* terminology established in the literature!

4.1.83 Remark.

What Definition 4.1.82 says in case 1. is —see (2) in 4.1.76— "*if, for some fixed order $<$ the following statement*

$$\emptyset \neq \mathbb{A} \to (\exists a \in \mathbb{A})\mathbb{A} \cap (a) >= \emptyset \tag{1}$$

is true in set theory, for *__any__* $\mathbb{A}$, *then we say that $<$ has MC*".

The following observation is very important *for future reference*:

If $\mathbb{A}$ is given via a defining property $F(x)$, as

$$\mathbb{A} \overset{Def}{=} \{x : F(x)\}$$

then (1) translates —in terms of $F(x)$— into

$$(\exists a)F(a) \to (\exists a)\Big(F(a) \wedge \neg(\exists y)\big(a > y \wedge F(y)\big)\Big) \tag{2'}$$

OR

$$(\exists a)F(a) \to (\exists a)\Big(F(a) \wedge \neg(\exists y)\big(y < a \wedge F(y)\big)\Big) \tag{2}$$

Conversely, for each formula ("property") $F(x)$ we get a class $\mathbb{A} = \{x : F(x)\}$ and thus —if $\mathbb{A}$ has MC with respect to $<$— we may express this fact as in (2) above.

Chapter 5

Functions

Feb. 28, 2022

At last! We consider here a *special case of relations* that we know as “functions”.

Many of you know already that a function is a relation with some special properties.

Let’s make all this official:

5.1. Preliminaries

5.1.1 Definition. A *function* $\mathbb{R}$ is a *single-valued* relation.

That is,

$$\text{whenever we have } \textit{both } x\mathbb{R}y \text{ and } x\mathbb{R}z$$

then

$$\text{we will also have } y = z$$

It is traditional to use, generically, lower case letters from among f, g, h, k when dealing with functions that are sets and $\mathbb{F}, \mathbb{G}, \mathbb{H}, \mathbb{K}$ for functions that are proper classes —with primes and/or subscripts if we run out of letters. □

The above definition does not care about *left* or *right fields*.

5.1.2 Remark. Another way of putting it, using the notation from 4.1.54, is:

A relation $\mathbb{R}$ is a function *iff* $(a)\mathbb{R}$ is either *empty* or contains *exactly one* element.

□

5.1.3 Example. The empty set is a relation of course, the empty set of *pairs*. It is also a function since

$$(x, y) \in \emptyset \wedge (x, z) \in \emptyset \to y = z$$

vacuously, by virtue of the left hand side of $\to$ being false. □

5.1.4 Definition. (Function-specific notations) Let $\mathbb{F}$ be a function.

1. First off, the *concepts AND notation* for domain, range, and —*in case of* a function $\mathbb{F} : \mathbb{A} \to \mathbb{B}$— left field, right field, field, total and onto *are inherited from those for relations without change*.

2. Even the notations "$a\mathbb{R}b$" and "$(a, b) \in \mathbb{R}$" transfer over to functions and are often useful and are employed!.

3. And yet, we have an annoying *difference* in notation:

For a relation $\mathbb{F}$ —or viewing a function $\mathbb{F}$ as a relation— the set

$$\{y : a\mathbb{F}y\} \tag{1}$$

is denoted by $(a)\mathbb{F}$ (see 4.1.54).

If $\mathbb{F}$ is a function, then the set in (1) is either *empty* or has one element only. In *Relational Notation* that is:

$$(a)\mathbb{F} = \begin{cases} \{y\} & \text{if } \mathbb{F} \text{ defined at } a \\ \emptyset & \text{if } \mathbb{F} \text{ undefined at } a \end{cases} \tag{2}$$

The *literature* in general[b] denotes (2) this way

$\mathbb{F}(a)= y$ ⟨note order reversal and brace removal!⟩

$\mathbb{F}(a)\uparrow$ ⟨$\mathbb{F}$ undefined at a⟩

[b] Not all the literature: The significant book [Kur63] writes "af" for (set) functions AND relations, omitting even the brackets around a.

Notation: Thus for a *function* $\mathbb{F}$,

$$a\mathbb{F}y \text{ iff } (a)\mathbb{F} = \{y\} \text{ iff } \boxed{\mathbb{F}(a) = y}$$

and

$$\neg(\exists y)a\mathbb{F}y \text{ iff } (a)\mathbb{F} = \emptyset \text{ iff } \boxed{\mathbb{F}(a)\uparrow}$$

In short, $\mathbb{F}(a)$ reports THE unique ELEMENT IN $\{y : a\mathbb{F}y\} \neq \emptyset$ —*not* the entire class— or reports nothing: $\mathbb{F}(a)\uparrow$ if $\{y : a\mathbb{F}y\} = \emptyset$

□

5.1.5 Example. In particular $\mathbb{F}(a) = \emptyset$ means $(a)\mathbb{F} = \{\emptyset\}$, that is, $(a, \emptyset) \in \mathbb{F}$ or $a\mathbb{F}\emptyset$ —not what one might hastily think it means!

Definitely, $\mathbb{F}(a) \downarrow$ here, with output the object "$\emptyset$", it is NOT $\mathbb{F}(a) \uparrow$

□

5.1.6 Definition. (Images) The class of *all* outputs of a function $\mathbb{F}$, *when the inputs come from a particular class* $\mathbb{X}$, is called the *image of* $\mathbb{X}$ *under* $\mathbb{F}$ and is denoted by $\mathbb{F}[\mathbb{X}]$.

Thus,

$$\mathbb{F}[\mathbb{X}] \overset{Def}{=} \{\mathbb{F}(x) : x \in \mathbb{X}\} \tag{1}$$

Note that careless notation like $\mathbb{F}(A)$ —where A is a set— will *not* do.

This notation means the input *IS* A —*not* members of A.

If I want the inputs to be *from INSIDE* A, then I *must use other than* the round brackets notation; I did.

☡ $\mathbb{F}(\mathbb{X})$ is *meaningless* for a *proper class* $\mathbb{X}$. I cannot have, for example, $(\mathbb{X}, y) \in \mathbb{F}$ since $(\mathbb{X}, y) = \{\mathbb{X}, \{\mathbb{X}, y\}\}$ is meaningless. ☡

The *inverse image* of a class $\mathbb{Y}$ under a function $\mathbb{F}$ is useful as well, that is, the class of *all* inputs that generate $\mathbb{F}$-outputs exclusively in $\mathbb{Y}$.

It is denoted by $\mathbb{F}^{-1}[\mathbb{Y}]$ and is defined as

$$\mathbb{F}^{-1}[\mathbb{Y}] \overset{Def}{=} \{x : \mathbb{F}(x) \in \mathbb{Y}\} \tag{2}$$

□

5.1.7 Theorem. *If $\mathbb{F}$ is a function, and A is a set, then $\mathbb{F}[A]$ is a set.*

Proof. Let

$$\mathbb{Y} = \mathbb{F}[A] \tag{†}$$

Thus, for every y, $y \in \mathbb{Y}$ iff for some $x \in A$, $\mathbb{F}(x) = y$.

In short, each $y \in \mathbb{Y}$ is labelled —in the sense of 4.1.42— by all the $x \in A$ with the property $\mathbb{F}(x) = y$.

Note that the described label-set is *valid* according to 4.1.42 since

- All members of $\mathbb{Y}$ receive labels from A: Indeed, by (†),

  If $y \in \mathbb{Y}$, the *label-set* for y —$A \cap \mathbb{F}^{-1}[\{y\}]$— is *nonempty* (1)

- The set $A \cap \mathbb{F}^{-1}[\{y\}]$ has no repeated members (it is a set!) thus the labels assigned to y are distinct, and more importantly
- If $y \neq y'$, both in $\mathbb{Y}$, then they receive non overlapping labels, that is, $\mathbb{F}^{-1}[\{y\}] \cap \mathbb{F}^{-1}[\{y'\}] = \emptyset$.

  Indeed, if $z \in \mathbb{F}^{-1}[\{y\}] \cap \mathbb{F}^{-1}[\{y'\}]$, then $\mathbb{F}(z) = y$ *and* $\mathbb{F}(z) = y'$; impossible for a function.

By Principle 3, $\mathbb{Y}$ —being labelled by the members of A— is a set too. □

Why did I use $A \cap \mathbb{F}^{-1}[\{y\}]$ instead of just $\mathbb{F}^{-1}[\{y\}]$ for the set of labels for y above?

Well, I wanted to be sure they all come from within A.

We are *not* told that A is the left field of A and said field $\mathbb{X}$ might properly contain $A \subsetneq \mathbb{X}$. In particular we may have an $x \in \mathbb{X} - A$ such that $\mathbb{F}(x) = y$.

This x is NOT a label for y FROM A!

5.1.8 Corollary. *If* $\mathbb{F}$ *is a function and* $\mathrm{dom}(\mathbb{F})$ *is a set, then* $\mathbb{F}$ *is a set.*

Proof. *Exercise*! □

Pause. So far we have been giving definitions regarding functions of *one* variable. Or have we?◀

Not really: We have already said that the multiple-input case is subsumed by our notation. If $\mathbb{F} : \mathbb{A} \to \mathbb{B}$ and $\mathbb{A}$ is a class of n-tuples, then $\mathbb{F}$ is a function of "n-variables".

The binary relation, that such an $\mathbb{F}$ is, contains pairs like $\big((\vec{x}_n), x_{n+1}\big)$.

However, we usually abuse the notation $\mathbb{F}\Big((\vec{x}_n)\Big)$ —or $\Big((\vec{x}_n)\Big)\mathbb{F}$— and write instead $\mathbb{F}(\vec{x}_n)$ —or $(\vec{x}_n)\mathbb{F}$— omitting *the brackets of the n-tuple* $(\vec{x}_n)$.

☡ **5.1.9 Remark.** Regarding, say, the definition of $\mathbb{F}[X]$ (5.1.6):

What if $\mathbb{F}(a)\uparrow$*? How do you "collect" an* undefined "value" *into a class?*

Well, you don't.

Both (1) and (2) in 5.1.6 have a rendering that is *independent* of the notation "$\mathbb{F}(a)$".

$$\mathbb{F}[\mathbb{X}] = \{y : (\exists x \in \mathbb{X})x\mathbb{F}y\} \tag{1'}$$

$$\mathbb{F}^{-1}[\mathbb{Y}] = \{x : (\exists y \in \mathbb{Y})x\mathbb{F}y\} \tag{2'}$$

□ ☡

In view of Theorem 5.1.7, all our functions in the rest of the section will be sets as we adhere to *set* left fields.

5.1.10 Example. Thus, $f[\{a\}] = \{f(x) : x \in \{a\}\} = \{f(x) : x = a\} = \{f(a)\}$.

Let now $g = \Big\{(1, 2), \Big(\{1, 2\}, 2\Big), (2, 7)\Big\}$, clearly a function. Thus, $g(\{1, 2\}) = 2$, but $g[\{1, 2\}] = \{2, 7\}$. Also, $g(5) \uparrow$ and thus $g[\{5\}] = \emptyset$.

On the other hand, $g^{-1}[\{2, 7\}] = \{1, \{1, 2\}, 2\}$ and $g^{-1}[\{2\}] = \{1, \{1, 2\}\}$, while $g^{-1}[\{8\}] = \emptyset$ since no input causes output 8. □

March 2, 2022

5.1.11 Remark. (Kleene Equality) When $f(a)\downarrow$, then $f(a) = f(a)$ as is naturally expected.

What about when $f(a)\uparrow$?

This begs a more general question that we settle as follows (following Kleene, [Kle43]):

When is $f(a) = g(b)$ where f, g are two functions?

$$f(a) = g(b) \overset{Def\ \ ([\text{Kle43}])}{\equiv} f(a)\uparrow \wedge\, g(b)\uparrow \vee (\exists y)\Big(f(a) = y \wedge\ g(b) = y\Big)$$

□

5.1.12 Example. Let $g = \{\langle 1, 2\rangle, \langle \{1, 2\}, 2\rangle, \langle 2, 7\rangle\}$.

Then, $g(1) = g(\{1, 2\})$ and $g(1) \neq g(2)$.

$g(3) = g(4)$ since both sides are undefined. □

5.1.13 Definition. A function f is 1-1 **iff** (i.e., the concept is short for) for all x, y and z, $f(x) = f(y) = z$ implies $x = y$.

Or, in relational notation,

$x f z \wedge y f z \rightarrow x = y$: distinct inputs <u>cannot</u> cause the same <u>output</u>

> So 1-1 ness is about <u>actual outputs</u> being distinct for distinct inputs.

□

Let f be $f : \mathbb{N} \to \mathbb{N}$.

Suppose $f(1)\uparrow$ and also $f(2)\uparrow$.

By 5.1.11, we have $f(1) = f(2)$. But $1 \neq 2$.

This is why we defined 1-1ness (5.1.13) by

$$\text{If } f(x) = f(y) = z, \text{ then } x = y$$

And this is what the literature that is *oblivious to nontotal functions* misses!

5.1.14 Example. $\{\langle 1,1\rangle\}$ and $\{\langle 1,1\rangle,\langle 2,7\rangle\}$ are 1-1: No output that is due to two distinct inputs.

$\emptyset$ is 1-1 vacuously.

$\{\langle 1,0\rangle,\langle 2,0\rangle\}$ is not 1-1. □

5.1.15 Exercise. Prove that if f is a 1-1 function, then the relation *converse* f^{-1} is a function (that is, a single-valued relation). □

5.1.16 Definition. (1-1 Correspondence) A function $f : A \to B$ is called a *1-1 correspondence* iff it is all three: 1-1, total, and onto.

Often we say that A and B are *in 1-1 correspondence* writing $A \sim B$, often omitting mention of the function that *is* the 1-1 correspondence. □

The terminology is derived from the fact that every element of A is paired with precisely one element of B and vice versa.

5.1.17 Exercise. Show that $\sim$ is a *symmetric* and *transitive* relation on sets. □

5.1.18 Remark. Composition of functions is inherited from the composition of relations.

It is the identical concept since a function IS a relation.

Thus, $f \circ g$ for two functions still means

$$x\, f \circ g\, y \text{ iff, for some } z,\ x\, f\, z\, g\, y \tag{1}$$

► **Note!**,
$f \circ g$ is also a function. Indeed, if we have

$$xf \circ gy \text{ and } xf \circ gy'$$

then

$$\text{for some } z, xfzgy \tag{2}$$

and

$$\text{for some } w, xfwgy' \tag{3}$$

As f is a function, (2) and (3) give $z = w$. In turn, this (since g is a function too!) gives $y = y'$. □

The notation (as in 4.1.54) "$(a)f$" for relations is uncommon[†] when applied to functions —but it IS correct— where "$f(a)$" may be more convenient and "normal".

However, the "normal" notation "$f(a)$" is awkward in connection with composition. Consider

$$x \rightarrow \boxed{f} \rightarrow z \rightarrow \boxed{g} \rightarrow y$$

that represents (1) on p.178 above, note that f **acts first**.

Its *result* $z = f(x)$ is then *inputed* to g —that is, we do $g(z) = g\Big(f(x)\Big)$ to obtain output y. Thus the *first* acting function f is "called" first with argument x and then g is called with argument $f(x)$.

[†] See however [Kur63].

Note that if we wrote "$(f \circ g)(a)$" this would imply —wrongly— that g acts first (first call) *being closest to the input!*

*To get around this misleading illusion we need a new notation (**below**) for functional composition*.

5.1.19 Definition. (Salvaging Notation “$f(a)$”)
We just learnt (5.1.18) that the composition of two functions produces a function.

The present definition is *about notation only*.

Let $f : A \to B$ and $g : B \to C$ be two functions. The Notation $f \circ g : A \to C$, their *relational composition*, is the one in 4.1.16.

For composition of *functions*, we have the alternative —so-called *functional notation for composition*:

“gf” stands for “$f \circ g$”; *note the order reversal* AND the absence of “$\circ$”, the composition symbol.

In particular we *write* $(gf)(a)$ *for* $(a)(f \circ g)$ —cf. 5.1.4— placing the input close to the function that uses it.

Thus let f and g be functions, hence as we saw (5.1.18), $f \circ g$ is a function as well.

Therefore

$$\begin{aligned}
\Big(gf\Big)(a) = b \text{ iff } & (a)(f \circ g) = \{b\} \text{ (}\textbf{Box}\text{ on p.181 via the lens of p.161)}\\
\text{iff } & a(f \circ g)b\\
\text{iff } & (a)f = \{c\} \wedge (c)g = \{b\}, \text{ for some } c\\
\text{iff } & f(a) = c \wedge g(c) = b, \text{ for some } c\\
\text{iff } & g\Big(f(a)\Big) = b
\end{aligned}$$

So the notation "gf" above works *if we want the input "(a)" to the right* —see also bottom of p.179. □

5.1.20 Theorem. *Functional composition is associative, that is,*

$$(gf)h = g(fh)$$

Proof. Exercise!

Hint. Note that by, 5.1.19, $(gf)h = h \circ (f \circ g)$. Take it from here.

□

5.1.21 Example. The *identity relation* on a set A is a function since $(a)\mathbf{1}_A$ is the *singleton* —meaning "one-element" set— $\{a\}$. □

The following interesting result connects the notions of ontoness and 1-1ness with the "algebra" of composition.

5.1.22 Theorem. *Let $f : A \to B$ and $g : B \to A$ be functions. If*

$$gf = \mathbf{1}_A \tag{1}$$

then g is onto while f is total and *1-1.*

5.1.23 Definition. Relating to (1) in the theorem above we say that g is a **left inverse** of f and f is a **right inverse** of g.

Using the indefinite article "a" because these are not in general unique! Stay tuned on this!

