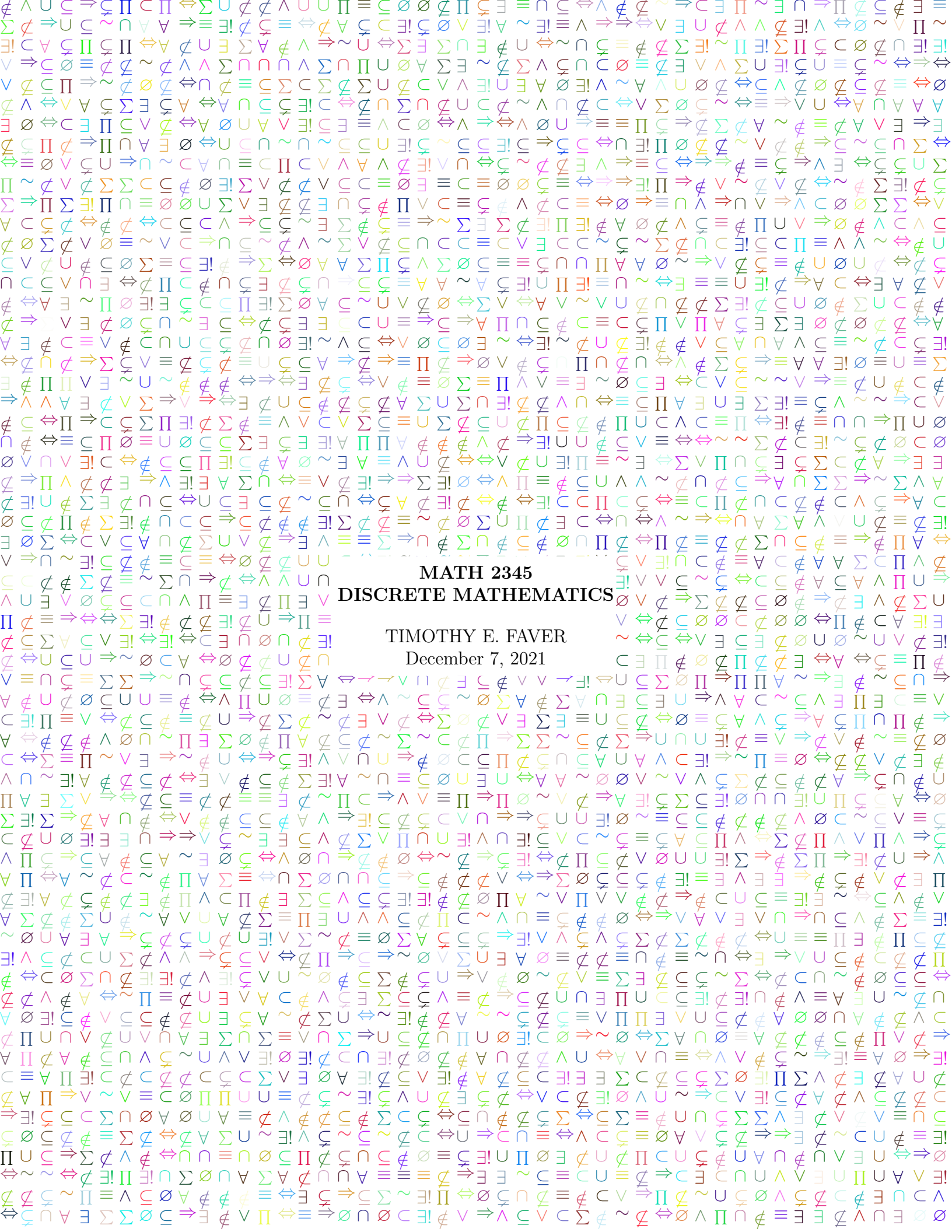
MATH 2345
DISCRETE MATHEMATICS

TIMOTHY E. FAVER

December 7, 2021



Contents

1. Introduction

"Better than 'indiscreet' mathematics."

—Paul Lukacs, Associate Professor of English, Loyola University Maryland

First of all, why "discrete"? Epp's preface[1] describes discrete math as "processes that consist of a sequence of individual steps." But this is true for much of math: first we do one step, then another, then another, and then maybe we solve the problem. What is particular to the realm of "discrete math" is less the step-by-step nature of the *processes* and more the *mathematical concepts and structures* to which the processes are applied. Namely, many of these concepts and structures are inherently discrete: they consist of distinct, separate components.

Consider the following situation.

1.0.1 Example.

There are 4 first-year students in an orientation small group. Each student shakes the hand of every other student in the group.

(i) *How many handshakes took place?*

(ii) *What is discrete about this?*

(iii) *Should students exchange unprotected handshakes with strangers?*

Solution. (i) Denote the four students impersonally by S_1, S_2, S_3, and S_4. Here are all the pairs of handshakes (no one shakes hands with oneself).

$$\begin{array}{llll} S_1 \longleftrightarrow S_2 & S_2 \longleftrightarrow S_1 & S_3 \longleftrightarrow S_1 & S_4 \longleftrightarrow S_1 \\ S_1 \longleftrightarrow S_3 & S_2 \longleftrightarrow S_3 & S_3 \longleftrightarrow S_2 & S_4 \longleftrightarrow S_2 \\ S_1 \longleftrightarrow S_4 & S_2 \longleftrightarrow S_4 & S_3 \longleftrightarrow S_4 & S_4 \longleftrightarrow S_3 \end{array}$$

It appears that 12 handshakes took place. But when S_1 shakes hands with, say, S_2, surely that is the same as when S_2 shakes hands with S_1. So, going down the columns above from left to right, we cancel any repetitions.

$$\begin{array}{llll} S_1 \longleftrightarrow S_2 & \cancel{S_2 \longleftrightarrow S_1} & \cancel{S_3 \longleftrightarrow S_1} & \cancel{S_4 \longleftrightarrow S_1} \\ S_1 \longleftrightarrow S_3 & S_2 \longleftrightarrow S_3 & \cancel{S_3 \longleftrightarrow S_2} & \cancel{S_4 \longleftrightarrow S_2} \\ S_1 \longleftrightarrow S_4 & S_2 \longleftrightarrow S_4 & S_3 \longleftrightarrow S_4 & \cancel{S_4 \longleftrightarrow S_3} \end{array}$$

We are left with only six handshakes occuring.

(ii) We are dealing with distinct entities and actions: four students and a finite number of handshakes. Each student exists separately from the other three, and each handshake takes place in isolation from the others. We do not have a continuum of handshakes or students;

[1] Our main course text: *Discrete Mathematics with Applications, Fifth Edition* by Susanna Epp.

we are not saying anything like "for each real number x satisfying $0 \leq x \leq 1$ there is a student $x \ldots$"

(iii) Meh. ▲

Here are generalizations of the previous example: given a set of n elements, how many ways can we pick k elements from that set, if the order in which we pick them does not matter? If the order does[2]? If we are allowed to pick the same element multiple times in each of the previous cases (perhaps each time we make a choice, we are choosing the color of an article of clothing, and we are allowed to repeat colors in our attire)?

For that matter, what is a set? An element? What does it mean for a set to have n elements? And why *can* we always pick elements out of a set? These are among the many questions that we shall address — although we will not be able to answer all of them, even ones as simple as "What is a set?"

In contrast to all of this, calculus is often the mathematics of *continuous* processes and components: if we know how a process behaves at time t, how similar is its behavior at time $t+t_1$, where t_1 is small? How large can we make t_1 before the behavior starts to change? We will not need any calculus in this course — indeed, we will barely need $(x+y)^2 = x^2+2xy+y^2$, although it will be helpful to know that $x + x = 2x$ — but the tools that we will develop in the areas of proof and logical reasoning will serve in many subsequent classes, discrete or continuous. In particular, the notions of what is a set, and what is a function, and how might we dare to go about *proving* something transcend this course and touch all other aspects of mathematics.

This is where we finished on Monday, August 16, 2021.

[2]Here is one easy case in the event that order does not matter $k = n$. How many ways can we pick n elements from a set of n elements, if the order in which we choose does not matter? In this case, we are just picking the entire set, and so we have exactly 1 choice.

2. Sentences, Statements, and Logic

2.1. Sentences and statements.

> Mathematics, like ice cream and politics, is discussed in sentences. This is not always immediately apparent, because mathematical sentences may be presented in eccentric formats.
>
> —Larry Gerstein, *Introduction to Mathematical Structures and Proofs*

We begin with some fundamentals of mathematical grammar, and we will think far more about writing and phrasing than calculating or symbol-pushing for now. We will not define the word "**sentence**[3]," but we will know it when we see it. Sentences need not contain any words spelled out in letters and may consist solely of symbols.

2.1.1 Example.

The following are all sentences.

(i) *All even numbers are multiples of 2.*

(ii) $2 + 2 = 5$.

(iii) $x^2 + y^2 = 1$.

Determining whether a given sentence is true or false (or neither true nor false, in which case it is not a statement) comprises much of mathematics.

2.1.2 Example.

Which of the sentences in Example 2.1.1 are true? False? Neither?

Solution. **(i)** "All even numbers are multiples of 2" — this sentence is true because this is the definition of an even number, assuming we agree on what the word "multiple" means (a number n is even if there is a number m such that $n = 2m$). Consequently, this sentence is a statement.

(ii) "$2 + 2 = 5$" — this sentence is false because we can do arithmetic. But it has a truth value, so it is a statement.

(iii) "$x^2 + y^2 = 1$" — this sentence is neither true nor false because we do not know what x and y are. Presumably x and y are numbers, so that squaring and adding them makes sense. Perhaps $x = 1$ and $y = 0$, in which case the sentence is a true statement; perhaps $x = y = 1$,

[3] We *could* try to define it. Perhaps, "a sentence is a grammatical structure containing a subject and a verb" — but what do "grammatical structure," "subject," and "verb" mean? Wikipedia has a nice definition under `Sentence (mathematical logic)`, but the first "sentence" in this definition contains at least four other concepts that we would need to define. It is not impossible to give a good definition of "sentence," just complicated.

in which case the sentence is a false statement. More information is needed before we can ascribe a truth value to this sentence. ▲

We are therefore led to consider those sentences which have definite truth values — **TRUE** or **FALSE**, although we will not define these words, either.

2.1.3 Definition.

A **STATEMENT** *is a sentence with a truth value, true or false. One and only one of the following possibilities holds for any statement:*

(i) *The statement is true.*

(ii) *The statement is false.*

2.2. Logical connectives and truth tables.

> "And it is also said," answered Frodo, "*Go not to the Elves for counsel, for they will say both no and yes.*"
>
> —J. R. R. Tolkien, *The Lord of the Rings*

There are a number of fundamental operations or "logical connectives" that we can perform on/apply to statements. Typically we will use letters like P, Q, and R to refer to statements. We might say something like the following: let P be the statement "$1 + 1 = 2$; then P is true." We might also say "It is the case that P." This latter option may sound abrupt and wrong; usually after "that" we expect a complete sentence — but remember that P now *is* a sentence.

2.2.1. The three fundamental connectives: $\sim$, $\wedge$, and $\vee$.

Our first operation on statements is the "not" operation. Attaching "not" to a statement should flip (or reverse, or invert) its truth value.

2.2.1 Definition.

If P is a statement, then the **NEGATION** *of P is the statement "It is not the case that P." We denote the negation of P by $\sim P$. If P is true, then $\sim P$ is (defined to be) false; if P is false, then $\sim P$ is (defined to be) true.*

2.2.2 Example.

If P is the statement "x is a positive real number," then $\sim P$ is the statement "It is not the case that x is a positive real number." Of course, this statement means the same as the statement "x is not a positive real number," and usually we can find several ways of phrasing a negation (with one perhaps being more palatable than another). In fact, $\sim P$ is true if and only if the more succinct statement "$x \leq 0$" is true.

We can summarize the behavior of the **NEGATION OPERATOR** $\sim$ in the following truth table.

P	$\sim P$
T	F
F	T

In general, a **TRUTH TABLE** is a rectangular array of symbols in which the top row consists of several statements, one per column, and in the columns below are the truth values of those statements, denoted T or F. Typically the statements in the left-most columns are the "given" statements and the ones to the right are statements "constructed" from those given statements.

Our next operation on statements is the "and" operation. We expect that if we combine two statements by "and," then both statements have to be true for the combination to be true. For example, the statements "2 is a positive number" and "2 is even" are both true, so the statement "2 is a positive number and 2 is even" (which we might condense as "2 is positive and even") is true. But the statement "2 is greater than 3" is false, so the statement "2 is positive and 2 is greater than 3" is false.

2.2.3 Definition.

The **CONJUNCTION** *of the statements P and Q is the statement "P and Q," which we denote by $P \wedge Q$. We define $P \wedge Q$ to be true when P and Q are both true and false if at least one of P or Q is false.*

Here is the truth table for $\wedge$.

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	F	F
F	T	F

Immediately we see that this table is larger than the one for $\sim P$. We now have two "given" statements, P and Q, and, with two values (T or F) each for P and Q, this leads to four cases to consider.

2.2.4 Remark.

When writing a truth table for an operation that involves multiple statements, start the first row with each given statement as true. Then change one truth value in each subsequent row; ideally, all but one of the truth values in any two consecutive rows should be the same. If there are n given statements involved, the table will have 2^n rows[4].

Our final operation on statements is the "or" operation. Often we use "or" in the context of **EXCLUSIVE** possibilities, like "A number is even or it is odd," or "The dinner options are steak or lobster." A number can be even or odd but not both; the conference budget allows participants to order steak or lobster but not both. Other times our "or" is **INCLUSIVE**: "The prerequisite for this course is MATH 1112 or MATH 1113" means that one must take

[4]We will eventually prove this when we study counting.

MATH 1112 or MATH 1113 to take this course, but one will not be barred from the course for having taken *both* MATH 1112 and MATH 1113[5]. We will use this inclusive "or" in mathematics.

How should we define the truth values of "P or Q" in terms of those of P and Q and respecting this inclusivity? Consider the statements

"2 is even," "2 is positive," and "2 is greater than 3."

The first two are true; the third is false. We intuitively know, then, that the statements "2 is even or 2 is positive" and "2 is even or 2 is greater than 3" are both true, since at least one of the "component" statements is true. However, a statement like "2 is greater than 3 or 2 is odd" is false since both components are false. This motivates the following definition of "or."

2.2.5 Definition.

The **DISJUNCTION** *of the statements P and Q is the statement "P or Q," which we denote by $P \vee Q$. We define $P \vee Q$ to be true if at least one of P or Q is true and false only if both P and Q are false.*

Here is the truth table for $\vee$.

P	Q	$P \vee Q$
T	T	T
T	F	T
F	F	F
F	T	T

2.2.6 Remark.

Our definitions of $\sim$, $\wedge$, and $\vee$ are just that: definitions. We have not used any prior mathematics to prove that these operators should have the truth values above; rather, we used our human intuition and experience to decide on these definitions.

This is where we finished on Wednesday, August 18, 2021.

2.2.2. Statement forms.

Out of the three operations $\sim$, $\wedge$, and $\vee$ we can form many statements more statements. In particular, we can iterate and combine $\sim$, $\wedge$, and $\vee$.

2.2.7 Definition.

A **STATEMENT FORM** *is an expression consisting of statements and logical connectives. If two statement forms $\mathcal{P}$ and $\mathcal{Q}$ always have the same truth values when their component statements[6] have the same truth values, then $\mathcal{P}$ and $\mathcal{Q}$ are* **EQUIVALENT**, *and we write*

[5] Of course, one may be barred from taking both of these courses in the first place.

$\mathcal{P} \equiv \mathcal{Q}$.

We will always use parentheses or square brackets to indicate order of operations[7]; when writing a truth table, evaluate the expressions inside nested parentheses/square brackets first. For example, we would never write $\sim \sim P$ but rather $\sim(\sim P)$.

2.2.8 Example.

Write the truth table for each of the following statement forms.

(i) $\sim(\sim P)$

(ii) $\sim(P \wedge Q)$

(iii) $P \vee (Q \vee R)$

Solution. (i) It is often helpful to break a statement into multiple parts and find the truth values for each of those parts separately (how precisely one does this is something of a matter of taste). In the table below we separate the ancillary part $\sim P$ with vertical lines from the given statement P and the final statement $\sim(\sim P)$.

P	$\sim P$	$\sim(\sim P)$
T	F	T
F	T	F

We observe that $P \equiv \sim(\sim P)$, and so two negatives do make a positive.

(ii)

P	Q	$P \wedge Q$	$\sim(P \wedge Q)$
T	T	T	F
T	F	F	T
F	F	F	T
F	T	F	T

(iii)

P	Q	R	$Q \vee R$	$P \vee (Q \vee R)$
T	T	T	T	T
T	T	F	T	T
T	F	F	F	T
F	F	F	F	F
F	F	T	T	T
F	T	T	T	T
F	T	F	T	T
T	F	T	T	T

▲

[6] If it feels like a statement form is a function whose inputs are statements and whose outputs are "true" or "false," that is essentially what is happening.

[7] Order of operations — "parentheses, exponentials, multiplication, division, addition, subtraction" ascribes meaning to $6 \div 2 \cdot 1 + 2$, but writing a number like this is sheer moral turpitude. We will not commit acts of moral turpitude with logical connectives.

With sufficient practice, determining the truth values of a statement form will feel as familiar as arithmetic. For that matter, let us momentarily recall some fundamental properties of arithmetic.

2.2.9 Theorem (Arithmetic).

Let a, b, and c be real numbers. Then

(i) *Commutativity of addition:* $a + b = b + a$

(ii) *Commutativity of multiplication:* $a \cdot b = b \cdot a$

(iii) *Associativity of addition:* $(a + b) + c = a + (b + c)$

(iv) *Associativity of multiplication:* $a \cdot (b \cdot c) = (a \cdot b) \cdot c$

(v) *Distribution of multiplication over addition:* $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$

In particular, "commutativity" is an "order" property, while "associativity" is a "grouping" property.

The logical connectives $\wedge$ and $\vee$ behave in much the same way.

2.2.10 Theorem ("Algebraic" properties of $\wedge$ and $\vee$).

Let P, Q, and R be statements. Then

(i) *Commutativity of* $\wedge$*:* $P \wedge Q \equiv Q \wedge P$

(ii) *Commutativity of* $\vee$*:* $P \vee Q \equiv Q \vee P$

(iii) *Associativity of* $\wedge$*:* $(P \wedge Q) \wedge R \equiv P \wedge (Q \wedge R)$

(iv) *Associativity of* $\vee$*:* $(P \vee Q) \vee R \equiv P \vee (Q \vee R)$

(v) *Distribution of* $\wedge$ *over* $\vee$*:* $P \wedge (Q \vee R) \equiv (P \wedge Q) \vee (P \wedge R)$

(vi) *Distribution of* $\vee$ *over* $\wedge$*:* $P \vee (Q \wedge R) \equiv (P \vee Q) \wedge (P \vee R)$

We leave the proofs as exercises with truth tables (note that each table will have 8 sets of truth values). However, the validity of these equivalences should be apparent by casting them into ordinary (but informal) language. For example, if P is true and Q is true, then "P and Q" is true. But the order in which we are saying truths shouldn't matter, so "Q and P" must be true, too. Be careful, however, not to attempt associativity with both $\wedge$ and $\vee$; for example, $(P \wedge Q) \vee R$ is not equivalent to $P \wedge (Q \vee R)$, and so the expression $P \wedge Q \vee R$ is undefined and meaningless.

2.2.11 Remark.

We have met our first theorem above in Theorem 2.2.10. What is a **THEOREM***? Broadly, any mathematical statement that has been proved (proven?) true. Depending on the writer's taste and the hierarchy of statements in the given document, a theorem may also be called a* **PROPOSITION***; the word* **LEMMA** *usually refers to an auxiliary theorem that helps to prove a "larger" theorem, while* **COROLLARY** *is usually a consequence of a theorem. (Sometimes the hard work goes into the lemmas, the really interesting results appear in the corollaries, and the theorems get all the praise.)*

And what is a **PROOF** *of a statement? Any convincing, correct argument that the statement is true. To say that this course will teach proof writing is grandiose, and absurd; this is a lifetime's work. But we will focus on broadly common techniques and strategies for proofs so that, in time, the command "Prove this statement" is not as intimidating and opaque as it might be right now.*

There are a number of equivalent statement forms that are worth remembering; see Theorem 2.1.1 in Epp's book. We highlight two particularly famous ones, the first of which we already established as part (ii) of Example 2.2.8.

2.2.12 Theorem (De Morgan).

Let P and Q be statements. Then

(i) $\sim(P \wedge Q) \equiv \sim P \vee \sim Q$.

(ii) $\sim(P \vee Q) \equiv \sim P \wedge \sim Q$.

We have used the word "theorem" for the first time. In general, a **THEOREM** is a statement that has been shown to be always true. For example, part (i) is really the statement "$\sim(P \wedge Q) \equiv \sim P \vee \sim Q$ for any statements P and Q." The **PROOF** of a theorem is any convincing and correct verification that its statement is true.

There are plenty of other logical connectives that we use in everyday language — for example, "but," "unless," "only if" — and that we will *not* use in our treatment of mathematics. For us, $\sim$, $\wedge$, and $\vee$ will largely suffice (along with three others that we will meet shortly). We close by mentioning two other words that we will use frequently in conjunction with statements, statement forms, and truth values.