□

Proof. (**of 5.1.22**)

About g**:** Our goal, *ontoness*, means that, for each $x \in A$, I can "solve the equation $g(y) = x$ for y".

Indeed I can: By definition of $\mathbf{1}_A$,

$$g\Big(f(x)\Big) \overset{5.1.19}{=} (gf)(x) \overset{(1)}{=} \mathbf{1}_A(x) = x$$

So to solve, take $y = f(x)$.

About f: As seen above, $x = g(f(x))$, for each $x \in A$.

Totalness: Since the above remark is *the same as* "$x\, f \circ g\, x$ is true", there must be a z such that $x\, f\, z$ and $z\, g\, x$.

But xfz says $f(x) = z$ and therefore $f(x) \downarrow$. This settles *totalness*.

1-1 ness: For the 1-1ness, let $f(a) = f(b)$.

Recall the definition of 1-1ness (5.1.13) but here it is redundant to say "let $\underline{f(a) = f(b) = c}$".

Applying g to both sides we get $g(f(a)) = g(f(b))$.

or

$$(gf)(a) = (gf)(b)$$

But this says $a = b$, by $gf = \mathbf{1}_A$, and we are done. □

5.1.24 Example. *The above is as much as can be proved.* For example, say $A = \{1, 2\}$ and $B = \{3, 4, 5, 6\}$.

Let $f : A \to B$ be $\{\langle 1, 4\rangle, \langle 2, 3\rangle\}$ and

$g : B \to A$ be $\{\langle 4, 1\rangle, \langle 3, 2\rangle, \langle 6, 1\rangle\}$, or in friendlier notation

$f(1) = 4$
$f(2) = 3$
and
$g(3) = 2$
$g(4) = 1$
$g(5)\uparrow$
$g(6) = 1$

Clearly, $gf = \mathbf{1}_A$ holds, but note:

(1) f is not onto.

(2) g is neither 1-1 nor total. □

☡ **5.1.25 Example.** With $A = \{1, 2\}$, $B = \{3, 4, 5, 6\}$ and $f : A \to B$ and $g : B \to A$ as in the previous example, consider also the functions $\tilde{f}$ and $\tilde{g}$ given by

$$\begin{aligned}
\tilde{f}(1) &= 6\\
\tilde{f}(2) &= 3\\
&\text{and}\\
\tilde{g}(3) &= 2\\
\tilde{g}(4) &= 1\\
\tilde{g}(5) &\uparrow\\
\tilde{g}(6) &= 1
\end{aligned}$$

Clearly, $\tilde{g}f = \mathbf{1}_A$ and $g\tilde{f} = \mathbf{1}_A$ hold, but note:

(1) $f \neq \tilde{f}$.

(2) $g \neq \tilde{g}$.

Thus, neither left nor right inverses need to be unique. The article "a" in the definition of said inverses was well-chosen. □ ☡

The following two partial converses of 5.1.22 are useful.

5.1.26 Theorem. *Let $f : A \to B$ be total and 1-1. Then there is an* onto *$g : B \to A$ such that $gf = \mathbf{1}_A$.*

Proof. Consider the *converse* relation (4.1.52) of f —that is, the *relation* f^{-1}— and call it g:

$$x\, g\, y \overset{\text{Def}}{\text{ iff }} y\, f\, x \tag{1}$$

By Exercise 5.1.15, $g : B \to A$ is a (possibly nontotal) function.

So we can write (1) as $g(x) = y$ iff $f(y) = x$, from which,

substituting $f(y)$ for x in $g(x)$ we get

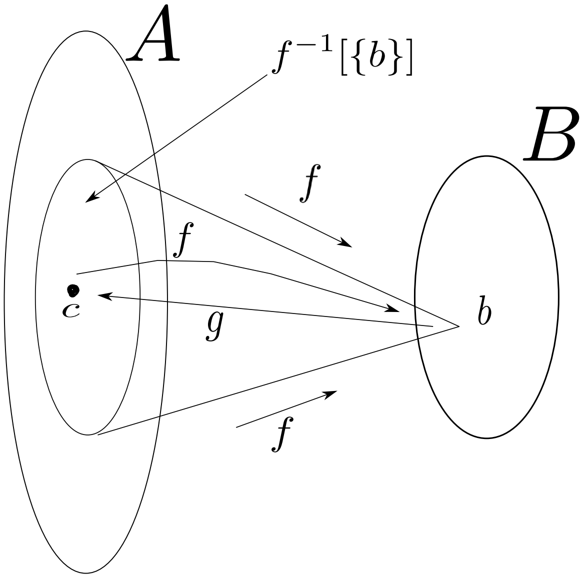
$g(f(y)) = y$, for all $y \in A$, that is $gf = \mathbf{1}_A$, hence g is onto by 5.1.22.

We got both statements that we needed to prove. □

March 4, 2022

5.1.27 Theorem. *Let $f : A \to B$ be onto. Then there is a* total and 1-1 *$g : B \to A$ such that $fg = \mathbf{1}_B$.*

Proof. By assumption, $\emptyset \neq f^{-1}[\{b\}] \subseteq A$, for all $b \in B$.



To define $g(b)$ **choose** *one* $c \in f^{-1}[\{b\}]$ and set $g(b) = c$.

► Do so for *each* $b \in B$. ◄

Since $f(c) = b$, we get $f(g(b)) = b$ for all $b \in B$, that is, $fg = \mathbf{1}_B$.

Hence g is 1-1 and total by 5.1.22. □

☡☡ **5.1.28 Remark. (Axiom of Choice)** The proof of 5.1.27 states

$$\underline{\textbf{choose}}\ \textit{one}\ c \in f^{-1}[\{b\}]$$

and that must be done for all $b \in B$ that may be *infinitely many*.

But how do you choose "the" c? If we were dealing with natural numbers I can see that (How?).

But not with the reals and not with arbitrary unspecified sets!

How do you DESCRIBE in a mathematical way the process of choosing ONE element out of each of (potentially) infinitely many nonempty sets?

How —for example (due to Russell)— do you describe the process of choosing ONE sock from each of infinitely many pairs?

True, you might sit there for an infinite amount of time, and pick ONE sock at random from each pair. But can you sit that long?

Incidentally you could describe a process of choosing from an infinite set of pairs of shoes!

In set theory one takes as an axiom that a SET of (results of) c-choices exists! They call it the "Axiom of Choice". □ ☡☡

5.2. Finite and Infinite Sets

Broadly speaking (that is, with very little detail contained in what I will say next) we have sets that are *finite* —intuitively meaning that we can "count" *all* their elements in a "finite amount" of "time" (but see the ☡-remark 5.2.3 below)— and those that are not, the *infinite* sets!

What is a mathematical way to say all this?

Any *counting process* of the elements of a finite set A will have us say out loud —every time we *pick*, or *point* at, an element of A— "0th", "1st", "2nd", etc.,

Once we reach and pick the *last* element of the set, we finally pronounce "nth", for some appropriate n that we reached in our counting (Again, see 5.2.3.)

Thus, mathematically, we *are pairing* each member of the set —or *label* each member of the set— with a member from $\{0, \ldots, n\}$.

Thus the following makes sense:

5.2.1 Definition. (Finite and infinite sets) A set A is *finite* iff it is either empty, or is in 1-1 correspondence with $\{x \in \mathbb{N} : x \leq n\}$. This "normalised" "small" set of natural numbers we usually denote by $\{0, 1, 2, \ldots, n\}$.

If a set is *not* finite, then it is —by definition— *infinite*. □

5.2.2 Example. For any n, $\{0, \ldots, n\}$ is finite since, trivially,

$$\{0, \ldots, n\} \sim \{0, \ldots, n\}$$

using the identity (Δ) function on the set $\{0, \ldots, n\}$. □

5.2.3 Remark. One must be careful when one attempts to explain finiteness via counting by a human.

For example, Achilles[†] could count *infinitely many objects* by constantly accelerating his counting process as follows:

He procrastinated for a *full second*, and then counted the first element. Then, he counted the second object *exactly after* $1/2$ a second from the first. Then he got to the third element $1/2^2$ seconds after the previous, ..., he counted the n th item at exactly $1/2^{n-1}$ seconds after the previous, and so on *forever*.

Hmm! It was *not* "forever", was it? After a total of 2 seconds he was done!

You see (as you can easily verify from your calculus knowledge (limits)),[‡]

$$1+\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^{n-1}}+\ldots=\frac{1}{1-1/2}=2$$

So "clock-time" is *not* a good determinant of finiteness! □

[†] OK, he was a demigod; but only "demi".

[‡] $1+\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^{n-1}}=\frac{1-1/2^n}{1-1/2}$. Now let n go to infinity at the limit.

March 7, 2022

5.2.4 Theorem. *If $X \subsetneq \{0, \ldots, n\}$, then there is no onto function $f : X \to \{0, \ldots, n\}$.*

☡ I am saying, *no such f, whether total or not*; totalness is immaterial. ☡

Proof. First off, the claim holds if $X = \emptyset$, since then any such f equals $\emptyset$ —no inputs, therefore no outputs!

The range of f is empty so it cannot be onto.

Let us otherwise proceed by way of contradiction, and assume that the theorem is *wrong*.

That is, **assume that** it *IS* possible to have such onto functions, for *some* n and *well-chosen* X.

So let n_0 be the *smallest* n that *contradicts* the theorem, and let X_0 be a *corresponding* set "X" that supports the contradiction, that is,

$$X_0 \subsetneq \{0, \ldots, n_0\} \text{ AND } f : X_0 \to \{0, \ldots, n_0\} \text{ is } \textit{onto} \qquad (1)$$

Firstly, we saw that $X_0 \neq \emptyset$, since $X_0 = \emptyset$ *does NOT FAIL the theorem*.

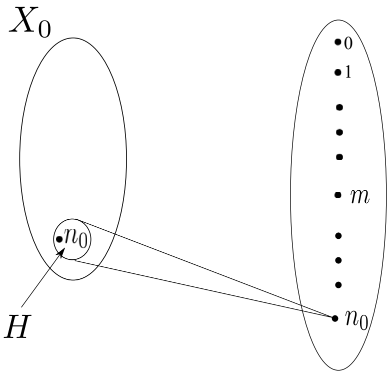
Secondly, $n_0 > 0$, since otherwise —i.e., *IF* $n_0 = 0$— $X_0 = \emptyset$ (Why?) and, as remarked, the latter *does NOT FAIL the theorem*.

Let us set $H = f^{-1}[\{n_0\}]$.

$\emptyset \neq H \subseteq X_0$; the $\neq$ by ontoness.

Case 1. $n_0 \in H$. Then removing all pairs (a, n_0) from f —all these have $a \in H$— we get a new function $f' : X_0 - H \to \{0, 1, \ldots, n_0 - 1\}$, which *is still onto* as we only removed inputs that cause output n_0 —and thus contradicts the theorem.

This contradicts minimality of n_0 since $n_0 - 1$ works too!



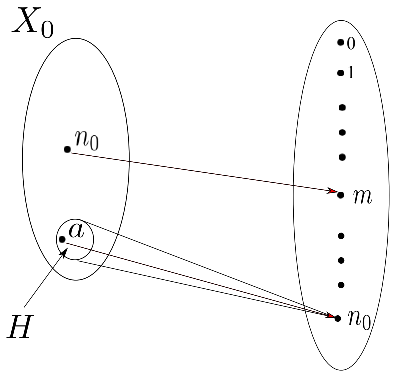
Case 2. $n_0 \notin H$.

(a) Subcase $n_0 \notin X_0$. So $X_0 \subsetneqq \{0, 1, \ldots, n_0\}$, thus $X_0 \subseteq \{0, 1, \ldots, n_0 - 1\}$. By $H \neq \emptyset$, $X_0 - H \subsetneqq \{0, 1, \ldots, n_0 - 1\}$. As in Case 1, $f' : X_0 - H \to \{0, 1, \ldots, n_0 - 1\}$ is onto. **NOTE** that $f(n_0)\uparrow$ in this case.

(b) $n_0 \in X_0$. We have two subcases:

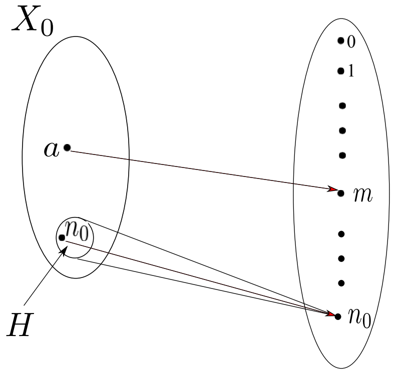
- $f(n_0)\uparrow$. Then we (almost) act as in Case 2(a):
  The new "X_0" is $(X_0 - H) - \{n_0\}$.
  We remove $n_0 \in X_0$ to ensure that the new "X_0" *will* be a subset of $\{0, 1, \ldots, n_0 - 1\}$ and *we get a contradiction exactly per Case 2(a)*.

- We have the picture below —that is, $f(n_0) = m \neq n_0$ for some m.



We simply transform the picture to the one below, "correcting" f to have $f(a) = m$ and $f(n_0) = n_0$, that is defining a new "f" that we will call f' by

$$f' = \Big(f - \{(n_0, m), (a, n_0)\}\Big) \cup \{(n_0, n_0), (a, m)\}$$



We get a contradiction per Case 1. □

5.2.5 Corollary. (Pigeon-Hole Principle) *If* $m < n$, *then* $\{0, \ldots, m\} \not\sim \{0, \ldots, n\}$.

Proof. If the conclusion fails then we have an onto $f : \{0, \ldots, m\} \to \{0, \ldots, n\}$, contradicting 5.2.4. □

Important!

5.2.6 Theorem. *If A is finite due to $A \sim \{0, 1, 2, \ldots n\}$ then there is* ***no justification of finiteness via another canonical set*** *$\{0, 1, 2, \ldots m\}$ with $n \neq m$.*

Proof. If $\{0, 1, 2, \ldots n\} \sim A \sim \{0, 1, 2, \ldots m\}$, then $\{0, 1, 2, \ldots n\} \sim \{0, 1, 2, \ldots m\}$ by 5.1.17, hence $n = m$, otherwise we contradict 5.2.5. □

5.2.7 Definition. Let $A \sim \{0, \ldots, n\}$. Since n is uniquely determined by A we say that A has $n + 1$ elements and write $|A| = n + 1$. □

March 11, 2022

5.2.8 Corollary. *There is no onto function from* $\{0, \ldots, n\}$ *to* $\mathbb{N}$.

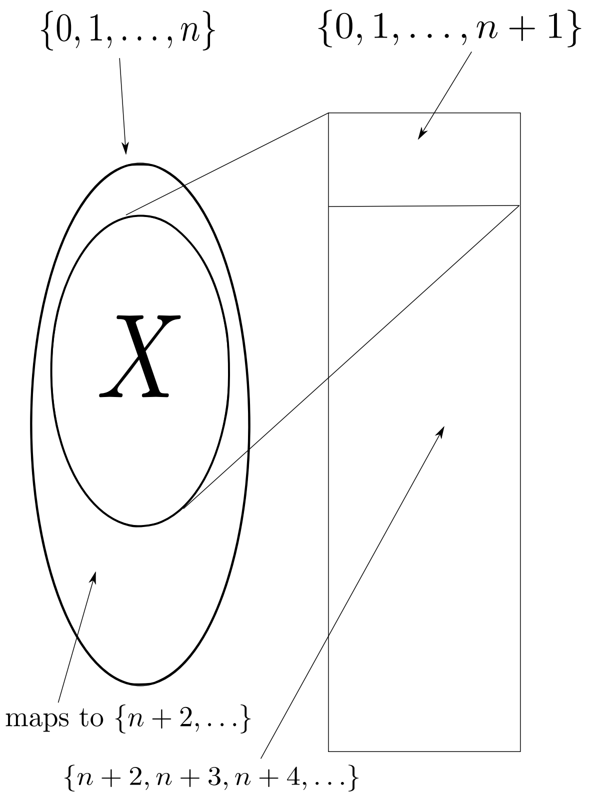
“$\underline{\text{For all } n \in \mathbb{N}}$, there is no…” is, of course, implied.

Proof. Fix an n. By way of contradiction, let $g : \{0, \ldots, n\} \to \mathbb{N}$ be onto.

Let X be the set of all inputs that map *onto* $\{0, \ldots, n+1\}$

$$X \overset{Def}{=} g^{-1}[\{0, 1, \ldots, n+1\}] \subseteq \{0, 1, \ldots, n\}$$

The existence of a g from $\{0, \ldots, n\}$ *onto* $\mathbb{N}$ implies an onto function from X to $\{0, \ldots, n+1\}$. But $X \subsetneq \{0, \ldots, n+1\}$. See the figure below.



This contradicts 5.2.4. □

5.2.9 Corollary. $\mathbb{N}$ *is infinite.*

Proof. By 5.2.1 the opposite case requires that there is an n and a function $f : \{0, 1, 2, \ldots, n\} \to \mathbb{N}$ that is a 1-1 correspondence. Impossible, since any such an f will fail to be *onto* $\mathbb{N}$. □

Our mathematical definitions have led to what we hoped they would:

For example, that $\mathbb{N}$ *is* infinite as we intuitively understand, notwithstanding Achilles's accelerated counting!

$\mathbb{N}$ is a "canonical" infinite set that we can use to *index* or *label* the members of many infinite sets.

Sets that can be indexed using natural number indices

$$a_0, a_1, \ldots$$

are called *countable*.

In the interest of *technical flexibility*, *we do not insist* that *all* members of $\mathbb{N}$ be used as indices.

We might enumerate with *gaps*:

$$b_5, b_9, b_{13}, b_{42}, \ldots$$

Thus, informally, a set A is *countable* if it is empty or (in the opposite case) if there is a way to index, hence enumerate, all its members in an array, utilising indices from $\mathbb{N}$. *See also 4.1.42 regarding indexing/labelling*.