2.2.13 Definition.

(i) *A statement form is a* **CONTRADICTION** *if it is always false regardless of the truth values of its component statement(s).*

(ii) *A statement form is a* **TAUTOLOGY** *if it is always true regardless of the truth values of its component statement(s).*

2.2.14 Example.

Show that $P \wedge (\sim P)$ is a contradiction. Can you think of a related tautology?

Solution. We construct a truth table.

P	$\sim P$	$P \wedge (\sim P)$
T	F	F
F	T	F

This is not surprising: a statement is either true or false and never both. Thus $P \wedge (\sim P)$ must be a contradiction.

Since a tautology is always supposed to be true, and since the negation of false is true, we claim that $\sim(P \wedge (\sim P))$ is a tautology. Appending an extra column for $\sim(P \wedge (\sim P))$ to the truth table above will quickly show that. We could also use De Morgan's laws (with $Q = \sim P$) first to obtain

$$\sim(P \wedge (\sim P)) \equiv (\sim P) \vee [\sim(\sim P)] \equiv (\sim P) \vee P$$

and then to construct the truth table for $(\sim P) \vee P$.

These results respect our intuition: given a statement P, it is always the (exclusive) case that either P is true or P is false, so either P is true or $\sim P$ is true, and never both. ▲

This is where we finished on Friday, August 20, 2021.

2.2.15 Remark.

There are a number of other connectives that we use in daily language — for example, "but," "while," "unless," and the "exclusive or" — that we could plausibly define as statement forms via the logical connectives $\sim$, $\wedge$, and $\vee$. We will not need such precise descriptions in this course, but a good exercise is to (1) think of the truth values that the statement form should have (this may be something of a matter of opinion) and (2) cook up a combination of statements and $\sim$, $\wedge$, and $\vee$ such that the corresponding truth table has the truth values from (1).

2.3. Conditional statements.

> "Please, remember — the alethiometer does not *forecast*; it says, '*If* certain things come about, *then* the consequences will be —' and so on."
>
> —Philip Pullman, *His Dark Materials*

Much of mathematics boils down to having two statements, P and Q, where P is known to be true, and where we want to show that Q must also be true, typically somehow as a consequence of the truth of P. (Perhaps in the absence of P the statement Q might not be true.) In other words, a frequent goal of mathematical life is to assume that the statement P is true and then to show that the statement "If P, then Q" is true.

But what does the statement "If P, then Q" mean?

- P implies Q?
- Q can be logically deduced from P?
- If P is true, then so is Q?
- It cannot be the case that Q is false when P is true?

These are all good candidates for explaining "If P, then Q," and there are surely others, but most of these beg further questions. What does "implies" mean? Or "logically deduced"? Or "then so"? Only the fourth option offers us the opportunity to cast "If P, then Q" in terms of prior logical connectives. If this statement means that we cannot have P true and Q false, then we expect that "If P, then Q" should mean that it is *not* the case that we have P and not Q.

2.3.1 Definition.

The **CONDITIONAL OF** Q **BY** P *is the statement "If P, then Q," and we denote it by*

$$(P \Longrightarrow Q) :\equiv \sim[P \wedge (\sim Q)].$$

The statement P is the **HYPOTHESIS**[8] *and the statement Q is the* **CONCLUSION**.

2.3.2 Remark.

We have previously used the symbol $\equiv$ to connect two equivalent sentential forms $\mathcal{P}$ and $\mathcal{Q}$. We now use the symbol $:\equiv$ to indicate that the new sentential form on the left of $:\equiv$ is being defined by the expression on the right side of $:\equiv$.

Let us calculate the truth table for $P \Longrightarrow Q$.

P	Q	$\sim Q$	$P \wedge (\sim Q)$	$\sim[P \wedge (\sim Q)]$
T	T	F	F	T
T	F	T	T	F
F	F	T	F	T
F	T	F	F	T

This truth table respects our desire that the following cases should coincide: (1) the truth of P, the truth of Q, and the truth of $P \Longrightarrow Q$ and (2) the truth of P, the falsity of Q, and the falsity of $P \Longrightarrow Q$. However, the bottom two rows contain the different cases when P is false. Perhaps surprisingly, $P \Longrightarrow Q$ is always true whenever P is false, regardless of whether or not P is true!

In other words, a false hypothesis and a true conclusion together make a true statement. This is surprising, and uncomfortable, but we should consider that our development of

[8] In mathematics, the word "hypothesis" is typically used to mean something that we are assuming to be true, not something whose validity we are trying to decide, whereas in the (other) sciences one might run an experiment to test the validity of a hypothesis.

$P \Longrightarrow Q$ was only based on how Q should behave if P is true; we made no requirements for the behavior of Q when P was false. Moreover, this behavior of $P \Longrightarrow Q$ forces us to acknowledge a difference between the *meaning* of the statement "If P, then Q" and the *process* of deducing Q from P.

2.3.3 Remark.

The symbol $\Longrightarrow$, contrary to how we typically use it, does not mean "follows logically from." That is, at this point in our development of logic, the statement $P \Longrightarrow Q$ has nothing to do with a rational argument telling us how to go from P to Q! We deduce the truth or falsity of $P \Longrightarrow Q$ by considering the truth/falsity of P and Q separately from each other.

2.3.4 Example.

Which of the following statements are true?

(i) *If Atlanta is a city in Georgia, then* $1 + 1 = 2$.

(ii) *If Washington, D.C., is a state, then* $1 + 1 = 2$.

(iii) *If* $1 + 1 = 2$, *then* $0 = 1$.

Solution. **(i)** Let P be the statement "Atlanta is a city in Georgia" and Q the statement $1 + 1 = 2$. Both are true, so $P \Longrightarrow Q$ is true.

(ii) Let P be the statement "Washington, D.C., is a state" and Q be the statement $1+1 = 2$. Then P is false, so $P \Longrightarrow Q$ is true. (Incidentally, Q is true, too.)

(iii) Let P be the statement $1 + 1 = 2$ and Q be the statement $0 = 1$. Then P is true and Q is false, so $P \Longrightarrow Q$ is false. ▲

2.3.5 Example.

Show that the statements $P \Longrightarrow (Q \vee R)$ and $[P \wedge (\sim Q)] \Longrightarrow R$ are equivalent.

Solution. Here is why we expect these statements to be equivalent: $P \Longrightarrow (Q \vee R)$ essentially means "If P, then Q or R." From our knowledge of if-then, we know that it cannot be the case that we have P and not at least one of Q and R; thus if we have P but not Q, we must have R.

We could write out truth tables, but let us demonstrate this more deftly using the algebra of logical connectives. First, by definition

$$[P \Longrightarrow (Q \vee R)] \equiv \sim[P \wedge (\sim(Q \vee R))].$$

We use De Morgan's laws to rewrite

$$\sim[P \wedge (\sim(Q \vee R))] \equiv (\sim P) \vee (\sim(\sim(Q \vee R))),$$

and then we find

$$(\sim P) \vee (\sim(\sim(Q \vee R))) \equiv (\sim P) \vee (Q \vee R).$$

All together,

$$[P \Longrightarrow (Q \vee R)] \equiv [(\sim P) \vee (Q \vee R)].$$

It seems that we have reduced our first statement as far as we can without "regrouping" or "factoring." Now we work on the other statement: by definition,

$$[(P \wedge (\sim Q)) \Longrightarrow R] \equiv \sim[(P \wedge (\sim Q)) \wedge (\sim R)].$$

We use De Morgan's laws:

$$\sim[(P \wedge (\sim Q)) \wedge (\sim R)] \equiv (\sim(P \wedge (\sim Q))) \vee (\sim(\sim R)).$$

We use De Morgan's laws again:

$$[(\sim(P \wedge (\sim Q))) \vee (\sim(\sim R))] \equiv ((\sim P) \vee (\sim(\sim Q))) \vee R \equiv (\sim P) \vee Q \vee R.$$

And this is what we had before. ▲

This is where we finished on Monday, August 23, 2021.

2.3.6 Example.

Suppose that I claim the following: "If it is raining, then I am wearing a raincoat." At any moment in time, one, and only one, of the following four circumstances must be the case. When am I a liar?

(i) *It is raining and I am not wearing a raincoat.*

(ii) *It is not raining and I am wearing a raincoat.*

(iii) *It is not raining and I am not wearing a raincoat.*

(iv) *It is raining and I am wearing a raincoat.*

Solution. **(i)** I am not doing what I promised: it is raining but I am not wearing a raincoat. So in this case I am a liar.

(ii) I never said what I would do when it is not raining, so I have not broken my word here.

(iii) Again, I never said anything about what I wear in times other than rain.

(iv) This is what I said I would do, and now I am doing it.

The four cases above suggest, at the level of our intuition rather than symbol-pushing with $\sim(P \wedge (\sim Q))$, why $P \Longrightarrow Q$ should be true whenever P is false. ▲

There is an equivalent form of the statement $P \Longrightarrow Q$ whose truth value is sometimes easier to ascertain.

2.3.7 Theorem.

The CONTRAPOSITIVE *of* $P \implies Q$ *is* $(\sim Q) \implies (\sim P)$*, and these statements are equivalent:*

$$[P \implies Q] \equiv [(\sim Q) \implies (\sim P)].$$

Proof. From the definition of $\implies$, we have

$$[(\sim Q) \implies (\sim P)] \equiv \sim[(\sim Q) \wedge (\sim(\sim P))].$$

We first simplify

$$[(\sim Q) \wedge (\sim(\sim P))] \equiv [(\sim Q) \wedge P].$$

Then

$$\sim[(\sim Q) \wedge (\sim(\sim P))] \equiv \sim[(\sim Q) \wedge P] \equiv \sim[P \wedge (\sim Q)] \equiv [P \implies Q].$$ ■

Flipping the statement $P \implies Q$ to the statement $Q \implies P$ or introducing negations to form $\sim P \implies \sim Q$ produces two different classes of statements from the original $P \implies Q$.

2.3.8 Definition.

The CONVERSE *of the statement* $P \implies Q$ *is the statement* $Q \implies P$*. The* INVERSE *of the statement* $P \implies Q$ *is the statement* $\sim P \implies \sim Q$*.*

2.3.9 Remark.

We will almost never use the words "converse" or "inverse" outside of meeting them here. One way to remember which is which is that the "inverse" of addition (a.k.a. subtraction) is "adding a negative," and so the inverse is the one with $\sim$*.*

2.3.10 Example.

Discuss the contrapositive, converse, and inverse of the statement "If it is raining, then I am wearing a raincoat."

Solution. This statement has the form $P \implies Q$, where P is the statement "It is raining," and Q is the statement "I am wearing a raincoat." We negate these statements: $\sim P$ is "It is not raining" and $\sim Q$ is "I am not wearing a raincoat."