It *is* allowed to *repeatedly list any element of A*, so that finite sets *are* countable.

For example, the set $\{42\}$:

$$42, 42, 42, \overbrace{\ldots}^{\text{42 forever}}$$

We may think that the enumeration above is done by assigning to "42" *all* of the members of $\mathbb{N}$ as indices, in other words, the enumeration is effected, for example, by the constant function $f : \mathbb{N} \to \{42\}$ given by $f(n) = 42$ for all $n \in \mathbb{N}$.

This is consistent with our earlier definition of indexing (4.1.42).

Now, mathematically,

5.2.10 Definition. (Countable Sets) We call a set A *countable* if $A = \emptyset$, or there is an *onto* function $f : \mathbb{N} \to A$.

We *do NOT* require f to be *total*.

This means that some or many indices from $\mathbb{N}$ need not be used in the enumeration.

If $f(n)\downarrow$, then we say that $f(n)$ is the nth element of A in the enumeration f.

We often write f_n instead of $f(n)$ and then call n a "subscript" or "index". □

Thus a nonempty set is countable iff it is the *range* of some function that has $\mathbb{N}$ as its *left field*.

BTW, since we allow f to be *nontotal*, the separate case "nonempty" in the Definition is unnecessary: $\emptyset$ *is* the range of the empty function that has $\mathbb{N}$ as its left field.

We said that the f that proves countability of a set A need not be total.

But such an f can always be "completed", by adding pairs to it, to get an f' such that $f' : \mathbb{N} \to A$ is onto *and* total. Here is how:

5.2.11 Proposition. *Let $f : \mathbb{N} \to A \neq \emptyset$*[†] *be onto. Then we can extend f to f' so that $f' : \mathbb{N} \to A$ is onto* and *total.*

Proof. Pick an $a \in A$ —possible since $A \neq \emptyset$— and keep it fixed. Now, our sought f' is given for all $n \in \mathbb{N}$ by cases as below:

$$f'(n) = \begin{cases} f(n) & \text{if } f(n) \downarrow \\ a & \text{if } f(n) \uparrow \end{cases}$$

□

[†] Since we are constructing a *total* onto function to A we need to assume the case $A \neq \emptyset$ as we cannot put any outputs into $\emptyset$.

Some set theorists also define sets that can be enumerated using *all* the elements of $\mathbb{N}$ as indices *without repetitions*.

5.2.12 Definition. (Enumerable or denumerable sets) A set A is *enumerable* iff $A \sim \mathbb{N}$. □

5.2.13 Example. Every enumerable set is countable, but the converse fails. For example, $\{1\}$ is countable but not enumerable due to 5.2.8.

$\{2n : n \in \mathbb{N}\}$ is enumerable, with $f(n) = 2n$ effecting the 1-1 correspondence $f : \mathbb{N} \to \{2n : n \in \mathbb{N}\}$. □

5.2.14 Theorem. *If A is an infinite subset of $\mathbb{N}$, then $A \sim \mathbb{N}$.*

Proof. We will build a 1-1 and total enumeration of A, presented in a finite manner as a (pseudo) program below, which enumerates all the members of A in strict ascending order and arranges them in an array

$$a(0), a(1), a(2), \ldots a(k-1), \ldots \tag{1}$$

```
$n$          $\leftarrow 0$
while        $A \neq \emptyset$
    $a(n)$   $\leftarrow \min A$  Comment. Inside the loop
                 $\emptyset \neq A \subseteq \mathbb{N}$, hence min exists.
    $A$      $\leftarrow A - \{a(n)\}$
    $n$      $\leftarrow n + 1$
end while
```

Note that the sequence $\{a(0), a(1), \ldots, a(m)\}$ is **strictly increasing** for any m, since $a(0)$ is smallest in A, $a(1)$ is smallest in $A - \{a(0)\}$ and hence the next higher than $a(0)$ in A, etc.

Will this loop ever exit?

Suppose that it does precisely at the time it *starts* (but does not complete: *loop detects* "A= $\emptyset$") the k-th pass through the loop.

Thus A became empty when we did $A \leftarrow A - \{a(k-1)\}$ in the *previous pass*, that is

$$A = \{a(0), a(1), \ldots, a(k-1)\}$$

and thus, since

$$a(0) < a(1) < \ldots < a(k-1)$$

we have that the function $f : \{0, \ldots, k-1\} \to A$ given by

$$f = \{(0, a(0)), (1, a(1)), \ldots (k-1, a(k-1))\}$$

is total, 1-1 and onto, thus, $A \sim \{0, \ldots, k-1\}$ **contradicting that** A **is infinite**!

Thus, we never exit the loop! We do fill the array "a"

Therefore, by the remark in the ☡ paragraph above, (1) enumerates A in strict ascending order, i.e.,

$$\text{if we define } f : \mathbb{N} \to A \text{ by } f(n) = a(n), \text{ for all } n$$

then f is 1-1 (by strict increasing property: distinct inputs cause distinct outputs), and is trivially total, and onto.

Why the latter?

Every $a \in A$ is reached in ascending order, and assigned an "n" from $\mathbb{N}$. □

5.2.15 Theorem. *Every infinite countable set is enumerable.*

Proof. Let $f : \mathbb{N} \to A$ be onto and total (cf. 5.2.11), where A is infinite.

Let $g : A \to \mathbb{N}$ such that $fg = \mathbf{1}_A$ (5.1.27).

Thus, g is *total* and 1-1 and moreover is *onto* $B = \mathrm{ran}(g)$.

We have the following configuration:

$$A \overset{\overset{g}{\to}}{\sim} B \subseteq \mathbb{N} \xrightarrow{f} A \tag{1}$$

Hence B is *infinite*, else we would have $A \sim B \sim \{0, \ldots, n\}$ (some n) and thus $A \sim \{0, \ldots, n\}$ (see Exercise 5.1.17) making A *finite*!

Thus, by 5.2.14, $B \sim \mathbb{N}$, hence $A \sim \mathbb{N}$ via $A \sim B \sim \mathbb{N}$ and 5.1.17 once more. □

So, if we can enumerate an infinite set at all, then we can enumerate it without repetitions.

We can linearise an infinite square matrix of elements in each location (i, j) by devising a traversal that will go through each (i, j) entry *once*, and will *not miss any entry*!

In the literature one often sees the method diagrammatically, see below, where arrows *clearly* indicate the sequence of traversing, with the understanding that we use the arrows by picking the first unused chain of arrows from left to right.

$$\begin{tikzcd}
(0,0) & (0,1) & (0,2) & (0,3) & \ldots \\
(1,0) \arrow[ur] & (1,1) \arrow[ur] & (1,2) \arrow[ur] & & \\
(2,0) \arrow[ur] & (2,1) \arrow[ur] & & & \\
(3,0) \arrow[ur] & & & & \\
\vdots & & & &
\end{tikzcd}$$

So the linearisation induces a 1-1 correspondence between $\mathbb{N}$ and the linearised sequence of matrix entries, that is, it shows that $\mathbb{N} \times \mathbb{N} \sim \mathbb{N}$.

In short,

5.2.16 Theorem. *The set $\mathbb{N} \times \mathbb{N}$ is countable. In fact, it is enumerable.*

Is there a "mathematical" way to do this? Well, the above IS mathematical, don't get me wrong, but is given in *outline*. It is kind of an argument in geometry, where we rely on drawings (figures).

READ ME! Here are the algebraic details:

Proof. (of 5.2.16 with an algebraic argument). Let us call $i+j+1$ the "*weight*" of a pair (i,j). The weight is the number of elements in the group:

$$(i+j,0),(i+j-1,1),(i+j-2,2),\ldots,(i,j),\ldots,(0,i+j)$$

Thus the diagrammatic enumeration proceeds by enumerating *groups* by increasing weight

$$1,2,3,4,5,\ldots$$

and in each group of weight k we enumerate in *ascending order of the second component.*

Thus the (i,j) th entry occupies position j *in its group* —the first position in the group being the 0 th, e.g., in the group of $(3,0)$ the first position is the 0 th— and this position *globally* is the number of elements in all groups *before* group $i+j+1$, *plus* j. Thus the first available position for the first entry of group (i,j) members is just after this many occupied positions:

$$1+2+3+\ldots(i+j)=\frac{(i+j)(i+j+1)}{2}$$

That is,

$$\text{global position of } (i,j) \text{ is this: } \frac{(i+j)(i+j+1)}{2}+j$$

The function f which for all i,j is given by

$$f(i,j)=\frac{(i+j)(i+j+1)}{2}+j$$

is the algebraic form of the above enumeration. □

There is an easier way to show that $\mathbb{N} \times \mathbb{N} \sim \mathbb{N}$ without diagrams:

By the unique factorisation of numbers into products of primes (Euclid) the function

$g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given for all m, n by $g(m, n) = 2^m 3^n$ is 1-1, since Euclid proved that $2^m 3^n = 2^{m'} 3^{n'}$ implies $m = m'$ and $n = n'$.

It is not onto as it never outputs, say, 5, but $\mathrm{ran}(g)$ is an *infinite* subset of $\mathbb{N}$ (Exercise!).

Thus, trivially,

$$\mathbb{N} \times \mathbb{N} \overset{via\ g}{\sim} \mathrm{ran}(g) \sim \mathbb{N}$$

the 2nd "$\sim$" by 5.2.14. END READ ME!

5.2.17 Exercise. If A and B are enumerable, so is $A \times B$.

Hint. So, $\mathbb{N} \sim A$ and $\mathbb{N} \sim B$. Can you show now that $\mathbb{N}\times\mathbb{N} \sim A\times B$? □

With little additional effort one can generalise to the case of $\bigtimes_{i=1}^{n} A_i$. ☡

5.2.18 Remark.

1. Let us collect a few more remarks on countable sets here. Suppose now that we start with a countable set A. Is every subset of A countable?

   Yes, because the composition of onto functions is onto. **Exercise!**

   5.2.19 Exercise. *What does composition of onto functions have to do with this?* Well, if $B \subseteq A$ then there is a *natural* onto function $g : A \to B$. Which one?

   Hint. Think "natural"! Get a *natural* total and 1-1 function $f : B \to A$ and then use f to get g. □

2. As a special case, if A is countable, then so is $A \cap B$ for any B, since $A \cap B \subseteq A$.

3. How about $A \cup B$? If both A and B are countable, then so is $A \cup B$. Indeed, and without inventing a new technique, let

$$a_0, a_1, \ldots$$

be an enumeration of A and

$$b_0, b_1, \ldots$$

for B. Now form an infinite matrix with the A-enumeration as the 1st row, while each remaining row is the same as the B-enumeration. Now linearise this matrix!

Of course, we may alternatively adapt the unfolding technique to an infinite matrix of just two rows. **How?**

4. **5.2.20 Exercise.** Let A be enumerable and an enumeration of A

$$a_0, a_1, a_2, \ldots \tag{1}$$

is given.

So, this is an enumeration without repetitions.

Use techniques we employed in this section to propose a new enumeration in which *every* a_i is listed *infinitely many times* (this is useful in some applications of logic). □

March 14, 2022

5.3. Diagonalisation and uncountable sets

5.3.1 Example. Suppose we have a 3×3 matrix

$$\begin{array}{ccc}1 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 1\end{array}$$

and we are asked:

Find a sequence of three numbers, *using only* 0 *or* 1, that does not *fit* as a row of the above matrix —i.e., is *different from all rows.*

Sure, you reply: Take 1 1 1. Or, take 0 0 0.

That is correct. But what if the matrix were big, say, $10^{350000} \times 10^{350000}$, or even *infinite*?

Is there a *finitely describable technique* that can produce an "unfit" row for *any* square matrix, even an *infinite* one?

□

Yes, it is Cantor's *diagonal method* or technique.

5.3.2 Definition. (Diagonalisation: How-to) Cantor noticed that *any row that fits in a matrix* M as the, say, i-th row, *intersects* the *main diagonal* —i.e., *the list of* $M(x, x)$ *entries*— at the same spot that the i-th column does.

That is, at entry $M(i, i)$.

Thus if we take the main diagonal —*a sequence that has the same length as any row*— and *make a copy of it changing every one of the original entries*, then it will *not* fit *anywhere in* M as a row!

Thus the *Main (Original) Diagonal* is the *sequence* of entries below:

$$\begin{array}{cccccc}
pos.\ 0 & pos.\ 1 & pos.\ 2 & & pos.\ i & \\
\downarrow & \downarrow & \downarrow & & \downarrow & \\
M(0,0), & M(1,1), & M(2,2), & \ldots, & M(i,i), & \ldots
\end{array}$$

The modified diagonal is (where we named "D" the array below):

$$\begin{array}{ccccccc}
 & pos.\ 0 & pos.\ 1 & pos.\ 2 & & pos.\ i & \\
 & \downarrow & \downarrow & \downarrow & & \downarrow & \\
D = & \overline{M(0,0)}, & \overline{M(1,1)}, & \overline{M(2,2)}, & \ldots, & \overline{M(i,i)}, & \ldots
\end{array}$$

where, *for all positions i*, $\overline{M(i,i)} \neq M(i,i)$.

Thus if D fits as row x, then the x-th element of D —$\overline{M(x,x)}$— will overlap the (original) x-th element of M —$M(x,x)$.

But these two are *different*! So, D does NOT FIT as the x-th row!

□

This HOW TO would give the alternative answer $0 \quad 1 \quad 0$ to our original question in 5.3.1.

5.3.3 Example. We have an infinite *matrix* M of 0-1 entries. Can we produce an infinite *sequence* of 0-1 entries that does not match *any* row in the matrix?

Yes, to get the counterpart of D above just define for all x:

$$\overline{M(x,x)} = 1 - M(x,x)$$

In words, take the main diagonal and *flip every entry* (0 to 1; 1 to 0).

Now refer to 5.3.2. □

5.3.4 Example. (Cantor) Let S denote the set of all *infinite sequences* —also called *infinite strings*— of 0s and 1s.

Pause. What is an *infinite sequence*? Our intuitive understanding of the term is captured mathematically by the concept of a total function f with left field (and hence domain) $\mathbb{N}$.

The n-th member of the sequence is $f(n)$.◀

Can we arrange *all* of S in an *infinite matrix*—one element per row?

No, since the preceding example shows that we would miss at least one infinite sequence (i.e., we *would fail to list it as a row*), because a sequence of infinitely many 0s and/or 1s can be found, that does not match any row!

□

But arranging all members of S as an infinite matrix —one element per row— is tantamount to saying that we can enumerate all the members of S using members of $\mathbb{N}$ as indices.

So we cannot do that. S is not countable!

5.3.5 Definition. (Uncountable Sets) A set that is *not* countable is called *uncountable*. □

So, an uncountable set is *neither* finite, *nor* enumerable —because "countable" means finite or enumerable.

Example 5.3.4 shows that uncountable sets exist. Here is a more interesting one.

5.3.6 Example. (Cantor) The set of real numbers in the interval

$$(0,1) \overset{\text{Def}}{=} \{x \in \mathbb{R} : 0 < x < 1\}$$

is uncountable. This is done via an elaboration of the argument in 5.3.4.

Think of a member of $(0,1)$, *in form*, as an infinite sequence of numbers from the set $\{0,1,2,3,4,5,6,7,8,9\}$ prefixed with a dot; that is, think of the number's decimal notation.

Some numbers have representations that end in 0s after a certain point. We call these representations *finite*. Every such number has also an "*infinite representation*" since the non zero digit d immediately to the left of the infinite tail of 0s can be converted to $d-1$, and the infinite tail into 9s, without changing the value of the number.

Allow only infinite representations.

Assume now *by way of contradiction* that a listing of all members of $(0,1)$ exists, *listing them via their infinite representations* —where the leading decimal point is omitted and all a_{ij} satisfy $0 \leq a_{ij} \leq 9$ (decimal digits).

$$\begin{array}{c} a_{00}a_{01}a_{02}a_{03}a_{04}\ldots \\ a_{10}a_{11}a_{12}a_{13}a_{14}\ldots \\ a_{20}a_{21}a_{22}a_{23}a_{24}\ldots \\ a_{30}a_{31}a_{32}a_{33}a_{34}\ldots \\ \vdots \end{array} \tag{1}$$

The "How To" of Definition 5.3.2 is applied now to obtain a

number

$$D = (.)\overline{a_{00}}\,\overline{a_{11}}\,\overline{a_{22}}\,\ldots\overline{a_{xx}}\,\ldots$$

where

$$\overline{a_{xx}} = \begin{cases} 2 & \text{if } a_{xx} = 0 \vee a_{xx} = 1 \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

Clearly (by 5.3.2) D does not fit in *any row i of (1)*, that is, the number it represents *is both*

- *In* $(0,1)$ —since its digits are 1 or 2 it is $0 < D < 1$,

  AND

- *Not in* $(0,1)$ —by the diagonalisation in (2).

This contradiction shows that we do *NOT* have the *enumeration* of all of $(0,1)$ depicted as (1): *The real interval is uncountable*. □

5.3.7 Example. (5.3.4 Revisited) Consider the set of *all* total functions from $\mathbb{N}$ to $\{0, 1\}$. Is this countable?

Well, if there is an enumeration of these one-variable functions

$$f_0, f_1, f_2, f_3, \ldots \tag{1}$$

consider the function $g : \mathbb{N} \to \{0, 1\}$ given by $g(x) = 1 - f_x(x)$.

Clearly, this *must* appear in the listing (1) since it has the correct left and right fields, and is total.

Too bad! If $g = f_i$ then $g(i) = f_i(i)$. By definition, it is also $1 - f_i(i)$.

A contradiction.

This is just version of 5.3.4; as already noted there, an infinite sequence of 0s and 1s is *just another way of viewing* a total function from $\mathbb{N}$ to $\{0, 1\}$. □

The *same* argument as above shows that the set of all functions from $\mathbb{N}$ to $\mathbb{N}$ is uncountable.

Taking $g(x) = f_x(x) + 1$ also works here to “systematically change the diagonal” $f_0(0), f_1(1), \ldots$ since we are not constrained to keep the function values in $\{0, 1\}$.

5.3.8 Remark. Worth Emphasizing. Here is how we constructed g: We have a set of *in principle available f-indices* for g —i.e., in principle, g is a f_i, for some i.

We want to make sure that *none of the indices applies* —i.e., that in fact g is *NOT* f_i for any such index.

A convenient method to do that is to inspect each available index, i, and using the diagonal method do this: *Ensure that g differs from f_i at input i*, by setting

$$g(i) = 1 - \overset{\textit{diag. entry}}{f_i(i)} \tag{1}$$

What we did: We think of F is the *name* of an infinite matrix. $F(i,j) = f_i(j)$ by definition (of F).

Thus (1) above defines an *altered* diagonal "g" for F that cannot fit as a *row* —as an "f_i" (5.3.2).

(1) ensures that $g \neq f_i$; period.

We say that *we cancelled the index* i as a *possible* "f-index" of g.

Since the process is applied *for each* i, *we have cancelled* all *possible indices for* g: For no i can we have $g = f_i$. □

5.3.9 Example. (Cantor) What about the set of all subsets of $\mathbb{N}$ — $\mathcal{P}(\mathbb{N})$ or $2^{\mathbb{N}}$?

Cantor showed that this is uncountable as well: If not, we have an enumeration of its members as

$$S_0, S_1, S_2, \ldots \tag{1}$$

Define the set

$$D \overset{Def}{=} \{x \in \mathbb{N} : x \notin S_x\} \tag{2}$$

So, $D \subseteq \mathbb{N}$, thus it must appear in the list (1) as an S_i. But then

$$i \in D \text{ iff } i \in S_i$$

by virtue of $D = S_i$.

However, also $i \in D$ iff $i \notin S_i$ by (2).

This contradiction establishes that a *legitimate subset of* $\mathbb{N}$*, namely* D*, is* not *an* S_i*.*

That is, $2^{\mathbb{N}}$ *cannot* be so enumerated; it is uncountable. □

☡ **5.3.10 Example. (Characteristic functions)** Let $S \subseteq \mathbb{N}$. We can *represent* S as a total function $c_S : \mathbb{N} \to \{0, 1\}$ given by:

$$c_S(x) = \begin{cases} 1 & \textbf{if } x \in S \\ 0 & \textbf{if } x \in \mathbb{N} - S \end{cases}$$

But then, if we write

$$c_i \text{ for } c_{S_i}$$

With reference to the "D" of 5.3.9, we note that

$$\begin{aligned} c_D(x) &= 1 \text{ iff } x \in D \\ &= 1 \text{ iff } x \notin S_x \ \langle \underline{\text{Comment}}\text{: (2) in 5.3.9} \rangle \\ &= 1 \text{ iff } c_x(x) = 0 \end{aligned}$$

So

$$c_D(x) = 1 - c_{S_x}(x)$$

Thus the argument in 5.3.9 is not new then, but is rather a "translation" of the *diagonalisation* over the list of 0/1 functions c_x, for $x = 0, 1, 2, \ldots$, in the manner of 5.3.7. □ ☡

Chapter 6

A Short Course on Predicate Logic

March 16, 2022

We have become somewhat proficient in using informal logic in our arguments about aspects of discrete mathematics, in particular proving statements like $\mathbb{A} \subseteq \mathbb{B}$ and $\mathbb{X} = \mathbb{Y}$, for any classes that we know something about their properties.

Although we have used quantifiers already —$\exists$ and $\forall$— we did so mostly viewing them as *symbolic abbreviations* of *English texts* about mathematics.

In this chapter we will expand our techniques in logic, extending them to include the correct syntactic —also called "formal"— manipulation of quantifiers.

This chapter also includes the WHAT and the HOW TO of the versatile *Induction* —or *mathematical induction*— technique used to prove properties of the natural numbers.

We know how to detect fallacious statements formulated in Boolean logic: Simply show by a truth table that the statement is not a tautology (or not a so-called *tautological implication*).

Correspondingly, we will show in the domain of quantifier logic not only how to *prove* statements that include quantifiers but also how to *disprove* false statements that happen to include quantifiers.

6.1. Enriching our proofs to manipulate quantifiers

Manipulation of quantifiers boils down to "*how can I remove a quantifier from the beginning of a formula?*" and "*how can I add a quantifier at the beginning of a formula?*"

Once we learn this technique we will be able to reason within mathematics with ease.

But first let us define once and for all what a mathematical proof *looks like*: its *correct*, *expected* *syntax* or *form*.

We will need some concepts to begin with.

1. The alphabet and structure of formulas. Formulas are strings and name statements of mathematics and computer science.

   The alphabet of *symbols* that we use to write down formulas contain, *at a minimum*,

   $$=, \neg, \wedge, \vee, \rightarrow, \equiv, (\,,\,), \forall, \exists, {}^{*}\text{object variables}^{\dagger}$$

   We *finitely generate* the infinite set of *object variables* using single letters, if necessary with primes and/or subscripts: $A, x, y'', w'''_{23}, u_{501}$.

[*] ∃ is introduced as an abbreviation of something more complex in 6.2.2.

[†] That is, variables that denote *objects* such as numbers, arrays, matrices, sets, trees, etc.

2. One normally works in a mathematical area of interest, or *mathematical theory* —such as Geometry, Set Theory, Number Theory, Algebra, Calculus, Theory of Computation— where one needs *additional symbols* to write down formulas, like

$$0, \emptyset, \in, \subseteq, \subsetneqq, \bigcap, \bigcup, \cup, \int, \circ, +, \times, \mu$$

   and many others.

3. Mathematicians as a rule get to recognise the *formulas (name statements)* and *terms (name objects)* in the math areas of their interest via practise without being necessarily taught the recursive definition of the syntax of these.

We will not spell out the syntax in these notes either (but see [Tou08] if you want to know!). Thus one learns to be content with getting to know formulas and terms by their behaviour and through use, rather than by their exact definition of syntax.

- *Terms* are "function calls", in the jargon of the computer savvy person.

  These calls take math objects as *inputs* and return math objects as *outputs*.

  Examples are: $\underline{x,\ A,\ \emptyset,\ 0,\ 5,\ 42}$, $x + y$, $x \times 3$, $0 \times x + 1$, $A \cap B$.

  > The underlined examples above are the simplest possible objects (terms): Constants and variables.
  >
  > More complex ones are build via function calls.

NOTE. One is told that × is stronger than +, so, notwithstanding the bracket-parsimonious notation "$0 \times x + 1$", we know it means "$(0 \times x) + 1$", so this call returns 1, no matter what we plugged into x.

- *Formulas* are *also* function calls, but their output is *restricted* (by their syntax that I will not define carefully!) to be one or the other of the truth values true or false ($\mathbf{t}$ or $\mathbf{f}$) but nothing else! Their input, just as in the case for terms, is any math object.

  Examples are:

  $2 < 3$ ($\mathbf{t}$),

  $(\forall x)x = x$ ($\mathbf{t}$),

  $(\forall x)x = 0$ ($\mathbf{f}$),

  $(\exists x)x = 0$ ($\mathbf{t}$),

  $x = 0$ neither true nor false; answer depends on the input we place in x!

  More: $x = x$ ($\mathbf{t}$) answer is independent of input; $x = 0 \rightarrow x = 0$ ($\mathbf{t}$) answer is independent of input;

  $x = 0 \rightarrow (\forall x)x = 0$ neither true nor false; answer depends on the input in x!

  The input variable is the *leftmost x*; the other two (x's)are *bound* and *unavailable* to accept inputs. See below.

- If an **occurrence** of formula variable *is* available for input it would normally be called "an occurrence as an *input variable*".

Rather, such occurrences are called *free occurrences* in the literature.

Non-input occurrences of a variable are called "bound".

Let's *emphasise*: It is not a *variable* x that is free or bound in a formula, but it is *the occurrences of said variable* that we are speaking of, as the immediately preceding example makes clear.

4. In $(\forall x)x = 0$ the variable x is non input, it is "*bound*" we say.

   Just like this: $\Sigma_{i=1}^4 i$, which means $1 + 2 + 3 + 4$ and "i" is an illusion! *Not* available for input:

   Something like $\Sigma_{3=1}^4 3$ is nonsense!

   Similar comment for $(\forall x)x = 42$. Neither of these two occurrences of x is free (available) for substitution in it.

   No wonder "bound" variables are sometimes called "apparent variables".

5. We call $\forall, \exists, \neg, \wedge, \vee, \rightarrow, \equiv$ the "*logical connectives*", the last 5 of them being called *Boolean connectives*.

   People avoid cluttering notation with too many brackets by agreeing that the *first 3 connectives* have the same "strength" or "priority"; the *highest*. The remaining connectives have priorities decreasing as we walk to the right.

   Thus, if A and B are (*denote*) formulas, then $\neg A \vee B$ means $(\neg A) \vee B$; $\neg$ wins the claim for A. If we want $(\forall x)$ to apply to the entire $A \rightarrow B$ we must write $(\forall x)(A \rightarrow B)$.

   What about $A \rightarrow B \rightarrow C$ and $A \equiv B \equiv C$? Brackets are *implied from right to left*: $A \rightarrow (B \rightarrow C)$ and $A \equiv (B \equiv C)$.

   And this? $(\exists y)(\forall x)\neg A$. Brackets are *implied, again, from right to left*: $(\exists y)\big((\forall x)(\neg A)\big)$.

   BTW, the part of a formula where a $(\forall x)$ or $(\exists x)$ *acts upon* —the "$(\ldots)$" in $(\forall x)(\ldots)$ and $(\exists x)(\ldots)$— is called their *scope*.

March 18, 2022

6. **Boolean deconstruction**. A formula like $(\forall x)A \to B$ can be *deconstructed* Boolean-wise into $(\forall x)A$ and B. If I knew more about B —say, e.g., it is $x = 3 \to x = 7$, then I can deconstruct further.

   So, now I have got the full deconstruction:

   $$(\forall x)A, \quad x = 3, \quad x = 7$$

   The last two have NO Boolean structure so deconstructing stops with them. How about $(\forall x)A$? This cannot be deconstructed either, even if A had Boolean structure!

☡ Such structure is locked up in the scope of $(\forall x)$. ☡

We call the formulas where deconstruction stops "*prime*".

A prime formula is one with no explicit Boolean structure, e.g., $x < 8$, or one of the form $(\forall x)A$ (A is the scope) or $(\exists x)A$ (A is the scope).

Every formula is either prime or can be deconstructed into prime components.

Here the Boolean (**block-**)structure of the original above is

$$\boxed{(\forall x)A} \to \boxed{x = 3} \to \boxed{x = 7}$$

Playing the "WHAT-IF game" of Boolean logic —that is, being ONLY ABLE to say that the "boxes" have as value *one of* **t** or **f** each, but I do not know which— I conclude (from truth tables) that the Boolean structure does *NOT* make the above true for all choices of "what if" values.

6.1.1 Remark. (Tautologies) A formula A is a *tautology* iff it **is true due to its Boolean structure**, according to truth tables (2.3.4) *no matter what are the "what-if" values of its prime formulas into which it is deconstructed* are assumed to be.

Note "Assumed to be":

We do **NOT** compute the *intrinsic* truth value of a prime formula that is part of A **when we check whether A is a tautology or not**.

▶ For example, $x = x$ is a prime formula and thus its assumed value could be ANY ONE of **t** or **f**.

Thus it is NOT a tautology, even though, it *intrinsically IS true*, no matter what the value of x! □

The term "tautology" is applied only to formulas that do have Boolean structure.

6.1.2 Example.

1. $(\forall x)A$ is *not* a tautology as it has two possible truth values (being a prime formula).
2. $x = 0 \rightarrow x = 0$ *is* a tautology. Which are its prime (sub) formulas?

□

3. $(\forall x)x = 0 \rightarrow x = 0$ is *not* a tautology. I repeat (**one last time**):

   To determine tautologyhood we *DO NOT evaluate prime formulas*; we just consider *each* of the two scenarios, **t** or **f**, for each prime formula and use truth tables to compute the overall truth value.

   If we DID evaluate $(\forall x)x = 0$ we would see that (say over the natural numbers, or reals, or complex numbers) it is false.[†] So the implication is true! <u>But we DON'T do that</u>!

   This one is **Not true** <u>as a Boolean formula</u>!

[†] If we are doing our mathematics restricted to the set $\{0\}$, then, in this "theory" the formula IS true!

☡ So, how do we show that $(\forall x)A$ is true (if it is)?

Well, in easy cases we try to see if A is true for all values of x — which boils down to

> "fix an (*arbitrary*, hence undisclosed) x" and show A is true for THAT x.

▶ That failing, we will use a proof (see Section 6.2).

Similarly for $(\exists x)A$. To show it is true (if it is) we try to see if A is true for some value of x.

Often we just *guess* one such value that works, say c, and verify the truth of A when $x = c$.

▶ That failing, we will use a proof. ☡

6.1.3 Definition. (Important! Tautological implication)

We say that the formulas $A_1, A_2, \ldots, A_n$ *tautologically imply* a formula B —in symbols $A_1, A_2, \ldots, A_n \models_{taut} B$— meaning

"the truth of $A_1 \wedge A_2 \wedge \ldots \wedge A_n$ implies the truth of B"

that is, by the truth table for →, that

$$A_1 \wedge A_2 \wedge \ldots \wedge A_n \to B \text{ is a tautology}$$

□

So, $\models_{taut}$ propagates truth from left to right.

6.1.4 Example. Here are some easy and some involved tautological implications. They can all be verified using truth tables, either building the tables in full, or taking shortcuts.

1. $A \models_{taut} A$
2. $A \models_{taut} A \vee B$
3. $A \models_{taut} B \to A$
4. $A, \neg A \models_{taut} B$ —any B. Because I do "*work*" only if $A \wedge \neg A$ is true! See 6.1.3.
5. $\mathbf{f} \models_{taut} B$ —any B. Because I do *work* only if lhs is true! See above.
6. Is this a valid tautological implication? $\underline{B, A \to B \models_{taut} A}$, where A and B are distinct.

   No, for if A is false and B is true, then the lhs is true, but the rhs is false!
7. Is this a valid tautological implication? $\underline{A, A \to B \models_{taut} B}$? Yes! Say $A = \mathbf{t}$ and $(A \to B) = \mathbf{t}$. Then, from the truth table of →, it *must* be $B = \mathbf{t}$.
8. How about this? $\underline{A, A \equiv B \models_{taut} B}$? Yes! Verify!
9. How about this? $\underline{A \vee B \equiv B \models_{taut} A \to B}$? Yes! I verify:

   First off, **assume** lhs of $\models_{taut}$ —that is, $A \vee B \equiv B$— is true.

   Two cases:

   - $B = \mathbf{f}$. Then I need the lhs of $\equiv$ to be true to satisfy the red "assume". So $A = \mathbf{f}$ as well and clearly the rhs of $\models_{taut}$ is true with these values.
   - $B = \mathbf{t}$. Then I need not worry about A on the lhs. The rhs of $\models_{taut}$ is true by truth table of →.

10. $A \wedge (\mathbf{f} \equiv A) \models_{taut} B$, for any B. Well, just note that the lhs of $\models_{taut}$ is $\mathbf{f}$ so we need to do no work with B to conclude that the implication is valid.

11. $$A \to B, C \to B \models_{taut} A \vee C \to B$$

    This is nicknamed "proof by cases" for the obvious reasons. Verify this tautological implication! □

6.2. Proofs and Theorems

The job of a mathematical proof is to start from assumed (axioms) truths and unfailingly preserve truth in all its steps as it is developed.

Thus, when the proof ends it will have produced a truth at its very last step. A theorem.

A proof is a finite sequence of formulas —it is our "mathematical argument"— where *each formula we write down*, one per line with a short explanation, is either

1. an "assumption/hypothesis[†]" or an *axiom*,

   OR

2. is obtained from formulas we wrote earlier *IN THIS PROOF* employing *some valid rule*.

[†] To be explained on p.261.

Am I allowed in step 1. above to write *an already proved theorem* A?

Yes, because doing so is equivalent to lengthening the proof by adding —instead of just A— $\boxed{\ldots, A}$, that is, the *entire proof of* A obtained from axioms only, *not invoking other theorems*.

Programming analogy: I am allowed to invoke macros in a program because this is equivalent to writing down explicitly the macro-expansion code.

What are our axioms, our starting assumptions, when we do proofs?

We have two types:

1. Axioms needed by Logic (*Logical Axioms*) that are common in all proof-work that we do in *mathematics* or *computer science.*

   ▶ For example, such is the "identity" axiom $x = x$ and the tautology $\neg A \vee A$.

   Both these configurations —"$x = x$" and "$\neg A \vee A$"— define *infinitely many axioms* as their "instances".

   The first allows us to use *ANY* object variable in place of "x" the second allows to use any "statement" (*formula*) in place of A.

2. Axioms needed to do MATH in some theory (*Mathematical axioms*).

   Here is a *sample* of axioms from a few *MATH* (*theories*):

   (i) i. Number theory for $\mathbb{N}$:
   - $x < y \vee x = y \vee x > y$ (*trichotomy*)
   - $\neg x < 0$ this axiom indicates that 0 is *minimal* in $\mathbb{N}$.