The contrapositive is $(\sim Q) \implies (\sim P)$, which is the statement "If I am not wearing a raincoat, then it is not raining." If I tell the truth when I say that "If it is raining, then I am wearing a raincoat," then the contrapositive should also be true: if I am not wearing a raincoat and it is raining, then I am a liar.

The converse is $Q \implies P$, which is the statement "If I am wearing a raincoat, then it is raining." Although we must take care not to associate a *causal* relationship between Q and P in the statement "If Q, then P" (i.e., we should not expect to be able to deduce P from Q), nonetheless this converse need not be true. After all, the mere act of donning a raincoat need not summon rain, and I may well want to wear my raincoat before a shower or after.

The inverse is $(\sim P) \implies (\sim Q)$, which is the statement "If it is not raining, then I am not wearing a raincoat." Again, it is certainly possible that I wear my raincoat even when it is not raining.

The point of this example is to suggest, in ordinary language, that an if-then statement may be true but its converse and inverse may be false. ▲

This is where we finished on Wednesday, August 25, 2021.

2.3.11 Example.

(i) *If the converse is true, must the original conditional be true?*

(ii) *If the inverse is true, must the original conditional be true?*

Solution. First we work out all four of the major statements involving "$\implies$."

Conditional	$P \implies Q$	$\sim P \vee Q$
Contrapositive	$\sim Q \implies \sim P$	$Q \vee \sim P$
Converse	$Q \implies P$	$\sim Q \vee P$
Inverse	$\sim P \implies \sim Q$	$P \vee \sim Q$

From the second column we see that the inverse is the contrapositive of the converse. Thus the conditional and the contrapositive are equivalent, and the converse and the inverse are equivalent. Importantly (and, sometimes, unfortunately) the original conditional and its converse/inverse are *not* equivalent.

Now we address the questions above.

(i) We know that $P \implies Q$ is false if and only if P is true and Q is false. But if Q is false, then $Q \implies P$ is automatically true. Thus the converse can be true while the conditional is false.

(ii) Again, no, since the inverse and the converse are equivalent. (Alternatively, let Q be false and P be true; then the converse is true and the original conditional is false.) ▲

We sometimes encounter the situation in which we know the truth values of both "If P, then Q" and "If Q, then P." So, we want a logical connective to encapsulate the statement "$P \implies Q$ and $Q \implies P$." This is easy to define.

2.3.12 Definition.

The **BICONDITIONAL OF** P **AND** Q *is the statement*

$$P \iff Q :\equiv (P \implies Q) \wedge (Q \implies P).$$

We sometimes say that Q *is* **NECESSARY AND SUFFICIENT** *for* P.

A good amount of mathematics involves "cleaning things up": we have some complicated statement P and we want to find a simpler statement Q that means the same as P. That

is, we want to find a statement Q that is necessary and sufficient for P. For example, the statements "x is a nonzero real number satisfying $x^2 - x = 0$" and "$x = 1$" are equivalent, and one is much more succinct than the other.

2.3.13 Example.

Show that $P \iff Q$ and $Q \iff P$ are equivalent.

Solution. To show that two sentential forms are equivalent, one recourse is always to write out the truth tables for each sentential form and show that when the given truth values (one might say, the "inputs") are the same, then the truth values of the sentential forms (the "outputs") are also the same. Here, however, we can use definitions:

$$Q \iff P \equiv (Q \Longrightarrow P) \wedge (P \Longrightarrow Q).$$

Now use the "commutativity" of $\wedge$ ($R \wedge S \equiv S \wedge R$) to find

$$(Q \Longrightarrow P) \wedge (P \Longrightarrow Q) \equiv (P \Longrightarrow Q) \wedge (Q \Longrightarrow P) \equiv P \iff Q. \qquad \blacktriangle$$

2.3.14 Remark.

(i) *We sometimes write P iff Q for $P \iff Q$.*

(ii) *The statement $P \iff Q$ is sometimes read as "P is* NECESSARY AND SUFFICIENT *for Q."*

(iii) *Epp uses $\rightarrow$ for $\Longrightarrow$, $\leftarrow$ for $\Longleftarrow$, and $\leftrightarrow$ for $\iff$. Epp also prefers lowercase letters for statements (i.e., p instead of P).*

2.4. Predicates and quantifiers.

The statements "Every number greater than or equal to 2 has a unique factorization into primes" and "There are two real numbers whose square is 4" involve two new notions of mathematical grammar. First, the mathematical "objects" in these sentences are not listed out explicitly — "every number" instead of 1, 2, 3, and so on and "two real numbers" instead of 2 and -2 — but rather referred to obliquely as members of an abstract set. Second, both statements "quantify" the existence and properties of certain mathematical objects; the first sentence encompasses a very large amount of objects ("every number"), and the second a much smaller set ("there are two"). This motivates our study both of predicates, which are statements with variables, and quantifiers, which are phrases that tell us "how much" or "how many" objects satisfy certain properties.

2.4.1. Predicates.

2.4.1 Definition.

A PREDICATE *is a sentence that contains finitely many symbols, called* VARIABLES*, which becomes a statement when the variables have specific values. The set of values for*

the predicate is the **DOMAIN** *of the predicate. In other words, a predicate is a sentence containing variables that becomes true or false when the variables are specified.*

2.4.2 Example.

(i) *The sentence "$x < 1$," where x is assumed to be a real number, is not a statement, since it does not have a definite truth value. But this is a predicate. Specifically, let $P(x)$ be the predicate "$x < 1$," where we will take the domain D to be the set of all real numbers. Then $P(x)$ is either true or false for a given $x \in D$. In particular, $P(0)$ is true, $P(1)$ is false, and $P(2)$ is false.*

(ii) *Let $P(x, y)$ be the predicate $x^2 + y^2 = 1$, where x and y are assumed to be numbers. Then $P(1, 0)$ and $P(0, 1)$ are true statements, while $P(1, 1)$ is false.'*

This is where we finished on Friday, August 27, 2021.

(iii) *Let D be the set of all students enrolled at KSU this fall and for a student x let $P(x)$ be the predicate "x is taking a math class." Then $P(x)$ is true whenever x is a student in this class. Given a particular student x, we know that $P(x)$ must be true or false for that student x, but we ourselves (as students, professors, and law-abiding citizens) may be unable to determine the truth or falsity of $P(x)$ without violating privacy rights and digging into restricted records.*

Suppose that $P(x)$ is a predicate in the single variable x and that D is the domain[9] for P. If x is one of these values[10], we will write $x \in D$, read "x is in D" or "x belongs to D." If y is a value that is not in D, then we write $y \notin D$.

In general, we should feel free to read and write more expansively "It is the case that $P(x)$" whenever we see a predicate $P(x)$ by itself. Of course, a sentence of the form "It is the case that $P(x)$" probably feels incomplete when spoken aloud, but remember that $P(x)$ is really a sentence by itself.

We will use the following sets as domains frequently enough that we give them special notation.

2.4.3 Definition.

(i) *We denote[11] the set of all* **REAL NUMBERS** *by $\mathbb{R}$.*

(ii) *We denote the set of all* **INTEGERS** *by $\mathbb{Z}$. Informally, $\mathbb{Z}$ consists of all numbers of the form 0, ± 1, ± 2, $\pm 3, \ldots$.*

(iii) *We denote the set of all positive integers by $\mathbb{N}$. That is, $\mathbb{N}$ consists of all numbers of*

[9] If this feels like we are dealing with a function, we are. This function assigns values in D to the truth values "true" or "false."

[10] Here we are using the symbol x to denote both the abstract variable x of the predicate $P(x)$ and particular elements of the domain D. This is essentially what we do with functions, after all.

the form 1, 2, 3, ... We often call the numbers in $\mathbb{N}$ *the* **NATURAL NUMBERS** *(and less commonly the* **WHOLE NUMBERS***). Importantly, for us,* $0 \notin \mathbb{N}$.

We will discuss the properties of $\mathbb{N}$ and $\mathbb{Z}$ in further detail in Section 4.

2.4.4 Example.

We illustrate some elementary set notation:

$$100 \in \mathbb{Z}, \qquad -25 \in \mathbb{Z}, \qquad 100 \in \mathbb{N}, \qquad 0 \notin N, \qquad \textit{and} \qquad 50 \in \mathbb{R}.$$

2.4.2. Quantifiers.

Suppose that $P(x)$ is a predicate with domain D. Perhaps it is the case that $P(x)$ is true for all $x \in D$, or maybe $P(x)$ is false for all $x \in D$, or maybe there are some x for which $P(x)$ is true and others for which $P(x)$ is false. The first two possibilities are "universal" in that they describe how $P(x)$ behaves for *all* x, while the third is "existential" — there *exist some* x for which $P(x)$ is true and some for which $P(x)$ is false. We begin the process of converting these "quantified" possibilities into mathematical notation.

2.4.5 Definition.

Let $P(x)$ *be a predicate with domain* D. *The* **EXISTENTIAL STATEMENT** *"There exists* $x \in D$ *such that* $P(x)$*" is true if there is at least one* $x \in D$ *such that* $P(x)$ *is true. It is false if* $P(x)$ *is false for all* $x \in D$. *We may abbreviate this statement by*

$$\exists x \in D \text{ such that } P(x) \quad \textit{or by} \quad \exists x \in D : P(x) \quad \textit{or by} \quad (\exists x \in D)P(x).$$

The symbol $\exists$ *is called the* **EXISTENTIAL QUANTIFIER**.

2.4.6 Example.

Decide whether the following existential statements are true or false.

(i) $\exists x \in \mathbb{R}$ *such that* $x^3 = 27$.

(ii) $\exists x \in \mathbb{R}$ *such that* $x^2 = 4$.

(iii) $\exists x \in \mathbb{R}$ *such that* $x^2 = -1$.

Solution. (i) We need to show that there exists at least one $x \in \mathbb{R}$ such that $x^3 = 27$. After a moment's thought we see that $x = 3$ works. This is, incidentally, the only such x that works, too.

(ii) We need to show that there exists at least one $x \in \mathbb{R}$ such that $x^2 = 4$. We can take either $x = 2$ or $x = -2$.

[11] The font style that we use for $\mathbb{R}$, $\mathbb{Z}$, and $\mathbb{N}$ is called "blackboard bold." Some authors denote these sets by ordinary bold capital letters, i.e., **R**, **Z**, and **N**. We will not do that.

(iii) We need to show that there exists at least one $x \in \mathbb{R}$ such that $x^2 = -1$. But we know that $x^2 \geq 0$ for all $x \in \mathbb{R}$, so we cannot find $x \in \mathbb{R}$ such that $x^2 = -1$. Thus this statement is false. ▲

It may be the case that $P(x)$ is a predicate with domain D for which there exists exactly one $x \in D$ such that $P(x)$ is true. We might also say that such an x is **UNIQUE**.