     Adding the previous one makes $<$ a total order, so 0 is also *minimum*.
   - Many others that we omit.

   ii. Euclidean Geometry:
   - From two distinct points passes *one and only one* line.
   - ("Axiom of parallels") From a point A off a line named k —both A and k being on the same plane— passes a unique line on said plane that is parallel to k.
   - Many others that we omit.

iii. Axiomatic Set Theory:

- For any set A, we have

  $$(\exists y)y \in A \to (\exists x)\Big(x \in A \wedge \neg(\exists z \in A)z \in x\Big)$$

  This is the so-called axiom of "foundation" from which one can prove things like $A \in A$ is always *false*.

  This axiom incarnates Principles 0-2 in an axiomatic set theory like "ZFC".

  It says that *IF* $A \neq \emptyset$ —this is "$(\exists y)y \in A$"— *THEN* there is some element in A —this is the part "$(\exists x)\Big(x \in A$"— *which contains no element of* A —this is the part "$\neg(\exists z \in A)z \in x$".

- And a few others —including the Axiom of Choice, acronym "AC"— that we omit. □

☡ Foundation above tells us, among other things, that we cannot contain all members of a chain

$$\ldots \in x'' \in x' \in x$$

in a set A. ☡

And then we have "*hypotheses*" or "*assumptions*".

Are those not just axioms of logic or math? Not necessarily!

You recall that to prove $A \subseteq B$ you go like this:

"Let $x \in A$ for some fixed x". This "$x \in A$" is a hypothesis from which you will prove (hopefully) $x \in B$.

It is *NOT* an axiom of logic nor one of mathematics!

March 21, 2022

6.2.1 Definition. (The Logical Axioms)

1. All tautologies; these need no defence as "start-up truths".
2. Formulas of the form $(\forall x)A[x] \to A[t]$, for any formula A, variable x and "object" t.

   This object can be as simple as an (object) variable y (might be same as x), constant c, or as complex as a "*function call*", $f\Big(g\big(y, h(z)\big), a, b, w\Big)$ where f accepts 4 inputs, g accepts 2 and h accepts one. y, z, w are variables while a and b are unspecified constants.

   The axiom is true in any theory as it "says" "if A is true for all (values of) x, then it is also true for the specific value t".

   The axiom works only if we take care that the free variables of t (if any) that we substitute into the x of $A[x]$ do not accidentally get caught ("captured") into the scope of a quantifier $(\forall z)$ or $(\exists z)$ lurking inside A.

   So, If $A[y]$ is $(\exists x)x = y$ we cannot take t to be $f(x)$ and do $A[f(x)]$. Taking t to be $f(z)$ is OK: $(\exists x)x = f(z)$.

BTW "$[x]$" indicates the free variable of interest to us. It does not imply that x actually occurs free in A nor does it imply that there may not be *other* free variables in A.

How do I indicate that x, y, z are precisely all the free variables ("inputs") of A? $A(x, y, z)$.

3. Formulas of the form $A[x] \to (\forall x)A[x]$, **for any formula A where the variable x does not occur free in it.**

   That is. the truth value of A is independent of the value of x and writing —or not writing— "$(\forall x)$" up in front makes no difference.

   For example say A is $3 = 3$. This axiom says then, "if $3 = 3$ is true, then so is $(\forall x)3 = 3$".

   Sure! $3 = 3$ does not depend on x. So saying "for all values of x we have $3 = 3$" is the same as saying just "we have $3 = 3$".

4. $x = x$ is the *identity* axiom, no matter what “x” I use to express it. So, $y = y$ and $w = w$ are also instances of the axiom.

5. $x = y \to y = x$ and $x = y \wedge y = z \to x = z$ are the *equality* axioms. They can be expressed equally well using variables other than x and y (e.g., u, v and w).

□

6.2.2 Remark. (The “$\exists$”) The symbol $\exists$ is an abbreviation:

For any formula A, $(\exists x)A[x]$ stands for or is short for $\neg(\forall x)\neg A[x]$.

We also get the tautology (hence theorem)

$$\vdash \overbrace{(\exists x)A}^{\text{using abbrev. of rhs}} \equiv \neg(\forall x)\neg A$$

□

The "rules of proving", or rules of inference. These are two up in front —you will find I am grossly miscounting:

6.2.3 Definition. (Rules of Inference)

The rules used in proofs are called *rules of inference* and are these two (actually the second contains infinitely many rules).

1. From $A[x]$ I may infer $(\forall x)A[x]$. Logicians write the up-in-front ("primary") rules as fractions without words:

   $$\frac{A[x]}{(\forall x)A[x]} \tag{1}$$

   this rule we call *generalisation*, or *Gen* in short.

2. I may *construct* (and use) using any tautological implication *that I have verified*, say, this one

   $$A_1, A_2, \ldots, A_n \models_{taut} B \tag{2}$$

   leads to the rule

   $$\frac{A_1, A_2, \ldots, A_n}{B}$$

   Seeing readily that $A, A \to B \models_{taut} B$, we have the rule

   $$\frac{A, A \to B}{B}$$

   This is a very popular rule, known as *modus ponens*, for short *MP*.

Worth Saying. So rules preserve truth.

Read a rule such as (1) or (2) as saying

> If you *already* wrote *all* the formulas of the “numerator” (*in any order*) in a proof, then it is *legitimate to write thereafter in the proof* the denominator formula (of the rule).
>
> We call the numerator *input* or *hypotheses* of the rule and call the denominator *result* or *conclusion*.

□

6.2.4 Remark.

1. The second “rule” above is a rule constructor.

   Any tautological implication we come up with is fair game:

   It leads to a *valid rule* since the name of the game (in a proof) is *preservation/propagation of truth.*

   This is NOT an invitation to learn and memorise infinitely many rules (!) but is rather a license to build your own rules as you go, *as long as you bothered to verify* the validity of the tautological implication they are derived from.

2. Gen is a rule that indeed propagates truth: If $A[x]$ is true, that *means* that it is so for all values of x —and all values of any other free variables on which A depends but I did not show in the $[\ldots]$ notation.

   But then so is $(\forall x)A[x]$ true, as it *says precisely the same thing*: “$A[x]$ is true, for all values of x and all values of any other free variables on which A depends but I did not show in the $[\ldots]$ notation”.

   The only difference between the two notations is that I added some notational *emphasis* in the second —$(\forall x)$.

3. Hmm. So is $\forall x$ redundant? Yes, but *only as a formula PREFIX*. In something like this

$$x = 0 \to (\forall x)x = 0 \tag{1}$$

it is **NOT** redundant!

Dropping $\forall$ we change the meaning of (1).

As is, (1) is *not* a true statement. For example, if the value of the "input x" (the left one) is 0, then it is false.

However dropping $\forall x$, (1) changes to $x = 0 \to x = 0$ which is a tautology; *always true*.

□ ☡

We saw the *shape of proofs* at the outset, on p.256.

6.2.5 Definition. (Theorems)

A theorem is a formula that **appears** in a proof.

Often one writes $\vdash A$ to symbolically say that A is a theorem. If we must indicate that we worked in some specific theory, say ZFC (set theory), then we may indicate this as

$$\vdash_{ZFC} A$$

If moreover we have had some "*non-axiom* hypotheses" (see box on p.261) that form a set Σ, then we may indicate so by writing

$$\Sigma \vdash_{ZFC} A$$

□

Why Σ —and not A, B, C?— for a set of (*non-axiom*) assumptions? Because we reserve upper case latin letters for formulas. For *sets* of formulas we use a *distinguishable* capital letter, so, we chose distinguishable Greek capital letters, such as $\Gamma, \Sigma, \Delta, \Phi, \Theta, \Psi, \Omega$. Obviously, Greek capital letters like A, B, E, Z will not do!

6.2.6 Remark. (Hilbert-style proofs) The proof concept as defined is known as a “Hilbert-style proof”.

We write them *vertically*, ONE formula per line, every formula consecutively numbered, with annotation to the right of each formula written (this is the “why did I write this?”).

Like this

$$\begin{array}{lll} 1) & F_1 & \langle \text{because} \rangle \\ 2) & F_2 & \langle \text{because} \rangle \\ \vdots & \vdots & \vdots \\ n) & F_n & \langle \text{because} \rangle \end{array}$$

□

6.2.7 Example. (New (derived) rules) A derived rule is one we were not given —in 6.2.3— to bootstrap logic, but we can still prove they propagate truth.

1. We have a new (derived) rule: $(\forall x)A[x] \vdash A[t]$.

   This is called *Specialisation*, or *Spec*.

   Aha! We used a *non-axiom assumption* here!

   I write a Hilbert proof to show that $A[t]$ is a theorem if $(\forall x)A[x]$ is a (non-axiom) hypothesis (assumption) —shortened to "hyp".

   1) $(\forall x)A[x]$ $\langle \text{hyp} \rangle$
   2) $(\forall x)A[x] \to A[t]$ $\langle \text{axiom} \rangle$
   3) $A[t]$ $\langle 1 + 2 + \text{MP} \rangle$

2. Taking t to be x we have $(\forall x)A[x] \vdash A[x]$, simply written as $(\forall x)A \vdash A$.

3. The *Dual Spec* derived rule:

$$A[t] \vdash (\exists x)A[x] \tag{1}$$

We prove it below, but first I must prove:

$$\vdash A[t] \to (\exists x)A[x] \tag{2}$$

Here it goes

1)	$(\forall x)\neg A[x] \to \neg A[t]$	$\langle$axiom$\rangle$
2)	$A[t] \to \neg(\forall x)\neg A[x]$	$\langle$1 + Taut. Impl.$\rangle$
2′)	$A[t] \to (\exists x)A[x]$	$\langle$2 + using abbreviation “$\exists$”$\rangle$

Now, Dual Spec:

1)	$A[t]$	$\langle$hyp$\rangle$
2)	$A[t] \to (\exists x)A[x]$	$\langle$proved above$\rangle$
3)	$(\exists x)A[x]$	$\langle$1 + 2 + MP$\rangle$

Taking t to be x we have $A[x] \vdash (\exists x)A[x]$, simply written as $A \vdash (\exists x)A$.

□

March 23, 2022

There are two principles of proof that we state without proving them (see [Tou03a, Tou08] if curious).

6.2.8 Remark. (Deduction theorem and proof by contradiction)

1. The *deduction theorem* (also known as "proof by assuming the antecedent") states, if

$$\Gamma, A \vdash B \tag{1}$$

then also $\Gamma \vdash A \to B$, provided that in the proof of (1), all free variables that appear in A were treated as constants (as we say, were "frozen") AT or BELOW the point where A was inserted as a hypothesis:

> This "freezing" applies to ALL formulas in the proof that contain the free variables of A —*NOT only to* A.
>
> We cannot apply $\forall$ to any such variable in any formula at or below the hypothesis A, nor can we substitute values in them.

6.2.9 Example. To show $A \subseteq B$ do $(\forall x)(x \in A \to x \in B)$ or —same thing— $x \in A \to x \in B$ for all x.

To do the latter we pick a fixed ("frozen"!) undisclosed x and assume $x \in A$ to ensure the "$\to$" goes through for all (chosen) x.

Then we proceed to show $x \in B$ for that same x.

Hey! This is an application of the DThm!

You see above, intuitively, why free variables like x must be frozen —"fixed and undisclosed" we said! □

The notation "Γ, A" is standard for the more elaborate $\Gamma \cup \{A\}$.

In practice, this principle is applied to prove $\Gamma \vdash A \rightarrow B$, by doing instead the "easier" (1).

Why "easier"?

(1) We are helped by an *extra hypothesis*, A, and

(2) the formula to prove, B, is *less complex* than $A \rightarrow B$.

2. **Proof by contradiction**. To prove $\Gamma \vdash A$ —where A has *no free variables* or, as we say, is *closed* or is a *sentence*— is equivalent to proving the "constant formula" $\mathbf{f}$ from hypothesis $\Gamma, \neg A$.

3. Here is another reason to work with non-axiom hypotheses: The Deduction Theorem, which we apply as follows:

   Suppose we want to prove "$\vdash A \to B \to C \to D$".

   So we go like this:

   - By DThm, it suffices to prove $A \vdash B \to C \to D$ instead (here "Γ" is $\{A\}$).
   - Again, by DThm, it suffices to prove $A, B \vdash C \to D$ instead (here "Γ" is $\{A, B\}$).
   - Again, by DThm, it suffices to prove $A, B, C \vdash D$ instead (here "Γ" is $\{A, B, C\}$).

□ ☡

☡☡ I referred you to [Tou08] for some things.

However, the short intro here adopted the so-called "strong generalisation" (as did [Tou03a] and many other such as [Men87, Sho67]) —because it is closest to intuition of the practicing mathematician or computer scientist: "Stating $A(x)$ means that $A(x)$ is true for all the values of x", which has the side-effect of obliging the deduction theorem to be restricted:

In proving B from Γ, A one *must* ensure that no variable of A was subject to *generalisation* or *substitution*.

[Tou08] trades some power of generalisation in order to get an easier to apply deduction theorem, *without restrictions*.

So this is a choice on what we want to be "easy", and what we want to "be not so easy". There are two options! ☡☡

6.2.10 Remark. (Ping-Pong) For any formulas A and B, the formula —where I am using way more brackets than I have to, ironically, to *improve* readability—

$$(A \equiv B) \equiv \Big((A \to B) \land (B \to A)\Big)$$

is a tautology (we say this when discussing "→").

Thus to prove the lhs of the $\equiv$ suffices to prove the rhs.

In turn, to prove the rhs it *suffices* to prove each of $A \to B$ and $B \to A$ *separately*. This last idea encapsulates the ping-pong approach to proving equivalences. □

Here are a few applications.

6.2.11 Example. 1. Establish $\vdash (\forall x)(A \wedge B) \equiv (\forall x)A \wedge (\forall x)B$.

By ping-pong.

- Prove $\vdash (\forall x)(A \wedge B) \rightarrow (\forall x)A \wedge (\forall x)B$. By DThm suffices to do $(\forall x)(A \wedge B) \vdash (\forall x)A \wedge (\forall x)B$ instead.

1)	$(\forall x)(A \wedge B)$	⟨hyp⟩
2)	$A \wedge B$	⟨1 + Spec⟩
3)	A	⟨2 + tautological implication⟩
4)	B	⟨2 + tautological implication⟩
5)	$(\forall x)A$	⟨3 + Gen; OK: x is not free in line 1⟩
6)	$(\forall x)B$	⟨4 + Gen; OK: x is not free in line 1⟩
7)	$(\forall x)A \wedge (\forall x)B$	⟨5 + 6 + tautological implication⟩

Why the note "OK: x is not free in line 1"?

Because I applied DThm and moved $(\forall x)(A \wedge B)$ to the left of "$\vdash$"(made it hyp).

DThm *requires all* free variables of this to be *frozen* from the point of insertion down.

In particular I am *NOT allowed* to invoke $(\forall x)$ for such a frozen free variable. It is essentially a constant!

So the annotation observes that x is not free in what we moved over by DThm; Good!

- Prove $\vdash (\forall x)A \wedge (\forall x)B \rightarrow (\forall x)(A \wedge B)$. By DThm suffices to do $(\forall x)A \wedge (\forall x)B \vdash (\forall x)(A \wedge B)$ instead.

  1) $(\forall x)A \wedge (\forall x)B$ $\langle$hyp$\rangle$
  2) $(\forall x)A$ $\langle$1 + tautological implication$\rangle$
  3) $(\forall x)B$ $\langle$1 + tautological implication$\rangle$

  Complete the above proof!

2. Prove $\vdash (\forall x)(\forall y)A \equiv (\forall y)(\forall x)A$.

   By ping-pong.

   (a) Prove $\vdash (\forall x)(\forall y)A \to (\forall y)(\forall x)A$.
   By DThm suffices to do $(\forall x)(\forall y)A \vdash (\forall y)(\forall x)A$ instead.

1)	$(\forall x)(\forall y)A$	$\langle$hyp$\rangle$
2)	$(\forall y)A$	$\langle 1$ + Spec$\rangle$
3)	A	$\langle 2$ + Spec$\rangle$
4)	$(\forall x)A$	$\langle 3$ + Gen; OK, no free x in line 1$\rangle$
5)	$(\forall y)(\forall x)A$	$\langle 4$ + Gen; OK, no free y in line 1$\rangle$

   (b) Prove $\vdash (\forall y)(\forall x)A \to (\forall x)(\forall y)A$.
   Exercise! □

3. Prove $\vdash (\forall x)(A \to B) \to (\forall x)A \to (\forall x)B$.

   By the DThm do instead $(\forall x)(A \to B) \vdash (\forall x)A \to (\forall x)B$. In fact, even easier is to do (by DThm)

$$(\forall x)(A \to B), (\forall x)A \vdash (\forall x)B$$

   Here it is:

1)	$(\forall x)(A \to B)$	$\langle$hyp from DThm$\rangle$
2)	$(\forall x)A$	$\langle$hyp from DThm$\rangle$
3)	$A \to B$	$\langle 1 +$ Spec$\rangle$
4)	A	$\langle 2 +$ Spec$\rangle$
5)	B	$\langle 3 + 4 +$ MP$\rangle$
6)	$(\forall x)B$	$\langle 5 +$ Gen; OK, no free x in 1 or 2$\rangle$

6.2.12 Exercise. Prove for any A and B — where x is not free in A— that $\vdash (\forall x)(A \to B) \to (A \to (\forall x)B)$. □

6.2.13 Exercise. Prove for any A and B — where x is not free in A— that $A \to B \vdash A \to (\forall x)B$. □

☡ We have seen how to *add* an $(\exists x)$ in front of a formula (6.2.7 3).

How about *removing* an $(\exists x)$-prefix? This is much more complex than removing a $(\forall x)$-prefix:

The technique can be *proved* to be correct (eg., [Tou03a]) but I will omit the proof here as I did omit the proof of the deduction theorem technique and of the proof by contradiction technique.