2.4.7 Definition.

The statement "There exists a unique $x \in D$ such that $P(x)$" is true if

(i) *The statement $\exists x \in D : P(x)$ is true.*

(ii) *If $P(x_1)$ and $P(x_2)$ are true for some x_1, $x_2 \in D$, then $x_1 = x_2$.*

The statement "There exists a unique $x \in D$ such that $P(x)$" is false if either (i) or (ii) above are false. We abbreviate this statement by

$$\exists! x \in D \text{ such that } P(x) \quad \text{or by} \quad \exists! x \in D : P(x) \quad \text{or by} \quad (\exists! x \in D) P(x).$$

2.4.8 Example.

Show that the statement $\exists! x \in \mathbb{R} : x + 1 = 2$ is true.

Solution. First we need to show existence ($\exists$): just take $x = 1$ and add $1 + 1 = 2$. Now we show uniqueness: $x = 1$ is the only value of x that works. Suppose that y is some (other) real number such that $y + 1 = 2$. Subtract 1 from both sides to find $y = 1$. ▲

2.4.9 Remark.

Proving $\exists!$ requires two steps: showing existence of at least one and then showing that that one is the only one. In other words, $\exists! = \exists +!$. Often the two steps can be done independently of each other. A common way to prove uniqueness is to assume that there are two values x_1 and x_2 for which the property under consideration is true and then to show, somehow, that $x_1 = x_2$.

This is where we finished on Monday, August 30, 2021.

2.4.10 Example.

The statement "$\exists! x \in \mathbb{R} : x^2 = 4$" is false. How can we change it to make it true?

Solution. When considering uniqueness questions for the truth of a predicate, a common problem is that the domain is too large. If we restrict the domain sufficiently, we may be able to exclude all but one value at which the predicate is true.

Here, the problem is that both $x = 2$ and $x = -2$ satisfy $x^2 = 4$. If we restrict ourselves to positive solutions of this equation, then we get uniqueness. Let $(0, \infty)$ denote, as usual,

the set of all positive numbers. Then the statement

$$\exists! x \in (0, \infty) : x^2 = 4$$

is true, and it holds true when $x = 2$. ▲

Yet another possibility for a predicate $P(x)$ with domain D is that $P(x)$ is true for every $x \in D$. For example, let $P(x)$ be the statement "$x^2 \geq 0$ with domain $\mathbb{R}$. Then, certainly, $P(x)$ is true for every $x \in \mathbb{R}$.

2.4.11 Definition.

The **UNIVERSAL STATEMENT** *"For every $x \in D$ it is the case that $P(x)$" is true if $P(x)$ is true for every $x \in D$. We abbreviate this statement by*

$$(\forall x \in D)P(x) \quad \text{or by} \quad \forall x \in D, P(x) \quad \text{or by} \quad \forall x \in D : P(x).$$

The symbol $\forall$ is the **UNIVERSAL QUANTIFIER**. *Sometimes we flip the order of the phrases in the universal statement and say "It is the case that $P(x)$ for all $x \in D$." An $x \in D$ for which $P(x)$ is false is a* **COUNTEREXAMPLE** *to the universal statement $\forall x \in D : P(x)$.*

2.4.12 Example.

Decide whether the following universal statements are true or false.

(i) $\forall x \in \mathbb{R} : x^2 + 1 > 0$.

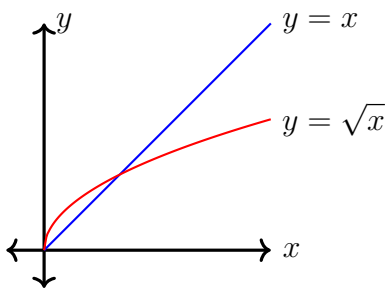
(ii) $\forall x \in \mathbb{R} : x > \sqrt{x}$.

(iii) *Every integer that is a multiple of 2 is also a multiple of 4.*

(iv) $\forall x > 2 : x > 1$.

Solution. (i)

(ii) If we think carefully, we will see (maybe from a graph) that this statement is false.



Take $x = 1/4$. Then $\sqrt{x} = 1/2 > 1/4 = x$.

(iii) The integer 6 is divisible by 2 but not by 4, so this statement is false. Here is how we can recast this statement with quantifiers: let $P(x)$ be the predicate "The integer x is a

multiple of 2" and $Q(x)$ be the predicate "The integer x is a multiple of 4." Then the (false) statement reads

$$\forall x \in \mathbb{Z} : P(x) \implies Q(x).$$

(iv) We are being coy with the notation here: instead of writing $x \in D$ for some domain D, we are saying "$\forall x > 2$." Here the domain D is the set of all real numbers strictly greater than 2, i.e., the interval $(2, \infty)$. So this statement could also be written as "$\forall x \in (2, \infty) : x > 1$."

Since $1 < 2$, if $x < 1$, then $x < 2$ as well. This statement is true. ▲

2.4.13 Remark.

The colon (:) is a valuable piece of punctuation. We pronounce it differently in an existential statement from how we do in a universal statement. Namely, in $\exists x \in D : P(x)$, the colon should be read as "such that." But in $\forall x \in D : P(x)$, the colon should be read as "it is the case that." Note that $\exists$ contains the verb of its sentence ("exists") but $\forall$ just starts a prepositional phrase.

This is where we finished on Wednesday, September 1, 2021 (Section 54).

2.4.3. Interaction of multiple quantifiers.

Many mathematical statements contain multiple quantifiers and predicates containing multiple variables, and the order in which each appears is important. We read such statements from left to right. For simplicity, typically we do not write a colon (:) after each quantifier but only after the last quantifier, before the predicate.

2.4.14 Example.

Rewrite the following statement using quantifiers and predicates: "Every nonnegative real number x has a nonnegative square root." Is this statement true? What happens if you reverse the order of the quantifiers (but keep the variables in the same order)?

Solution. This statement is equivalent to

$$\forall x \in [0, \infty)\ \exists y \in [0, \infty) : x = y^2.$$

This statement is true if we believe that the square root function $\sqrt{\cdot}$ is defined on the nonnegative real numbers. We might let $P(x, y)$ be the predicate "$x = y^2$" with domain consisting[12] of all numbers $x \in [0, \infty)$ and $y \in [0, \infty)$. Then the statement reads

$$\forall x \in [0, \infty)\ \exists y \in [0, \infty) : P(x, y).$$

By the way, sometimes when working with intervals we omit the $\in$ symbol and interval notation in a quantified statement and instead use inequalities. Thus we could also write

$$\forall x \geq 0\ \exists y \geq 0 : P(x, y).$$

[12]Later, when we have the language of ordered pairs, we will say more succinctly that the domain is the set of all $(x, y) \in [0, \infty) \times [0, \infty)$.

If we reverse the order of the quantifiers (i.e., we swap $\forall$ and $\exists$ but keep x and y in the order above), then we get the statement

$$\exists y \in [0, \infty)\ \forall x \in [0, \infty) : P(x, y).$$

In words, "There exists $x \geq 0$ such that for all $y \geq 0$ it is the case that $x = y^2$." This statement is certainly not true, as it is saying that all nonnegative numbers y square to the same number x.

Here is a counterexample: suppose there is such an x. Then this x must "work" for all y, and so we could take $y = 0$ to see that $x = 0^2 = 0$. But we could also take $y = 1$ to find $x = 1^2 = 1$. Hence $0 = 1$, which is absurd. ▲

This is where we finished on Wednesday, September 1, 2021 (Section 53).

2.4.15 Example.

If x is a positive real number that is close to 0, then the square x^2 will be close to 0, too. (Think about the graph of $y = x^2$ for $x > 0$ but close to 0. Of course, the same works for $x < 0$, too.) Analyze this idea in terms of quantifiers and predicates.

Solution. To force x^2 to be at most a certain (small) distance from 0, we just have to take x to be sufficiently small. The following statement is a quantification of this idea:

$$\forall \epsilon > 0\ \exists \delta > 0 : (0 < x < \delta \implies x^2 < \epsilon). \tag{2.4.1}$$

Here we have a predicate $P(x, \delta, \epsilon)$ that depends on three variables, x, δ, and ϵ. That is, $P(x, \delta, \epsilon)$ is the predicate $0 < x < \delta \implies x^2 < \epsilon$, which we of course read as "If $0 < x < \delta$, then $x^2 < \epsilon$." The domain of this predicate consists of all numbers $\epsilon > 0$, all numbers $\delta > 0$, and all numbers $0 < x < \delta$. This is a complicated domain, and the statement (2.4.1) is *not* an obvious expression of our original idea.

How would we show that the statement (2.4.1) is true? We would first have to take an arbitrary $\epsilon > 0$. Then we would have to find a number $\delta > 0$ and show that whenever $0 < x < \delta$, then it is also the case that $x^2 < \epsilon$.

Working backward, this is not too hard to do: we have $x^2 < \epsilon$ if and only if $x < \sqrt{\epsilon}$. So if $0 < x < \sqrt{\epsilon}$, then $x^2 < \epsilon$. And so we take $\delta = \sqrt{\epsilon}$. ▲

2.4.4. Negating statements with quantifiers and predicates.

If $P(x)$ is a predicate, then we will write $\sim P(x)$ for its negation. That is, if we take $P(x)$ as an abbreviation for "It is the case that $P(x)$," then $\sim P(x)$ means "It is not the case that $P(x)$." For example, if $P(x)$ is the predicate "$x + 1 = 2$" with domain $\mathbb{R}$, then $\sim P(x)$ means "It is not the case that $x + 1 = 2$," which we can write equivalently, and more compactly, as $x + 1 \neq 2$.

We now apply negations to quantified statements. Throughout, let $P(x)$ be a predicate with domain D.

First we negate the existential quantifier. Recall that $\exists x \in D : P(x)$ means that "There exists $x \in D$ for which it is the case that $P(x)$." What does the negation of this statement,

$$\sim(\exists x \in D : P(x)),$$

mean? First, it means "It is not the case that there exists $x \in D$ such that $P(x)$." So, if $x \in D$, then it cannot be the case that $P(x)$, and so it must be the case that $\sim P(x)$. Since x was an arbitrary element of D, it must be the case that $\sim P(x)$ for all $x \in D$. And so we have established the following.

2.4.16 Theorem.

Let $P(x)$ be a predicate with domain D. Then

$$\sim(\exists x \in D : P(x)) \equiv \forall x \in D : \sim P(x).$$

2.4.17 Example.

(i) *Write the statement "There exists a real number whose square is -1" using quantifiers and mathematical symbols but no English words.*

(ii) *Write the negation of the statement "There exists a real number whose square is -1" using quantifiers and mathematical symbols but no English words.*

(iii) *Write the negation of the statement "There exists a real number whose square is -1" using English words but no mathematical symbols.*

Solution. (i) $\exists x \in \mathbb{R} : x^2 = -1$.

(ii) Let $P(x)$ be the predicate "$x^2 = -1$." Then our statement compresses to $\exists x \in \mathbb{R} : P(x)$, and so its negation is the statement "$\forall x \in \mathbb{R} : \sim P(x)$. We see $\sim P(x)$ is the equivalent to the statement $x^2 \neq -1$, and so the negation is also the statement "$\forall x \in \mathbb{R} : x^2 \neq -1$."

(iii) Two possibilities for the negation are "All real numbers do not square to -1" and "No real number has square equal to -1." There are many others. ▲

Now we negate the universal quantifier. Recall that $\forall x \in D : P(x)$ means that "For all $x \in D$ it is the case that $P(x)$." Thus $\sim(\forall x \in D : P(x))$ means that "It is not the case that for all $x \in D$, it is the case that $P(x)$." And so there must be some element $x \in D$ for which it is not the case that $P(x)$. We conclude the following.

2.4.18 Theorem.

Let $P(x)$ be a predicate with domain D. Then

$$\sim(\forall x \in D : P(x)) \equiv \exists x \in D : \sim P(x).$$

2.4.19 Example.

Let P be the statement "Every real number is positive." (This is a false statement.)

(i) *Write P using quantifiers and mathematical symbols but no English words.*

(ii) *Write $\sim P$ using quantifiers and mathematical symbols but no English words.*

(iii) *Write $\sim P$ using English words but no mathematical symbols.*

Solution. **(i)** $\forall x \in \mathbb{R} : x > 0$.

(ii) We just write it all out in symbols:

$$\sim(\forall x \in \mathbb{R} : x > 0) \equiv \exists x \in \mathbb{R} : \sim(x > 0) \equiv \exists x \in \mathbb{R} : x \leq 0.$$

The negation of the statement "x is positive" is "It is not the case that x is positive" or "x is not positive." From our implicit understanding of the order structure of the real numbers, this means that x could be 0 or negative: $x = 0$ or $x < 0$, which is the same as $x \leq 0$. In particular, "x is not positive" is *not* equivalent to "x is negative."

(iii) "There exists a number $x \in \mathbb{R}$ such that x is less than or equal to zero." Sometimes this is called being "nonpositive." ▲

2.4.20 Example.

Let P be the statement "Every KSU student is a math major." Let Q be the statement "No KSU student is a math major." Do we have $\sim P \equiv Q$?

Solution. The statement $\sim P$ is "It is not the case that every KSU is a math major," and so $\sim P$ will be true if we can find just one KSU student who is not a math major. But $\sim P$ will continue to be true if we find two KSU students, one who is a math major, and one who is not. However, in this case Q. And so it is possible for $\sim P$ to be true while Q is false, and so $\sim P$ and Q are not equivalent.

Here is another way to view these statements using quantifiers. Let $M(x)$ be the predicate "x is a math major," where the domain of R is the set of all KSU students. Then P is the statement $\forall x \in D : M(x)$. Hence $\sim P$ is the statement

$$\sim(\forall x \in D : M(x)) \equiv \exists x \in D : \sim M(x).$$

Now, $\sim M(x)$ is the predicate "It is not the case that $M(x)$," which is to say, "It is not the case that x is a math major." Thus $\sim P$ is the statement "There exists $x \in D$ such that it is not the case that $M(x)$." More plainly, $\sim P$ is the statement "There exists a KSU student who is not a math major."

On the other hand, Q is the statement "For all KSU students, no student is a math major." (More normally, Q is the statement "No KSU student is a math major." The horror!) That is, Q is the statement $\forall x \in D : \sim M(x)$. And so we see that $\sim P$ and Q hinge on different quantifiers: $\sim P$ is existential, while Q is universal. ▲

This is where we finished on Friday, September 3, 2021 (Section 53).

Negating the quantifier $\exists!$ is somewhat different. This is because $\exists!$ is really the conjunction of two statements: existence and uniqueness. Thus to negate $\exists!$, one would have to show that either the object in question does not exist at all, or multiple versions of it exist. This is not rendered so nicely by flipping $\exists!$ to another symbol (like $\sim\forall \mapsto \exists$ and $\sim\exists \mapsto \forall$).

2.4.21 Example.

Let $P \equiv \exists! x \in \mathbb{R} : x^3 = 27$. (This is true.) How can we express $\sim P$?

Solution. The statement P means two things: there exists an $x \in \mathbb{R}$ such that $x^3 = 27$ and that this x is unique. So

$$\exists! x \in \mathbb{R} : x^3 = 27 \equiv (\exists x \in \mathbb{R} : x^3 = 27) \wedge (\text{“This } x \text{ is unique.”})$$

The negation is then

$$\sim(\exists! x \in \mathbb{R} : x^3 = 27) \equiv \sim(\exists x \in \mathbb{R} : x^3 = 27) \vee \sim(\text{“This } x \text{ is unique.”}).$$

We negate the existential statement easily:

$$\sim(\exists x \in \mathbb{R} : x^3 = 27) \equiv \forall x \in \mathbb{R} : x^3 \neq 27.$$

To negate the uniqueness statement, we should think about what the statement “This x is not unique” means in the context of cubes equaling 27. Lack of uniqueness means that at least two distinct numbers cube to 27. We might write this symbolically as

$$(\exists x_1 \in \mathbb{R} : x_1^3 = 27) \wedge (\exists x_2 \in \mathbb{R} : x_2^3 = 27 \wedge x_1 \neq x_2).$$

This is hideous, so we condense[13] the two existence statements into one:

$$\exists x_1, x_2 \in \mathbb{R} : x_1 \neq x_2 \wedge x_1^3 = x_2^3 = 27./$$

And so the negation of our original statement is

$$\sim(\exists! x \in \mathbb{R} : x^3 = 27) \equiv (\forall x \in \mathbb{R} : x^3 \neq 27) \vee (\exists x_1, x_2 \in \mathbb{R} : x_1 \neq x_2 \wedge x_1^3 = x_2^3 = 27).$$

This, too, is hideous, but it captures the negation of $\exists!$: either there exists nothing satisfy the property, or there exist more than one “thing” satisfying the property. ▲

[13] What we really want here, and what we have not developed, is the language of the ordered pair. We do not want to say “There exists $x_1 \in \mathbb{R}$ such that there exists $x_2 \in \mathbb{R}$ such that $x_1 \neq x_2$ and $x_1^3 = x_2^3 = 27$.” This suggests a dependence of x_2 on x_1. Rather, we want the *simultaneous* existence of x_1 and x_2 satisfying $x_1 \neq x_2$ and $x_1^3 = x_2^3 = 27$. A cleaner way to state this is to say that there exists a *pair* (x_1, x_2) such that $x_1 \neq x_2$ and $x_1^3 = x_2^3 = 27$. But what is an ordered pair? We must develop some set theory to understand that.

Suppose now that $P(x, y)$ is a predicate with variables $x \in D$ and $y \in E$. How should we negate a statement involving P with multiple quantifiers, like

$$\forall x \in D\ \exists y \in E : P(x, y)?$$

For a given x, let $Q(x, y)$ be the statement $\exists y \in E : P(x, y)$. Then our original statement is $\forall x \in D : Q(x, y)$, and so its negation is

$$\sim(\forall x \in D : Q(x, y)) \equiv \exists x \in D : \sim Q(x, y).$$

Now we negate $Q(x, y)$:

$$\sim Q(x, y) \equiv \sim(\exists y \in E : P(x, y)) \equiv \forall y \in E : \sim P(x, y).$$

Does this make sense in plain English? The original statement $\forall x \in D \exists y \in E : P(x, y)$ means that for all $x \in D$, there is some $y \in E$ for which it is the case that $P(x, y)$. So to negate this statement, there must exist some $x \in D$ such that for all $y \in E$, it is not the case that $P(x, y)$.

If we replace $P(x, y)$ with $\sim P(x, y)$, we find

$$\sim(\forall x \in D\ \exists y \in E : \sim P(x, y)) \equiv \exists x \in D\ \forall y \in E : \sim(\sim P(x, y)) \equiv \exists x \in D\ \forall y \in E : P(x, y).$$

Negating both sides of the equivalence above, we obtain

$$\sim(\exists x \in D\ \forall y \in E : P(x, y)) \equiv \forall x \in D\ \exists y \in E : \sim P(x, y).$$

2.4.22 Theorem.

Let $P(x, y)$ be a predicate with values $x \in D$ and $y \in E$. Then

$$\sim(\forall x \in D\ \exists y \in E : P(x, y)) \equiv \exists x \in D\ \forall y \in E : \sim P(x, y)$$

and

$$\sim(\exists x \in D\ \forall y \in E : P(x, y)) \equiv \forall x \in D\ \exists y \in E : \sim P(x, y).$$

This is where we finished on Friday, September 3, 2021 (Section 54).

2.4.23 Example.

Let D be the set of KSU students and E the set of math classes offered at KSU. Let $P(x, y)$ be the predicate "The student x has taken the math class y."

(i) *The statement "$\forall x \in D\ \exists y \in E : P(x, y)$" means that "For all KSU students x, there is a math class y such that student x has taken the class y." More informally, it means "Every KSU student has taken some math class." And so the negation of this statement means that some KSU student has never taken a math class!*

(ii) *The informal statement "Every KSU student has taken the same math class" is different: it means "There is a math class that every KSU student has taken," or, symbolically, "$\exists y \in E\ \forall x \in D : P(x, y)$."*

This is where we finished on Wednesday, September 8, 2021 (Section 53).

2.4.24 Corollary.

Suppose that P is a statement that contains an arbitrary (but finite) number of quantifiers, including possibly the same quantifier repeated multiple times in a row. Then $\sim P$ is formed by flipping each $\exists$ to $\forall$ and each $\forall$ to $\exists$ and last negating the predicate at the end.

2.4.25 Example.

Negate each of the following statements.

(i) $\forall x_1 \in D_1\ \forall x_2 \in D_2\ \exists x_3 \in D_3 : P(x_1, x_2, x_3)$.

(ii) $\exists x_1 \in D_1\ \exists x_2 \in D_2\ \forall x_3 \in D_3\ \exists x_4 \in D_4 : P(x_1, x_2, x_3, x_4)$.

(iii) *Every state in the U.S. has a university at which there is a student majoring in math.*

Solution. **(i)** $\exists x_1 \in D_1\ \exists x_2 \in D_2\ \forall x_3 \in D_3 : \sim P(x_1, x_2, x_3)$.

(ii) $\forall x_1 \in D_1\ \forall x_2 \in D_2\ \exists x_3 \in D_3\ \forall x_4 \in D_4 : \exists P(x_1, x_2, x_3, x_4)$.