I could say "see [Tou03a] if you want to learn the proof", but this reference is too advanced for a first year course on discrete math.

So, why not say "look at [Tou08]"?

Because, unfortunately, these two books have chosen *incompatible* "generalisation" rules, which results to *incompatible deduction theorem versions*.

The proof of the technique of eliminating $\exists$*-prefixes* relies on the deduction theorem. ☡

Technique of removing an $\exists$-prefix: Suppose I have that $(\exists x)A[x]$ is true —either as an **assumption** or a **theorem I proved earlier**— and I want to prove B.

Then I **assume** that —for *some* constant c that does not occur in B— $A[c]$ is true.

In words, "Let c be a value (constant!) that makes $A[c]$ true".

That is, I **add** $A[c]$ for an NEW constant c *NOT* in B as a *NEW* non-axiom hypothesis.

People annotate this step in a proof as "aux. hyp. related to $(\exists x)A[x]$."

Now I proceed to prove B using all that is known to me —that is, the axioms of the theory $\mathcal{T}$ and the non-axiom hypotheses Γ *and the non-axiom hypothesis* $A[c]$.

I do so by using all free (input-) variables of $A[c]$ as constants in my proof.[b]

[b] This is a side-effect of using the deduction theorem in the proof of correctness of the theorem below that justifies this technique.

6.2.14 Metatheorem. (Aux. Hyp. Metatheorem) *Suppose I work within theory $\mathcal{T}$ and hypotheses Γ and I have proved*

$$\Gamma \vdash_{\mathcal{T}} (\exists x)A[x]$$

Suppose next I add the hypothesis $A[c]$ to Γ, where c is a NEW *constant that is not part of B, and I manage to prove*

$$\Gamma, A[c] \vdash_{\mathcal{T}} B \tag{1}$$

where

(1) *All free variables of $A[c]$ were frozen throughout the proof*

(2) *c does not appear in the formulas of Γ or in the MATH axioms of $\mathcal{T}$.*

Then obtaining (1) under all stated restrictions constitutes a proof of $\Gamma \vdash_{\mathcal{T}} B$.

The “big deal” in 6.2.14 is that normally if you add a hypothesis X to hypotheses Γ and prove

$$\Gamma, X \vdash_{\mathcal{T}} B$$

then you cannot in general get rid of the dependence of the theorem B on the added hypothesis X.

Not so with the technique of Metatheorem 6.2.14: You get

$$\Gamma \vdash_{\mathcal{T}} B$$

as if you never assumed or used $A[c]$!

That is why they call it “auxiliary hypothesis”. Once it helps you to prove B it *drops out*; it does not stay around to get credit!

6.2.15 Example. Prove $\vdash (\exists y)(\forall x)A[x,y] \to (\forall x)(\exists y)A[x,y]$.

By the DThm it suffices to prove $(\exists y)(\forall x)A[x,y] \vdash (\forall x)(\exists y)A[x,y]$ instead.

1) $(\exists y)(\forall x)A[x,y]$ $\langle$hyp via DThm$\rangle$
2) $(\forall x)A[x,c]$ $\langle$aux. hyp. related to 1; for constant c not in the conclusion$\rangle$
3) $A[x,c]$ $\langle$2 + Spec$\rangle$
4) $(\exists y)A[x,y]$ $\langle$3 + Dual Spec$\rangle$
5) $(\forall x)(\exists y)A[x,y]$ $\langle$4 + Gen; OK, no free x in lines 1(DThm) and 2(aux. hyp)$\rangle$

Worth Noting: The "Γ" here is $\{(\exists y)(\forall x)A[x,y]\}$ thus we *do have* $\Gamma \vdash (\exists y)(\forall x)A[x,y]$ as required by 6.2.14.

What I am invoking here is the trivial $X \vdash X$ that is verified by the 1-line proof "1) X $\langle$*hyp*$\rangle$".

□

March 25, 2022

6.2.16 Example. Can I also prove the converse of the above? That is

$$\vdash (\forall x)(\exists y)A[x,y] \to (\exists y)(\forall x)A[x,y] \tag{1}$$

I will try.

By the DThm it suffices to prove $(\forall x)(\exists y)A[x,y] \vdash (\exists y)(\forall x)A[x,y]$ instead.

1)	$(\forall x)(\exists y)A[x,y]$	$\langle$hyp via DThm$\rangle$
2)	$(\exists y)A[x,y]$	$\langle 1 + \text{Spec}\rangle$
3)	$A[x,c]$	$\langle$aux. hyp. for 2; c not in the conclusion$\rangle$
4)	$(\forall x)A[x,c]$	$\langle 3 + \text{Gen}$; Stop! Forbidden! Illegal "$(\forall x)$": I should treat the free x of aux. hyp. as a constant on lines 3 and 4!$\rangle$

Still, can anyone PROVE (1); even if I cannot?

A question like this, *if you are to answer "NO"*, must be resolved by offering a **counterexample**.

That is, a special case of A for which I can clearly see that the claim is false.

Here is one such:

$$\underbrace{(\forall x)(\exists y)\overbrace{x=y}^{\text{“the } A\text{”}}}_{\mathbf{t}} \to \underbrace{(\exists y)(\forall x)\overbrace{x=y}^{\text{“the } A\text{”}}}_{\mathbf{f}} \tag{1}$$

□

Here is another non-theorem. We have the axiom $A \to (\forall x)A$ if x is not free in A. Can we relax the restriction?

No. If we had $\vdash A \to (\forall x)A$ with no restrictions then look at the special case

$$x = 0 \to (\forall x)x = 0 \tag{2}$$

on $\mathbb{N}$.

We already saw that this is not true for all x —not a theorem then!

In fact over $\mathbb{N}$, (2) is false if I take x to be 0: $\overbrace{0 = 0}^{\mathbf{t}} \to \overbrace{(\forall x)x = 0}^{\mathbf{f}}$.

6.2.17 Exercise. (Important "confusion remover") One might be confused by the act of *adding the hypothesis* $A(c)$ whenever we have $(\exists x)A(x)$.

Some lapse of judgement might construe this as an implication:

$$(\exists x)A(x) \to A(c) \tag{1}$$

This is false!! NOT a theorem!!

Working over the natural numbers, Prove by finding a very simple $A(x)$ and a specific appropriate constant c that (1) fails for this A and c so it is NOT a theorem! □

6.2.18 Exercise. (Important "confusion remover" #2) Prove by an *EASY* counterexample that $(\exists x)A[x] \to A[x]$ is not provable either.

□

Another useful principle that can be proved, but we will not do so, is that one can *replace equivalents-by-equivalents.* That is, if C is some formula, and if I have

1. $A \equiv B$, via proof, or via assumption, and also
2. A is a subformula of C

then I can *replace* one (or more) occurrence(s) of A in C (as subformula(s)) by B and call the resulting formula C'.

I will be guaranteed the conclusion $C \equiv C'$.

That is, from $A \equiv B$, I can prove $C \equiv C'$.

This principle is called the *equivalence theorem*.

Let's do a couple of ad hoc additional examples before we move to the section on Induction.

6.2.19 Example. $A \to B \vdash (\forall x)A \to (\forall x)B$.

By the DThm it suffices to prove $A \to B, (\forall x)A \vdash (\forall x)B$ instead.

1)	$A \to B$	$\langle$hyp$\rangle$
2)	$(\forall x)A$	$\langle$hyp from DThm$\rangle$
3)	A	$\langle 2$ + Spec$\rangle$
4)	B	$\langle 1 + 3$ + MP$\rangle$
5)	$(\forall x)B$	$\langle 4$ + Gen; OK as the DThm hyp. (line 2) has no free $x\rangle$

□

6.2.20 Example. (Substitution Theorem) We have $A[x] \vdash A[t]$ for any term t.

Indeed,

1)	$A[x]$	$\langle$hyp$\rangle$
2)	$(\forall x)A[x]$	$\langle 1$ + Gen$\rangle$
3)	$A[t]$	$\langle 2$ + Spec$\rangle$

□

6.2.21 Example. We have $A \to B \vdash (\exists x)A \to (\exists x)B$.

Proof via DThm, that is, prove

$$A \to B, (\exists x)A \vdash (\exists x)B$$

instead.

1)	$A[x] \to B[x]$	$\langle$hyp$\rangle$
2)	$(\exists x)A[x]$	$\langle$hyp via DThm$\rangle$
3)	$A[c]$	$\langle$aux. hyp. for 2$\rangle$
4)	$A[c] \to B[c]$	$\langle 1$ + 6.2.20; OK no free x in lines #2, 3$\rangle$
5)	$B[c]$	$\langle 3 + 4$ + MP$\rangle$
6)	$(\exists x)B[x]$	$\langle 5$ + Dual Spec$\rangle$

□

6.2.22 Example. READ ME! Refer to 6.2.2. Let us apply it to $\neg A$ for arbitrary A. We get

$$\vdash (\exists x)\neg A \equiv \neg(\forall x)\neg\neg A \tag{1}$$

Since $A \equiv \neg\neg A$ is a tautology, hence a theorem

Pause. Why "hence a theorem"?◀

we apply the *equivalence theorem* (p.292) and tautological implication[‡] and obtain:

$$\vdash \neg(\forall x)A \equiv (\exists x)\neg A \tag{2}$$

Applying another tautological implication to (2) we move the leftmost $\neg$ just past the "$\equiv$" and get

$$\vdash (\forall x)A \equiv \neg(\exists x)\neg A$$

which is of the same form as 6.2.2 with the roles of $\exists$ and $\forall$ reversed.

□

[‡] We have (1) and $\vdash \neg(\forall x)\neg\neg A \equiv \neg(\forall x)A$ as the lhs of the implication.

6.2.23 Example. $A \equiv B \vdash (\forall x)A \equiv (\forall x)B$.

True due to the equivalence theorem! "C" is "$(\forall x)A$". We replaced (one occurrence of) A by B in C, and we have assumed as starting point that $A \equiv B$. □

6.2.24 Exercise. Prove $A \equiv B \vdash (\forall x)A \equiv (\forall x)B$ without relying on the equivalence theorem. Rather use 6.2.19 in your proof, remembering the ping-pong tautology (6.2.10). □

6.3. Induction

In Remark 4.1.83 we concluded with a formulation —(2) on p.154— of the *minimal condition* (MC) for any order $<$ as (†) below:

Since any class $\mathbb{A}$ is given as $\mathbb{A} = \{x : F[x]\}$ for some $F[x]$ we have

The statement "*some order $<$ has MC*" is captured by the statement

For any "property", that is, formula $F[x]$, we have that the following is true

$$(\exists a)F[a] \rightarrow (\exists a)\Big(F[a] \wedge \neg(\exists y)\big(y < a \wedge F[y]\big)\Big) \qquad (\dagger)$$

This specific "$<$" is a total order (*satisfies trichotomy*) and thus the concepts *minimal* and *minimum* coincide as we know (4.1.81)

We will prove that MC (†) is equivalent to the *Principle* of so-called "*Strong Induction*" or "*Course-of-Values Induction*" —CVI— on $\mathbb{N}$.

In the proof I will use three easy THEOREMS (they are tautologies, hence axioms, hence theorems) as well as 6.2.2.

Theorem 1. $\vdash \neg A \vee B \equiv A \rightarrow B$

Theorem 2. $A \rightarrow B \equiv \neg B \rightarrow \neg A$ (contrapositive).

Theorem 3. $\neg(A \vee B) \equiv \neg A \wedge \neg B$ and also $\neg(A \wedge B) \equiv \neg A \vee \neg B$.

These are the well-known "de Morgan" equivalences.[§]. The first intuitively says "$A \vee B$ is false iff *both* A and B are false". The second intuitively says "$A \wedge B$ is false iff at least one of A and B is false".

Our proof below is written as a *conjunctional* $\Leftrightarrow$-chain, written vertically with annotation "$\langle \ldots \rangle$" to the right of *each* $\Leftrightarrow$.

Such proofs are called *Equational* ([DS90, GS94, Tou08]).

[§] Often called "de Morgan Laws".

March 28, 2022

So let $P[x]$ be an arbitrary "property" (formula!) of the variable x.

We start our "Equational proof" with (†) at the top of the chain, but where we replaced the arbitrary "F" there with "$\neg P$" here.

$$
\begin{aligned}
&(\exists a)\neg P[a] \rightarrow (\exists a)\Big(\neg P[a] \wedge \neg(\exists y)\big(y < a \wedge \neg P[y]\big)\Big)\\
\Leftrightarrow\ & \langle \text{using 6.2.2 and equiv. thm (p.292) removing double negations} \rangle\\
&\neg(\forall a)P[a] \rightarrow \neg(\forall a)\neg\Big(\neg P[a] \wedge (\forall y)\neg\big(y < a \wedge \neg P[y]\big)\Big)\\
\Leftrightarrow\ & \langle \text{contrapositive} \rangle\\
&(\forall a)\neg\Big(\neg P[a] \wedge (\forall y)\neg\big(y < a \wedge \neg P[y]\big)\Big) \rightarrow (\forall a)P[a]\\
\Leftrightarrow\ & \langle \text{two applications of de Morgan and equiv. thm} \rangle\\
&(\forall a)\Big(P[a] \vee \neg(\forall y)\big(\neg y < a \vee P[y]\big)\Big) \rightarrow (\forall a)P[a]\\
\Leftrightarrow\ & \langle \text{Theorem 1 above + equiv. thm (twice)} \rangle\\
&(\forall a)\Big(\quad \underbrace{(\forall y)\big(y < a \rightarrow P[y]\big) \rightarrow P[a]}_{\text{If, \underline{for all $y < a$, $P[y]$ is true}, then I can \underline{prove} $P[a]$}} \quad\Big) \rightarrow (\forall a)P[a] \qquad (‡)
\end{aligned}
$$

We have proved that (†) —the least principle for $\mathbb{N}$— of the previous page is equivalent to (‡).

(‡) embodies the *Strong Induction* or *Course-of-Values Induction* (CVI) Principle:

To prove $(\forall a)P[a]$, where P is a property over $\mathbb{N}$ and a is a $\mathbb{N}$-variable, it suffices to do TWO steps:

1. Fix —but do not disclose— an a (a is arbitrary but fixed).
2. On the (blue) hypothesis (underlined, line (‡); it is called *Induction Hypothesis* or *I.H.*) that $P[y]$ is true for all $y < a$, I must prove that so is $P[a]$.

   This last step is called the *Induction Step*, or *I.S.*

Hmm! All inductions have a "Basis", typically at 0. Doesn't this one? It does!

For $a = 0$, $(\forall y)(y < a \to P[y])$ is *true* since $y < a$ is *false*.

So the Basis is hiding in the I.S. as a "boundary case" (not included in the general "conditional" argument)

$$\textbf{If } \underbrace{(\forall y)\big(y < a \to P[y]\big)}_{I.H.} \textbf{ then } P[a]$$

since the "condition" (I.H.) does not help for $a = 0$.

The I.H. here is true absolutely (not a hypothetical "IF"), so truth tables oblige me to prove $P[0]$ absolutely, not conditionally; and do so from first principles —i.e., with no help from an I.H.

There is another simpler induction principle that we call, well, *simple* induction:

$$\frac{P[0], P[x] \to P[x+1]}{P[x]} \tag{SI}$$

"(SI)" for Simple Induction. That is, to prove $P[x]$ for all x (denominator) do *three* things:

Step 1. Prove/verify $P[0]$

Step 2. **Assume** $P[x]$ for fixed ("frozen") x (unspecified!).

Step 3. **prove** $P[x+1]$ for that same (previously frozen) x.

The assumption is the I.H. for simple induction.

The I.S. is the step that proves $P[x+1]$.

Note that what is described here is precisely an application of the Deduction theorem towards proving "$P[x] \to P[x+1]$", that is, **proving the implication for any given** x.

Step 4. If you have done **Step** 1. through **Step** 3. above, then you **have proved** $P[x]$ (for all x is implied!)

Is the principle (SI) *correct*? I.e., if I do all that the numerator of (SI) asks me to do (*equivalently*, **Steps** 1. – 3.), then do I *really* get that the denominator is true (for all x implied)? *YES!*

6.3.1 Theorem. *The validity of (SI) is a consequence of MC (least principle) on $\mathbb{N}$.*

Proof. Suppose (SI) is *not* correct.

Then, for some property $P[x]$, *despite having completed* **Steps** *1. – 3., $P[x]$ is* not true *for all x!*

Then,

$$\boxed{\text{let } n \in \mathbb{N} \text{ be } \textit{smallest} \text{ such that } P[n] \text{ is } \textit{false.}}$$

Now, $n > 0$ since I *did* verify the truth of $P[0]$ (**Step** 1.).

Thus, $n - 1 \geq 0$.

But then, when I proved "$P[x] \to P[x+1]$ for all x (in $\mathbb{N}$)" —in **Steps** 2. and 3.— this includes **proving**

$$P[n-1] \to P[n] \tag{4}$$

By the smallest-ness of n, $P[n-1]$ is *true*, hence $P[n]$ is true by (4).

I have just contradicted that $P[n]$ is false!

(SI) works if MC does! □

In fact, MC and SI are equivalent principles.

6.3.2 Theorem. *Conversely to the previous theorem (6.3.1), if* SI *on* $\mathbb{N}$ *works, then* $\mathbb{N}$ *has MC.*

Proof. By contradiction, I assume I have SI, but that *MC fails*.

So, there is a nonempty subset of $\mathbb{N}$, S, that has no least element.

I will get a contradiction by showing that $\overline{S} \overset{Def}{=} \mathbb{N} - S$ is all of $\mathbb{N}$ (hence $S = \emptyset$).