(iii) The signal words here are "every," which suggests the quantifier "for all," and "has" and "there is," which suggest the quantifier "there exist." The sentence contains three "objects" that are modified by the (implicit) quantifiers; states, universities, and students. There is something of a successive dependence among these objects. The universities are located within particular states, and the students are attending particular universities.

So, let S be the set of all U.S. states and, for $s \in S$, let $U(s)$ be the set of all universities within state s. For a given university u, let $P(u)$ be the set of all students at university u. Finally, for a student p at the university u in the state s, let $M(p, u, s)$ be the predicate "p is a math major at the university u in the state s." Then our statement is equivalent to

$$\forall s \in S\ \exists u \in U(s)\ \exists p \in P(u) : M(p, u, s).$$

Hence its negation is

$$\exists s \in S\ \forall u \in U(s)\ \forall p \in P(u) : \sim M(p, u, s).$$

Less ornately, the negation is the statement "There is a state in the U.S. in which no university has any math majors."

What is the domain of M? The sentence "p is a math major at the university u in the state s" makes sense whenever p is a student, u is a university, and s is a state. Now, for any

particular student p, the sentence "p is a math major at Loyola University Maryland in the state of Georgia" is false because Loyola MD is not in Georgia. But the point is that M has a *truth value* p is a student, u is a university, and s is a state. And so one can argue that the domain of M is the set of all students p, all universities u, and all states s. We will revisit this set in the context of abstract set theory later. For M to be a *true* statement, it would have to be the case that p is a math major, p is a student at university u, and university u is located within state s.

▲

This is where we finished on Wednesday, September 8, 2021 (Section 54).

2.4.26 Example.

Let $P(x)$ and $Q(x)$ be predicates with domain D.

(i) *Negate the statement*

$$\forall x \in D : P(x) \Longrightarrow Q(x).$$

Do not use $\Longrightarrow$ in your final expression of the negation.

(ii) *Write out both the original statement and the negation in words with no existential or logical connective symbols, just words.*

Solution. **(i)** First we negate as usual:

$$\sim(\forall x \in D : P(x) \Longrightarrow Q(x)) \equiv \exists x \in D : \sim(P(x) \Longrightarrow Q(x)).$$

Now, recall that $P(x) \Longrightarrow Q(x) \equiv \sim(P(x) \wedge \sim Q(x))$. Thus

$$\sim(P(x) \Longrightarrow Q(x)) \equiv \sim(\sim(P(x) \wedge \sim Q(x))) \equiv P(x) \wedge \sim Q(x).$$

And so

$$\sim(\forall x \in D : P(x) \Longrightarrow Q(x)) \equiv \exists x \in D : P(x) \wedge \sim Q(x).$$

(ii) The original statement means that for all x in the set D, if it is the case that $P(x)$, then it is also the case that $Q(x)$. The negation means that there exists x in D for which it is the case that $P(x)$ but not $Q(x)$. ▲

3. Proofs

3.1. Philosophical matters.

> "Mary, absorbed and happy as she fooled around with the lacquer to make her spyglass; fooling around was something she'd never been able to explain to her colleague Oliver Payne, who needed to know where he was going before he got there. Back in Oxford, she gave three of her precious wheel-tree seeds to a scientist at the Botanic Garden, a nice man who understood the importance of fooling around. The seedlings are growing well, but she refuses to tell him where they came from."
>
> —Philip Pullman, *His Dark Materials*

We have said that a **proof** is a correct and convincing argument that a statement is true. An argument that convinces its given audience of a certain truth but that contains (possibly unnoticed) errors cannot really be a proof because an outside observer who finds these errors will fail to be convinced of the truth, too. But a correct argument that does not convince its audience of the truth — possibly because the exposition of the argument is poor, or inappropriate for the given audience (maybe too advanced or too simplistic) — certainly is not a proof, either. When we write proofs, we must be careful both that the steps in our reasoning are correct and that our exposition is understandable.

There is no guaranteed way to write a proof. Very often all that we can do at first is play — or fumble, or explore, or fool around — without any definite sense of what will come precisely next, although always with a final goal in mind. Here are some questions that we should always ask in the process of constructing proofs.

1. [Starting the journey] What are the hypotheses? What are we assuming? What are we allowed to know?

2. [Ending the journey] What are the conclusions? What do we want to be true because of the hypotheses?

3. [Along the way] What do all the words mean? What does this situation resemble?

3.2. Proofs and symbolic logic.

Much of mathematics boils down to the following situation. We have statements P and Q and we know (or we assume) that P is true. Then we want to deduce that Q must be true. In ordinary language, we would say something like "If P, then Q." The process of establishing the truth of Q directly from the truth of P is broadly, and unsurprisingly, called **direct proof**.

In our mathematical language, we know that the statement "If P, then Q," or $P \implies Q$, need not have anything to do with a "logical" connection between P and Q, even though we

often (sloppily, but naturally) pronounce $\Longrightarrow$ as "implies." After all, the statement "If Z is the first letter of the alphabet, then 2 is even" is a true statement. Nonetheless, there is a useful connection between the statement $P \Longrightarrow Q$ and the process of establishing the truth of Q from the truth of P.

Recall that if P is true, then the statement $P \Longrightarrow Q$ is true if Q is true and false if Q is false. Thus if we start out assuming that P is true, then establishing the truth of Q is equivalent to establishing the truth of $P \Longrightarrow Q$. This is useful because $P \Longrightarrow Q$ is logically equivalent to other statements, namely

$$\sim(P \wedge \sim Q) \quad \text{and} \quad \sim Q \Longrightarrow \sim P.$$

Thus if one of the following three statements is true, then the other two are true as well:

$$P \Longrightarrow Q, \qquad \sim(P \wedge \sim Q), \quad \text{and} \quad \sim Q \Longrightarrow \sim P. \tag{3.2.1}$$

To show that $\sim(P \wedge \sim Q)$ is true, it suffices to show that $P \wedge \sim Q$ is false. One way to do this is to assume that $P \wedge \sim Q$ is true and conclude that another obviously false statement must be true. This is **PROOF BY CONTRADICTION**. For example, we might end up showing that a statement of the form $R \wedge (\sim R)$ is true; symbolically, this would be showing that the if-then statement

$$(P \wedge \sim Q) \Longrightarrow (R \wedge \sim R)$$

is true for some statement R.

Proving instead that the contrapositive statement

$$\sim Q \Longrightarrow \sim P$$

is, unsurprisingly, called **PROOF BY CONTRAPOSITIVE**. For concrete P and Q it may well be the case that proving the contrapositive is a much easier task than proving the original conditional.

Knowing that the three statements in (3.2.1) are logically equivalent does not help us in the slightest to establish the truth of one of them for given, concrete, particular statements P and Q. This is best illustrated by examples — lots and lots of examples — to which we now turn.

4. (Very) Elementary Number Theory

> "In Eregion long ago many Elven-rings were made, magic rings as you call them, and they were, of course, of various kinds: some more potent and some less. The lesser rings were only essays in the craft before it was full grown, and to the Elven-smiths they were but trifles — yet still to my mind dangerous for mortals."
>
> —J. R. R. Tolkien, *The Lord of the Rings*

Our purpose in studying (very) elementary number theory is not so much to learn deep properties of the integers but to produce some "essays in the craft" of proof-writing in the context of fairly uncomplicated[14] definitions, notation, and concepts. Hopefully the notions of even and odd, prime and composite numbers have been familiar to us for years. In our subsequent topics we will encounter far more unfamiliar definitions, notation, and concepts that we will have to master along with developing our proof-writing techniques.

Here in (very) elementary number theory we will pass through proofs in three stages: a fumbling, catch-as-catch-can, awkward stage of trying to figure out the flow of our logic; a formal, precise writing of the proof; and a detailed parsing of the style and writing techniques involved in that formal proof.

4.1. Fundamental definitions and properties of $\mathbb{N}$ and $\mathbb{Z}$.

We recall that the **integers** are all numbers of the form 0, ± 1, $\pm 2, \ldots$, and we denote the set of integers by $\mathbb{Z}$. The **natural numbers** are the positive integers (1, 2, 3, ...), and we denote the set of natural numbers by $\mathbb{N}$. We emphasize that $0 \notin \mathbb{N}$.

We assume all the "usual" rules of arithmetic (cf. Appendix A1 in Epp's text) and do not belabor them here. When dealing with "abstract" integers, we denote multiplication by juxtaposition, i.e., the product of m and n is mn. When dealing with "concrete" integers, we denote multiplication by $\cdot$: the product of 2 and 3 is $2 \cdot 3 = 6$.

4.2. Parity.

4.2.1 Definition.

(i) *An integer n is* **even** *if there exists an integer j such that $n = 2j$.*

(ii) *An integer n is* **odd** *if there exists an integer k such that $n = 2k + 1$.*

The **parity** *of an integer is its state of being even or odd.*

This is where we finished on Friday, September 10, 2021.

This definition does not immediately require that any integer *has* a parity, i.e., that an integer must be even or odd. We will prove this later. This definition also does not

[14] Taken out of, or within, context, this could be deeply insulting to number theorists everywhere.

immediately require parity is *exclusive*, i.e., that an integer cannot be both even and odd[15]. And, finally, this definition does not say anything about how parity alternates, i.e., that successive integers n and $n+1$ cannot have the same parity. We will, once again, have to (get to) prove this.

4.2.2 Example.

The sum of two even integers is even.

Proof Sketch.

- We introduce notation: let m_1 and m_2 be even integers.
- We state our goal: we want to show that $m_1 + m_2$ is even.
- We rephrase our goal: we want to find an integer N such that $m_1 + m_2 = 2N$.
- We ask what we know: that m_1 and m_2 are even.
- We ask what this means: that $m_1 = 2n_1$ and $m_2 = 2n_2$ for some integers n_1 and n_2. (Note that we have no right to assume $n_1 = n_2$, for then $m_1 = m_2$. But this is not one of our hypotheses.)
- We relate what we know to our goal: we express the sum $m_1 + m_2$ as

$$m_1 + m_2 = 2n_1 + 2n_2 = 2(n_1 + n_2).$$

- We recognize what we have found: we put $N = n_1 + n_2$, so N is an integer and $m_1 + m_2 = 2N$.
- We recognize that we are done.

Formal Proof. Let m_1 and m_2 be even. Then there are integers m_1 and m_2 such that $m_1 = 2n_1$ and $m_2 = 2n_2$. We add

$$m_1 + m_2 = 2n_1 + 2n_2 = 2(n_1 + n_2)$$

and put $N = n_1 + n_2$. Then N is an integer such that $m_1 + m_2 = 2N$, and so $m_1 + m_2$ is even.