I apply *SI* to the property

$$P(x) \overset{Def}{\equiv} \{0, 1, \ldots, x\} \subseteq \overline{S}$$

1. Basis. $P(0)$ says $\{0\} \subseteq \overline{S}$ which is equivalent to $0 \in \overline{S}$; true since if $0 \in S$ that would contradict assumption on S.
2. Fix x and assume (I.H.) $P(x)$.
3. $P(x+1)$ says $\{0, 1, \ldots, x, x+1\} \subseteq \overline{S}$. To prove this note:

   By 2., we have $\{0, 1, \ldots, x\} \subseteq \overline{S}$ so if $x + 1 \in S$ instead, then it would be smallest, contradicting hypothesis about S.

   Thus I MUST have also $\{0, 1, \ldots, x, x + 1\} \subseteq \overline{S}$ —and hence $P(x + 1)$.

   By SI, I have $P(x)$ true for all x, thus $\{0, 1, \ldots, x\} \subseteq \overline{S}$ for all x. In particular, $x \in \overline{S}$ for all x

   But then $S = \emptyset$. A contradiction! □

Since we have CVI equivalent to MC we now have

6.3.3 Corollary. *All three of CVI, SI and MC are equivalent principles over $\mathbb{N}$.*

Within set theory we can define —actually construct— *each* natural number as a set: $\emptyset$ for "0", and if we have defined/constructed as far as "n", then "$n+1$" is constructed as $n \cup \{n\}$.

Thus the first few natural numbers "are"

$$\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \ldots$$

We then can prove that the set of all of them —in modern literature denoted by ω rather than $\mathbb{N}$— satisfies MC (as we argued intuitively sometime ago), and thus we have all three by the corollary above.

However the self-standing theory of numbers (without the help, or even the concepts) of set theory —known as *Peano arithmetic* in short PA— assumes SI as an axiom.

You can then prove via a modification of our proofs (which are situated within set theory) Corollary 6.3.3.

6.3.1. Induction Practise

To begin with, there are “properties” to prove that are valid for all $n \geq k$ for some constant $k > 0$.

This is the domain where we have to stay in during the proof.

Thus for those the I.H. MUST “pick a fixed unspecified $n \geq k$”.

The points $n = 0, 1, \ldots, k-1$ are outside the domain so are “illegal”.

Thus the Basis of the induction must be for $n = k$.

As an example, the smallest n where $n + 3 < 2^n$ is true is $n = 3$ (verify!).

We can prove by induction

$$n + 3 < 2^n, \text{ for } n \geq 3$$

verifying as Basis the case $n = 3$.

Another example:

“n has a prime factor” is *erratic* for $n < 2$. For $n = 1$ it is *false* and for $n = 0$ it is *true* (every number is a factor of zero).

So one must take as domain of truth of the quoted blue property the set $\{n \in \mathbb{N} : n \geq 2\}$. 2 is the Basis.

6.3.4 Example. This is the "classical first example of induction use" in the discrete math bibliography! Prove that

$$0+1+2+\ldots+n=\frac{n(n+1)}{2} \tag{1}$$

So, the property to prove is the statement (1).

One must learn to not have to rename the various "properties" that we encounter as "$P[n]$".

I will use SI. So let us do the *Basis*. Boundary case is $n=0$. We verify: $lhs=0$. $rhs=(0\times 1)/2=0$. Good!

Fix n and take the expression (1) as I.H.

Do the I.S. Prove:

$$0+1+2+\ldots+n+(n+1)=\frac{(n+1)(n+2)}{2}$$

Here it goes

$$\begin{aligned}0+1+2+\ldots+n+(n+1) &\overset{\text{using I.H.}}{=} \frac{n(n+1)}{2}+(n+1)\\ &=(n+1)(n/2+1)\\ &=\frac{(n+1)(n+2)}{2}\end{aligned}$$

□

I will write more concisely in the examples that follow.

6.3.5 Example. Same as above but doing away with the "0+". Again, I use SI.

$$1 + 2 + \ldots + n = \frac{n(n+1)}{2} \tag{1}$$

- *Basis.* $n = 1$: (1) becomes $1 = (1 \times 2)/2$. True.
- Take (1) as I.H. with fixed n.
- I.S.:

$$\begin{aligned} 1 + 2 + \ldots + n + (n+1) &\overset{\text{using I.H.}}{=} \frac{n(n+1)}{2} + (n+1) \\ &= (n+1)(n/2+1) \\ &= \frac{(n+1)(n+2)}{2} \end{aligned}$$

□

March 30, 2022

6.3.6 Example. Prove

$$1 + 2 + 2^2 + \ldots + 2^n = 2^{n+1} - 1 \tag{1}$$

By SI.

- Basis. $n = 0$. $1 = 2^0 = 2^1 - 1$. True.
- As I.H. take (1) for fixed n.
- I.S.

$$\begin{aligned} 1 + 2 + 2^2 + \ldots + 2^n + 2^{n+1} &\overset{\text{using I.H.}}{=} 2^{n+1} - 1 + 2^{n+1} \\ &= 2 \cdot 2^{n+1} - 1 \\ &= 2^{n+2} - 1 \end{aligned}$$

□

6.3.7 Example. (Euclid) Every natural number $n \geq 2$ has a prime factor.

p is prime iff

1. $p > 1$

   AND

2. The only divisors of p are 1 and p

I do CVI (as you will see why!)

- *Basis*: For $n = 2$ we are done since 2 is a prime and $2 = 2 \times 1$.[†]
- I.H. Fix an n and assume the claim for all k, such that $2 \leq k < n$.
- I.S.: Prove for n: Two subcases:
  1. If n is prime, then OK! n divides n.
  2. If not, then $n = a \times b$, where $a \geq 2$ **and** $b \geq 2$. By I.H. a has a prime factor, thus so does $n = a \cdot b$. □

You see? Do you know many natural numbers n such that $n - 1$ divides n?! Only then an I.H. on $n - 1$ would be useful.

Only 2 has as factor $2 - 1$, so $2 = 2 \times 1$ but this is our Basis and has been handled directly (not looking at I.H.)

[†] You will recall that a number $\mathbb{N} \ni n > 1$ is a *prime* iff its **only** factors are 1 and n.

6.3.8 Example. (Euclid) Every natural number $n \geq 0$ is expressible base-10 as an expression

$$n = a_m 10^m + a_{m-1}10^{m-1} + \cdots + a_1 10 + a_0 \tag{1}$$

$$\text{where each } a_i \text{ satisfies } 0 \leq a_i < 10 \tag{2}$$

Proof by CVI again. You will see why.

- *Basis.* For $n = 0$ the expression "0" has the form of the rhs of (1) *and* satisfies inequality (2).
- Fix an $n > 0$ and assume (I.H.) that if $k < n$, then k can be expressed as in (1).
- For the I.S. express the $n > 0$ of the I.H. using Euclid's theorem (4.1.49) as

  $$n = 10q + r$$

  where $0 \leq r < 10$. By the I.H. —since $q < n$— let

  $$q = b_t 10^t + b_{t-1}10^{t-1} + \cdots + b_1 10 + b_0$$

  with $0 \leq b_j < 10$.

  NOTE. If we wanted to use SI instead of CVI (in theory possible by Corollary 6.3.3) this particular analysis does NOT naturally lead to SI since in $n = 10q + r$ —with $0 \leq r < 10$— I cannot expect that $q = n - 1$ to benefit from a "SI-type I.H."

  Then

  $$\begin{aligned} n &= 10q + r \\ &= 10\Big(b_t 10^t + b_{t-1}10^{t-1} + \cdots + b_1 10 + b_0\Big) + r \\ &= b_t 10^{t+1} + b_t 10^t + \cdots + b_1 10^2 + b_0 10 + r \end{aligned}$$

We see n has the right form since $0 \leq r < 10$. □

6.3.9 Example. An inequality. Let p_n denote the n-th prime number, for $n \geq 0$. Thus $p_0 = 2$, $p_1 = 3$, $p_2 = 5$, etc.

We prove that

$$p_n \leq 2^{2^n} \tag{1}$$

I use CVI on n. This is a bit of a rabbit out of a hat if you never read Euclid's proof that there are infinitely many primes.

- Basis $p_0 = 2 \leq 2^{2^0} = 2^1 = 2$.
- Fix $n > 0$ and take (1) as I.H.
- The I.S.: I will work with the fixed n above and the expression (product of primes, plus 1; this is inspired from Euclid's proof quoted above).

  $$p_0 p_1 p_2 \cdots p_n + 1$$

  I have

  $$\begin{aligned} p_0 p_1 p_2 \cdots p_n + 1 &\leq 2^{2^0} 2^{2^1} 2^{2^2} \cdots 2^{2^n} + 1 && \text{by I.H.} \\ &= 2^{2^0+2^1+2^2+\cdots+2^n} + 1 && \text{algebra} \\ &= 2^{2^{n+1}-1} + 1 && \text{by 6.3.6} \\ &< 2^{2^{n+1}-1} + 2^{2^{n+1}-1} && \text{smallest } n \text{ possible is } n = 1 \\ &= 2^1 \cdot 2^{2^{n+1}-1} \\ &= 2^{2^{n+1}} \end{aligned}$$

Now we have two cases on $q = p_0 p_1 p_2 \cdots p_n + 1$

1. q is a prime. Because of the "$+1$" q is different from all p_i in the product, so q is p_{n+1} or p_{n+2} or p_{n+3} or $\ldots$
   Since the sequence of primes is strictly increasing, p_{n+1} *is the least that q can be*.
   Thus
   $$p_{n+1} \le p_0 p_1 p_2 \cdots p_n + 1 \le 2^{2^{n+1}}$$
   in this case.

2. q is composite. By 6.3.7 some prime r divides q. Now, none of the
   $$p_0, p_1, p_2, \cdots, p_n$$
   divides q because of the "$+1$".

   Thus r is different from all of them, so it must be one of p_{n+1} or p_{n+2} or p_{n+3} or $\ldots$
   Thus,
   $$p_{n+1} \le r < q = p_0 p_1 p_2 \cdots p_n + 1 \le 2^{2^{n+1}}$$

Done! □

6.3.10 Example. Let

$$\begin{aligned} b_1 &= 3,\ b_2 = 6 \\ b_k &= b_{k-1} + b_{k-2}, \text{ for } k \geq 3 \end{aligned}$$

Prove by induction that b_n $\underline{\text{is divisible by 3 for } n \geq 1.}$ (Be careful to distinguish between what is *basis* and what are *cases* arising from the **induction step**!)

Proof. So the boundary condition is (from the underlined part above) $n = 1$. This is the *Basis*.

1. *Basis*: For $n = 1$, I have $a_1 = 3$ and this is *divisible by 3*. We are good.
2. *I.H.* Fix n and **assume claim** for all $1 \leq k < n$.
3. *I.S.* **Prove claim** for the above fixed n. There are two cases, as the I.H. is *not useable* for $n = 2$.

   Why? Because it would require entries b_0 and b_1.

   The red entry does not exist since the sequence starts with b_1. So,

   Case 1. $n = 2$. DIRECTLY. I am OK as $b_2 = 6$; it *is* divisible by 3.

   Case 2. $n > 2$. Is b_n divisible by 3? Well, $b_n = b_{n-1} + b_{n-2}$ in this case. By I.H. (valid for all $1 \leq k < n$) I have that $b_{n-1} = 3t$ and $b_{n-2} = 3r$, for some integers t, r. Thus, $b_n = 3(t + r)$. Done! □

Here are a few additional exercises for you to try —please do try!

6.3.11 Exercise.

1. Prove that $2^{2n+1} + 3^{2n+1}$ is divisible by 5 for all $n \geq 0$.
2. Using induction prove that $1^3 + 2^3 + \ldots + n^3 = \left[\dfrac{n(n+1)}{2}\right]^2$, for $n \geq 1$.
3. Using induction prove that $\sum_{i=1}^{n+1} i2^i = n2^{n+2} + 2$, for $n \geq 0$.
4. Using induction prove that $\sqrt{n} < \dfrac{1}{\sqrt{1}} + \dfrac{1}{\sqrt{2}} + \ldots + \dfrac{1}{\sqrt{n}}$, for $n \geq 2$.
5. Let

$$\begin{aligned} b_0 &= 1,\ b_1 = 2,\ b_3 = 3 \\ b_k &= b_{k-1} + b_{k-2} + b_{k-3}, \text{ for } k \geq 3 \end{aligned}$$

   Prove by induction that $b_n \leq 3^n$ for $n \geq 0$. (Once again, be careful to distinguish between what is *basis* and what are *cases* arising from the **induction step**!) □

Chapter 7

Inductively defined sets; Structural induction

An example of an inductively defined set is the following.

Suppose you want to define *by finite means*, and define *precisely*, the set of all *"simple" arithmetical expressions* that use the numbers $1, 2, 3$, the operations $+$ and $\times$, and round brackets.

Then you would define:

The set of said *simple arithmetical expressions* is the *smallest* set ($\subseteq$-smallest) that

1. Contains each of $1, 2$ and 3.
2. If it contains expressions E and E', then it also contains $(E + E')$ and $(E \times E')$.

Some folks would add a 3rd requirement "nothing else is in the set unless so demonstrated using 1. 2. above" and they omit "smallest".

Really?!

How exactly would you so “demonstrate”?

In a recursive definition you ought to be able to make your *recursive calls* and not have to trace back why the object you constructed exists!

We will prove in Theorem 7.2.5 that indeed there *is* an iterative way to show that a *particular* simple arithmetic expression was formed correctly by our recursion, but that defeats the beauty of recursion.

Besides, until we reach said theorem we don't even know how to prove that "nothing else is in the set unless "so demonstrated" (!!) using 1. 2. above".

WHAT does such a "demonstration" look like?

So it is nonsense to add such a statement in the bottom of the definition as a (redundant) afterthought.

Before we get to the general definitions, let us finesse our construction and propose some terminology.

(a) First off, in step 1. above we say that $1, 2$ and 3 are *the initial objects* of our recursive/inductive definition.

(b) In step 2. we say that $(E+E')$ is obtained by an *operation* (on strings) that is available to us, depicted as a "blackbox" below, which we named "+".

$$\begin{tikzcd}
E \arrow[dr] & & \\
& \boxed{\quad + \quad} \arrow[r] & (E+E') \\
E' \arrow[ur] & &
\end{tikzcd}$$

In words, the operation *concatenates from left to right the strings*

"(", "E", "+", "E'", and ")"

Similar comments for the operation "×".

(c) Both operations in this example are single-valued, that is, *functions*.

It is preferable to be slightly more general and allow operations that are just relations but sets nevertheless, but not necessarily functions.

Such an operation $O(x_1, \ldots, x_n, y)$ is n-ary —n inputs, $x_1, \ldots, x_n$— with output variable y.

7.0.1 Definition. (Closed Under) We say that a class of objects S is *closed under* a **relation** (operation) —which could be a function— $O(x_1, \ldots, x_n, y)$ meaning that for *all* input values $x_1, \ldots, x_n$ in S, *all* the obtained values y are also in S. □

We are ready for the general definition:

7.0.2 Definition. Given a set of *initial objects* $\mathcal{I}$ and a *set* of *operations* —each one being a set— $\mathcal{O} = \{O_0, O_1, O_2, \ldots\}$.

The object $\text{Cl}(\mathcal{I}, \mathcal{O})$ is called a *closure of* $\mathcal{I}$ *under* $\mathcal{O}$ —or the class *inductively defined by the pair* $(\mathcal{I}, \mathcal{O})$— and *denotes* the $\subseteq$-smallest *class*[†] S that satisfies

1. $\mathcal{I} \subseteq S$.

2. S is *closed under* all *operations in* $\mathcal{O}$, or simply, closed under $\mathcal{O}$.

3. The "smallest" part means: Any class T that satisfies 1. and 2. also satisfies $S \subseteq T$.

The set $\mathcal{O}$ may be infinite but is countable: That is, the numbering $i \mapsto O_i$ is from $\mathbb{N}$ onto $\mathcal{O}$. □

[†] Let's say "class" until we learn that the closure is actually a *set*.

Nice definition, but does $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$ *exist* given $\mathcal{I}$ and $\mathcal{O}$? Yes. But first,

7.0.3 Theorem. *For any choice of $\mathcal{I}$ and $\mathcal{O}$,* if $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$ exists, *then it is unique.*

Proof. Say the definition of $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$ ambiguously —i.e., may have more than one value— leads to (or produces) two classes, S and T.

Then, letting S pose as closure, we get $S \subseteq T$ from 7.0.2.

Then, letting T pose as closure, we get $T \subseteq S$, again from 7.0.2. Thus $S = T$. □

7.0.4 Theorem. *For any choice of $\mathcal{I}$ and $\mathcal{O}$ with the restrictions of Definition 7.0.2 the class* $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$ *exists and is a set.*

Proof. We have to check and note a few things.

1. By 4.1.5, for each O_i, $\mathrm{ran}(O_i)$ is a set.
2. The class $\mathbb{F} = \{\mathrm{ran}(O_i) : i = 0, 1, 2 \ldots\}$ is a set. This is so by **Principle** 3, since I can index all members of $\mathbb{F}$ by assigning unique indices from $\mathbb{N}$ to each of its members (and $\mathbb{N}$ is a set by **Principle** 0).
3. By 2. above and 2.4.17, $\bigcup \mathbb{F}$ is a set, and so is $T = \mathcal{I} \cup \bigcup \mathbb{F}$.
4. T contains $\mathcal{I}$ as a subset and is $\mathcal{O}$-closed since any O_i-output —*no matter where the inputs come from*— is in $\mathrm{ran}(O_i) \subseteq \bigcup \mathbb{F}$.
5. The family of sets $\mathbb{G} = \{S : \mathcal{I} \subseteq S \wedge S \text{ is } \mathcal{O}\text{-closed}\}$ contains the set T as a member. Thus (cf. 2.4.18)

$$C \overset{Def}{=} \left(\bigcap \mathbb{G}\right) \subseteq T$$

is a set.

Since all sets S in $\mathbb{G}$ contain $\mathcal{I}$ and are $\mathcal{O}$-closed, so is C (*Exercise*!).

But $C \subseteq S$ for all such sets S the way it (C) is defined.

So it is $\subseteq$-smallest.

By Definition 7.0.2(3) we must be sure that C is also the smallest among all CLASSES $\mathbb{A}$ that contain $\mathcal{I}$ and are $\mathcal{O}$-closed:

So let $\mathbb{A}$ be such a class.

The SET $\mathbb{A} \cap T$ —where T is that in the proof above— contains $\mathcal{I}$ and is $\mathcal{O}$-closed since each of T and $\mathbb{A}$ do (*Exercise*!) hence is in $\mathbb{G}$.

Thus, by the last boxed statement above,

$$C \subseteq \mathbb{A} \cap T \subseteq \mathbb{A}$$

Thus C satisfies 1–3 in 7.0.2 and hence $C = \mathrm{Cl}(\mathcal{I}, \mathcal{O})$. □

7.1. Induction over a closure

7.1.1 Definition. Let a pair $(\mathcal{I}, \mathcal{O})$ be given as above.

We say that a property

$P[x]$ propagates with $\mathcal{O}$ *iff* for each $O_i(x_1, \ldots, x_n, y) \in \mathcal{O}$, *if whenever all the inputs in the x_i* satisfy $P[x]$ (i.e., $P[x_i]$ is true for each argument x_i), then all output values returned by y —for said inputs— satisfy $P[y]$ as well.

We can also say that the property $P[x]$ is preserved by $\mathcal{O}$.

Recall that for each assignment of values to the inputs $x_1, \ldots, x_n$ we *may have more than one output values in y*; for all such values $P[y]$ is true. □

7.1.2 Lemma. *For all $(\mathcal{I}, \mathcal{O})$ and a property $P[x]$, if the latter propagates with $\mathcal{O}$, then the class $\mathbb{A} = \{x : P[x]\}$ is closed under $\mathcal{O}$ (is $\mathcal{O}$-closed).*

Proof. So let $O_i(x_1, \ldots, x_n, y) \in \mathcal{O}$. Let $a_1, \ldots, a_n$ be all in $\mathbb{A}$. Thus

$$P[a_i], \text{ for all } i = 1, \ldots, n$$

By assumption, if $O_i(a_1, \ldots, a_n, b)$, then $P[b]$ is true, hence $b \in \mathbb{A}$. □

7.1.3 Theorem. *Let* $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$ *and a property* $P[x]$ *be given. Suppose we have done the following steps:*

1. *We showed that for each* $a \in \mathcal{I}$, $P[a]$ *is true.*
2. *We showed that* $P[x]$ *propagates with* $\mathcal{O}$.

Then every $a \in \mathrm{Cl}(\mathcal{I}, \mathcal{O})$ *has property* $P[x]$.

Naturally, the technique encapsulated by 1. and 2. of 7.1.3 is called "induction over $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$" or "structural induction" over $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$.

Note that for each $O_i \in \mathcal{O}$ the "propagation of property $P[x]$" will take the form of an I.H. (assume the inputs of O_i have the property) followed by an I.S. (then the outputs of O_i have the property):

Proof. (of 7.1.3) Let us write

$$\mathbb{A} \overset{Def}{=} \{x : P[x]\}$$

Thus, 1. in 7.1.3 translates to

$$\mathcal{I} \subseteq \mathbb{A} \tag{$*$}$$

2. in 7.1.3 yield

$$\mathbb{A} \text{ is } \mathcal{O}\text{-closed} \tag{$**$}$$

Hence, by 7.0.2 (3) and $(*)$ and $(**)$,

$$\mathrm{Cl}(\mathcal{I}, \mathcal{O}) \subseteq \mathbb{A}$$

The last inclusion immediately translates to

$$\text{“}\underline{x \in \mathrm{Cl}(\mathcal{I}, \mathcal{O}) \text{ implies } P[x] \text{ is true}}\text{”} \qquad \square$$

7.1.4 Example. Let $S = \mathrm{Cl}(\mathcal{I}, \mathcal{O})$ where $\mathcal{I} = \{0\}$ and $\mathcal{O}$ contains just one operation, $x + 1 = y$, where y is the output variable. That is,

$$n \longrightarrow \boxed{x + 1 = y} \longrightarrow n + 1 \tag{1}$$

is our only operation. By induction over S, I can show $S \subseteq \mathbb{N}$.

The "$P[x]$" is "$x \in \mathbb{N}$".

So $P[0]$ is true. I verified the property for $\mathcal{I}$. That the property propagates with our operation is captured by (1) above (if $n \in \mathbb{N}$, then $n + 1 \in \mathbb{N}$). Done!

Can we show also $\mathbb{N} \subseteq \mathrm{Cl}(\mathcal{I}, \mathcal{O})$? **Yes**:

In this direction I do SI over $\mathbb{N}$ on variable n.

The property, let's call it $Q[x]$, now is "$x \in \mathrm{Cl}(\mathcal{I}, \mathcal{O})$".

For $n = 0$, $n \in \mathrm{Cl}(\mathcal{I}, \mathcal{O})$ since $0 \in \mathcal{I} \subseteq \mathrm{Cl}(\mathcal{I}, \mathcal{O})$ by 7.0.2.

Now, say (*I.H.*) $n \in \mathrm{Cl}(\mathcal{I}, \mathcal{O})$. Since $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$ is closed under the operation $x + 1 = y$, we have $n + 1 \in \mathrm{Cl}(\mathcal{I}, \mathcal{O})$ by 7.0.2.

So,

$$\mathrm{Cl}(\mathcal{I}, \mathcal{O}) = \mathbb{N} \qquad \square$$

☡ Thus the induction over a closure generalises *SI*. ☡

7.2. Closure vs. definition by stages

We will see in this section that there is also a *by-stages* or *by-steps* way to obtain $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$.

7.2.1 Definition. (Derivations) An *$(\mathcal{I}, \mathcal{O})$-derivation* —or just *derivation* if we know which $(\mathcal{I}, \mathcal{O})$ we are talking about— is a *finite sequence of objects*

$$d_1, d_2, d_3, \ldots, d_i, \ldots, d_n \tag{1}$$

satisfying:

Each d_i is

1. A member of $\mathcal{I}$,

   or

2. For some j, one of the results of $O_j(x_1, \ldots, x_k, y)$ with inputs $a_1, \ldots, a_k$ that are found in the derivation (1) *to the left of* d_i.

n is called the *length of the derivation*. Every d_i is called an *$(\mathcal{I}, \mathcal{O})$-derived* object, or just *derived*, if the $(\mathcal{I}, \mathcal{O})$ is understood. □

☡ Clearly, the concept of a derivation abstracts, thus generalises, the concept of *proof*, while a derived object abstracts the concept of a *theorem*. Initial objects abstract the concept of *axiom*. ☡

7.2.2 Example. For the $(\mathcal{I}, \mathcal{O})$ of 7.1.4, here are some derivations:

$$0$$

$$0, 0, 0$$

$$0, 1, 0, 1, 0, 1, 1, 1, 1, 0$$

Nothing says we cannot repeat a d_i in a derivation! Lastly here is an "efficient" derivation with no redundant steps: $0, 1, 2, 3, 4, 5$. □

7.2.3 Proposition. *If* $d_1, d_2, d_3, \ldots, d_i, \ldots, d_n, d_{n+1}, \ldots, d_m$ *is a* $(\mathcal{I}, \mathcal{O})$*-derivation, then so is* $d_1, d_2, d_3, \ldots, d_i, \ldots, d_n$.

Proof. Each d_i is validated in a derivation either outright (i.e., is in $\mathcal{I}$) or *by looking to the left*!

What we may remove a "(red)tail" that is to the *right* of d_i, this does not affect the validity of that entry. □

7.2.4 Proposition. *If* $d_1, d_2, \ldots, d_n$ *and* $e_1, e_2, \ldots, e_m$ *are* $(\mathcal{I}, \mathcal{O})$*-derivations, then so is* $d_1, d_2, \ldots, d_n, e_1, e_2, \ldots, e_m$.

Proof. Traversing $d_1, d_2, \ldots, d_n$ and $e_1, e_2, \ldots, e_m$ in

$$d_1, d_2, \ldots, d_n, e_1, e_2, \ldots, e_m$$

from left to right we validate each d_i and each e_j giving precisely the same validation *reason* as we would in each sequence $d_1, d_2, \ldots, d_n$ and $e_1, e_2, \ldots, e_m$ as *standalone*.

These reasons are local to each sequence. □

April 4, 2022

An object x is called $(\mathcal{I}, \mathcal{O})$-derived iff it appears in some $(\mathcal{I}, \mathcal{O})$-derivation.

By the "tail Lemma" (7.2.3) we can always assume that a derived object appears at the end of an $(\mathcal{I}, \mathcal{O})$-derivation *as we can remove the tail that follows it*.

We next prove that defining a set S as a $(\mathcal{I}, \mathcal{O})$-closure is equivalent to defining S as the set of all $(\mathcal{I}, \mathcal{O})$-derived objects.

7.2.5 Theorem. *For any initial sets of objects and operations on objects ($\mathcal{I}$ and $\mathcal{O}$) we have that* $\mathrm{Cl}(\mathcal{I}, \mathcal{O}) = \{x : x$ *is* $(\mathcal{I}, \mathcal{O})$*-derived*$\}$.

Proof. Let us write $D = \{x : x \text{ is } (\mathcal{I}, \mathcal{O})\text{-derived}\}$ and prove that $\mathrm{Cl}(\mathcal{I}, \mathcal{O}) = D$. We have two directions:

1. $\mathrm{Cl}(\mathcal{I}, \mathcal{O}) \subseteq D$: By induction over $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$. The property to prove is "$x \in D$".

   - Let $x \in \mathcal{I}$. Then x is derived via the one-member derivation

     $$x$$

     So $x \in D$. Thus all $x \in \mathcal{I}$ have the property.

   - The property "$x \in D$" propagates with each $O_k(\vec{x}_n, y) \in \mathcal{O}$: So let each of the x_i have a derivation $\boxed{\ldots, x_i}$. We show that so does y.

     Concatenating all these derivations we get a derivation (7.2.4)

     $$\boxed{\ldots, x_1}, \ldots, \boxed{\ldots, x_i}, \ldots, \boxed{\ldots, x_n} \tag{1}$$

     But then so is

     $$\boxed{\ldots, x_1}, \ldots, \boxed{\ldots, x_i}, \ldots, \boxed{\ldots, x_n}, y \tag{2}$$

     by 7.2.1, case 2. That is, y is *derived*, hence $y \in D$ is proved (I.S.).

2. $D \subseteq \mathrm{Cl}(\mathcal{I}, \mathcal{O})$: Let $x \in D$. This time we do good old-fashioned CVI over $\mathbb{N}$ on the length n of a derivation of x, toward showing that $x \in \mathrm{Cl}(\mathcal{I}, \mathcal{O})$ —this is the "property of x" that we prove.

   Basis. $n = 1$. The only way to have a 1-element derivation is that $x \in \mathcal{I}$.

   Thus, $x \in \mathcal{I} \subseteq \mathrm{Cl}(\mathcal{I}, \mathcal{O})$ by 7.0.2.

   I.H. Fix n and Assume the claim for x derived with length $k < n$ —i.e., such an x is in $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$.

   I.S. Prove that the claim holds when x has a derivation of length n.

   Consider such a derivation

$$a_1, \ldots a_i, \ldots, a_k, \ldots, \overset{a_n}{\overset{\|}{x}}$$

   If $x \in \mathcal{I}$, then we are done by the *Basis*.

   Otherwise, say x is the result of an operation (relation) $O_r \in \mathcal{O}$, applied on entries to the left of x, that is, say that $O_r(\ldots, x)$ is true —where we did not (have to) show the inputs.

   By the I.H. the inputs of O_r[*] all are in $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$. Now, since this closure is closed under $O_r(\ldots, x)$, we have that the output x is in $\mathrm{Cl}(\mathcal{I}, \mathcal{O})$ too. □

[*] They all have derivations of lengths $< n$ since they are to the left of x.

So now we have two *equivalent* (7.2.5) approaches to defining inductively defined sets S:

As $S = \mathrm{Cl}(\mathcal{I}, \mathcal{O})$ or as $S = \{x : x \text{ is } (\mathcal{I}, \mathcal{O})\text{-derived}\}$.

The first approach is best when you want to prove *properties of all members of the set S*. The second is best when you want to *show $x \in S$*, for some specific x.

7.2.6 Example. Let $A = \{a, b\}$.

Let $\mathcal{I} = \{\lambda\}$, let $\mathcal{O}$ consist of one operation R:

$$X \longrightarrow \boxed{R} \longrightarrow aXb \tag{3}$$

We claim that $\mathrm{Cl}(\mathcal{I}, \mathcal{O}) = \{a^n b^n : n \geq 0\}$, where for any string X,

$$X^n \overset{Def}{=} \underbrace{XX\ldots X}_{n \text{ copies of } X}$$

If $n = 0$, "0 copies of X" means λ.

Let us write $S = \{a^n b^n : n \geq 0\}$.

1. For $\mathrm{Cl}(\mathcal{I}, \mathcal{O}) \subseteq S$ we do induction over the closure to prove that any $x \in \mathrm{Cl}(\mathcal{I}, \mathcal{O})$ satisfies $x \in S$ ("the property").

   - Well, if $x \in \mathcal{I}$ then $x = \lambda = a^0 b^0$. Done.

   - The property propagates with R.

     For example, say $X = a^n b^n \in S$. Using (3) we see that the output, aXb, is $a^{n+1} b^{n+1} \in S$. *The property does propagate!* Done.

2. For $\mathrm{Cl}(\mathcal{I},\mathcal{O}) \supseteq S$ we could do induction over $\mathbb{N}$ on n in $x = a^n b^n$ (arbitrary member of S) to prove that any $x \in S$ satisfies $x \in \mathrm{Cl}(\mathcal{I},\mathcal{O})$ ("the property").

   But this would be proving again (!!) 7.2.5 rather than USING it!

So will use 7.2.5:

Here is a derivation of $x = a^n b^n$ for each $n \geq 0$, thus it is in $\mathrm{Cl}(\mathcal{I},\mathcal{O})$ for any $n \geq 0$.

$$\lambda = a^0b^0, a^1b^1, a^2b^2, a^3b^3, \ldots, \overbrace{a^{i+1}b^{i+1}}^{\text{application of rule to } a^ib^i}, \ldots, a^nb^n$$

7.2.7 Example. “Can we show also $\mathbb{N} \subseteq \mathrm{Cl}(\mathcal{I}, \mathcal{O})$? **Yes**” asks (and answers) Example 7.1.4.

Unlike there let us show this here by a derivation, avoiding the SI induction we did there:

To show $n \in \mathrm{Cl}(\{0\}, \{x \mapsto x + 1\})$.

Well, here is a derivation of n:

$$0, 1, 2, \ldots, \quad \overbrace{i + 1}^{\text{via Op } i \mapsto i+1} \quad , \ldots n$$

□

7.2.8 Example. Regarding the example of simple arithmetic expressions (p.315) we prove that each such expression has equal numbers of left and right brackets.

Here $\mathcal{I} = \{1, 2, 3\}$ and the two operations (single-valued) are

$$\begin{array}{c} E \\ \longrightarrow \\ \longrightarrow \\ E' \end{array} \boxed{\quad + \quad} \longrightarrow (E + E') \tag{1}$$

and

$$\begin{array}{c} E \\ \longrightarrow \\ \longrightarrow \\ E' \end{array} \boxed{\quad \times \quad} \longrightarrow (E \times E') \tag{2}$$

Well, each of the members of $\mathcal{I}$ has the claimed property.

The property is preserved by each of (1) and (2). For example,

$$\boxed{\text{If } E \text{ and } E' \text{ have the property so does ``}(E + E')\text{''}}$$

since we added just one left bracket and one right bracket to the already existing brackets of E and E'.

Similarly for $\times$. □

Bibliography

[Dav65] M. Davis, *The undecidable*, Raven Press, Hewlett, NY, 1965.

[DS90] Edsger W. Dijkstra and Carel S. Scholten, *Predicate Calculus and Program Semantics*, Springer-Verlag, New York, 1990.

[GS94] David Gries and Fred B. Schneider, *A Logical Approach to Discrete Math*, Springer-Verlag, New York, 1994.

[Kle43] S.C. Kleene, *Recursive predicates and quantifiers*, Transactions of the Amer. Math. Soc. **53** (1943), 41–73, [Also in [Dav65], 255–287].

[Kur63] A.G. Kurosh, *Lectures on General Algebra*, Chelsea Publishing Company, New York, 1963.

[Men87] Elliott Mendelson, *Introduction to Mathematical Logic*, 3rd ed., Wadsworth & Brooks, Monterey, CA, 1987.

[Sho67] Joseph R. Shoenfield, *Mathematical Logic*, Addison-Wesley, Reading, MA, 1967.

[Tou03a] G. Tourlakis, *Lectures in Logic and Set Theory, Volume 1: Mathematical Logic*, Cambridge University Press, Cambridge, 2003.

[Tou03b] ———, *Lectures in Logic and Set Theory, Volume 2: Set Theory*, Cambridge University Press, Cambridge, 2003.

[Tou08] ______, *Mathematical Logic*, John Wiley & Sons, Hoboken, NJ, 2008.