[15] If an integer must be even or odd, and an integer cannot be both even and odd, then any given integer is even and not odd, or odd and not even. We will actually cheat and use this fact before we have proven it. Namely, it is fairly easy to show, using only Definition 4.2.1, that an integer is not both even and odd. It is somewhat harder, using only this definition, to show that an integer must be even or odd. This will be easier with the division algorithm, which we will meet later.

Proof Analysis. Let m_1 and m_2 be even. [State the hypothesis.] Then there are integers m_1 and m_2 such that $m_1 = 2n_1$ and $m_2 = 2n_2$. [State the essential consequence of the hypothesis.] We add

$$m_1 + m_2 = 2n_1 + 2n_2 = 2(n_1 + n_2)$$

and put $N = n_1 + n_2$. [Do math. Since this is a computation that takes more than one equals sign, we display it on a separate line. Note that we did not jump from saying that m_1 and m_2 are even to this calculation; we introduced n_1 and n_2 first.] Then N is an integer such that $m_1 + m_2 = 2N$, and so $m_1 + m_2$ is even. [Draw the final conclusions. Since $m_1 + m_2 = 2N$ is a short computation, we do not put it on a separate line.]

4.2.3 Example.

If the square of an integer is odd, that integer is also odd.

Proof Sketch.

- Introduce notation/hypothesis: let n be an integer such that n^2 is odd.
- State the desired conclusion: n is odd.
- Rephrase the hypothesis/ask what we know: there is an integer k such that $n^2 = 2k + 1$.
- Rephrase the conclusion to see what we want to prove: we want an integer j such that $n = 2j + 1$.
- Try to connect the hypothesis to the conclusion: we know $n = \sqrt{n^2} = \sqrt{2k + 1}$, but this does not factor in any natural way to yield an expression of the form $n = 2j + 1$.
- Give up for a while.
- Try a different proof direction: the contrapositive. We must show that if an integer is not odd, then its square is also not odd. This has the advantage of allowing us to assume something about the integer rather than its more complicated square.
- Cheat: assume for the moment that "not odd" really means "even." This still needs proof, and we better hope that that proof does not depend on squares of integers, lest our reasoning be circular. (It is unlikely that it will: what could the value of n^2 possibly have to do with n being simultaneously even and odd?)
- Rephrase the contrapositive that we must prove: if m is an even integer, then m^2 is also an even integer. (Here we are accepting that all integers are either even or odd.)
- Rephrase the hypothesis/conclusion of the contrapositive: if $m = 2j$, then there is

k such that $m^2 = 2k$.

- Work with the hypothesis: if $m = 2j$, then

$$m^2 = (2j)^2 = 2^2 j^2 = 4j^2 = 2(2j^2).$$

- Think about what we calculated: if we put $k = 2j^2$, then $m^2 = 2k$, and so m^2 is even.

Formal Proof. We prove the contrapositive. Assume that n is an even integer; we will show that n^2 is also even. Since n is even, we may write $n = 2j$ for some integer j. Then

$$m^2 = 2(2j^2),$$

where $2j^2$ is, of course, an integer. Thus m^2 is even.

Proof Analysis. We prove the contrapositive. [State at the start that we are using the contrapositive instead of a direct proof. We do not need to state why we are proving the contrapositive, although sometimes it may be enlightening for the reader to know why.] Assume that n is an even integer; we will show that n^2 is also even. [State the hypothesis that we will use for the contrapositive. It is helpful, but not necessary, to state the conclusion that we will prove, since this conclusion is not the same as the one in the original statement. We use the phrase "will show," or even just "show," to emphasize that we do not yet know that n^2 is even, but rather that this is the result that we will obtain,] Since n is even, we may write $n = 2j$ for some integer j. [Deduce a useful consequence of the hypothesis.] Then

$$m^2 = 2(2j^2),$$

where $2j^2$ is, of course, an integer. [Do math. Here we have compressed a lot of the arithmetic that we did in the proof sketch. How much calculation one shows depends on the expected maturity of one's readers and the intricacy, relatively speaking, of the calculation involved. It may also be appropriate to state in words, rather than mathematical symbols, some of the more complicated details of a calculation.] Thus m^2 is even. [Wrap up the proof by drawing attention to the fact that the equality $m^2 = 2(2j^2)$ means that m^2 is even.]

This is where we finished on Monday, September 13, 2021 (Section 53).

4.2.4 Example.

An integer is not both even and odd.

Proof Sketch.

• When trying to prove a "not" statement, a good strategy is to see what goes wrong if that statement is negated. Why must it not be the case that an integer is both even and odd? In other words, we elect the method of contradiction.

• Introduce notation and hypotheses: suppose that n is an integer that is both even and odd.

• Deduce a notational consequence of these hypotheses: there are integers j and k such that $n = 2j$ (evenness) and $n = 2k + 1$ (oddness).

• Think about what could go wrong: we have two disparate quantities equal to the same quantity,

$$2j = n = 2k + 1.$$

Thus the two quantities are equal.

• Try to learn something more about the new notation: since $2j = 2k + 1$, we have

$$2j - 2k = 1 \implies 2(j - k) = 1.$$

• Ask what this tells us: that 1 is even, which is intuitively absurd. We might also observe, after some different algebra above, that $j = k + 1/2$. But adding $1/2$ to an integer k cannot yield an *integer* j.

• Ask if we can quit because we know that 1 is not even: no, because we have not conclusively proved that 1 is not even. (Did we see a proof anywhere? No. Could we write a special proof that 1 is not even? Probably.) Instead, let us try some different algebra: since $2(j - k) = 1$, we have $j - k = 1/2$. But $1/2$ is not an integer, since $0 < 1/2 < 1$, while $j - k$ is an integer, since j and k are both integers. This is our contradiction.

This is where we finished on Monday, September 13, 2021 (Section 54).

The work above motivates introducing the following axiom of integers. Broadly, an **AXIOM** is a statement that we assume to be true and do not prove.

4.2.5 Axiom ("Whole number property" of $\mathbb{Z}$).

For each $n \in \mathbb{Z}$, there does not exist an $m \in \mathbb{Z}$ satisfying $n < m < n + 1$.

Intuitively, $\mathbb{Z}$ contains no "fractions" or "mixed numbers." Since we only "defined" $\mathbb{Z}$ as a list, we find it important to single this fact out, and not to try to prove it.

Formal Proof. Suppose, by way of contradiction, that n is an integer that is both

even and odd. Then there are integers j and k such that both

$$n = 2j \quad \text{and} \quad n = 2k + 1.$$

Consequently,

$$2j = 2k + 1.$$

We subtract and factor to find

$$j - k = \frac{1}{2}.$$

This is impossible, since j and k are both integers and therefore $j - k$ must be an integer. But $1/2$ is not an integer, and so we have reached a contradiction.

Proof Analysis. Suppose, by way of contradiction, that n is an integer that is both even and odd. [The phrase "by way of contradiction" is a popular expression to use at the start of a proof by contradiction. We could also start by saying "We assume that n is an integer that is both even and odd and derive a contradiction." The point is to emphasize from the start the direction of our proof: toward a contradiction. Then when we find that contradiction, we can just stop!] Then there are integers j and k such that both

$$n = 2j \quad \text{and} \quad n = 2k + 1.$$

[Derive a consequence of the (contradiction-inducing) hypothesis. We display the two equalities on a separate line to highlight their importance.] Consequently,

$$2j = 2k + 1.$$

[Derive another consequence.] We subtract and factor to find

$$j - k = \frac{1}{2}.$$

[Derive yet another consequence. Here we suppress showing a lot of the arithmetic and condense it into the words "subtract and factor."] This is impossible, since j and k are both integers and therefore $j - k$ must be an integer. But $1/2$ is not an integer, and so we have reached a contradiction. [We explain why we have reached a contradiction. Since we said at the start that we would be seeking a contradiction, and now we have found one, we really do not need to say any more.]

This is where we finished on Wednesday, September 15, 2021.

4.2.6 Example.

Consecutive integers cannot have the same parity. We state this more precisely as follows.

(i) *An integer $n \in \mathbb{Z}$ is even if and only if $n+1$ is odd.*

(ii) *An integer $n \in \mathbb{Z}$ is odd if and only if $n+1$ is even.*

Proof Sketch.

• We start by proving (i), and this amounts to establishing the truth of two if-then statements:

$$n \text{ even} \implies n+1 \text{ odd} \quad \text{and} \quad n+1 \text{ odd} \implies n \text{ even}.$$

• Let us begin with the first statement: suppose n is even. We want to show that $n+1$ is odd. All that we really have are definitions: we can write $n = 2j$ for some integer j, and we want to write $n+1 = 2k$ for some integer k.

• We could start with what we know and substitute $n = 2j$ to find $n+1 = 2j+1$, which is what we want. Or we could start with what we want and assume $n+1 = 2k+1$. Then we subtract and find $n = 2k$. Either way, it looks like $k = j$.

• In fact, fooling around as we did seems to point the way toward proving the "reverse" statement "If $n+1$ is odd, then n is even." We assume that $n+1$ is odd, write $n+1 = 2k+1$, and subtract to find $n = 2k$, so n is even.

• To prove part (ii), we could more or less repeat the work above, just changing some of the words around and writing $+1$ in different parts. But that should feel like overkill. Instead, use the equivalence of the statements $P \iff Q$ and $\sim P \iff \sim Q$. Continue to assume that "not even" is "odd" and "even" is "not odd."

Formal Proof. ($\implies$) First suppose n is even. Then we can write $n = 2j$ for some integer j, and so $n+1 = 2j+1$. By definition, $n+1$ is odd.

($\impliedby$) Now suppose $n+1$ is odd. Then $n+1 = 2j+1$ for some integer j, and so $n = 2j$. By definition, n is even.

Proof Analysis. [When proving an "if and only if" statement, it is often, although not always, helpful to write out the two "directions" separately. It can be particularly helpful to the reader to indicate in the text with ($\implies$) the beginning of the "forward" proof and with ($\impliedby$) the beginning of the "reverse" proof. Typically one writes each of these parts in separate paragraphs.]

($\implies$) First suppose n is even. [Clearly state what we are assuming here, so the reader knows in what direction we are proceeding.] Then we can write $n = 2j$ for some integer

j, and so $n+1 = 2j+1$. By definition, $n+1$ is odd. [This is a very short proof, so to make the ending less abrupt we remind the reader of the definition of an odd integer (which we presume the reader knows!), so that we do not need to do any more work after achieving the equality $n+1 = 2j+1$.]

($\Longleftarrow$) Now suppose $n+1$ is odd. [Again we state what we are assuming. We use the transition word "now" to start this sentence to alert the reader to a jump or change in the direction of our thinking.] Then $n+1 = 2j+1$ for some integer j, and so $n = 2j$. By definition, n is even.

4.3. Divisibility.

4.3.1. Fundamentals.

The notion of divisibility extends the notion of parity — we expect that a number is even if it is divisible by 2, but one may divide by many other interesting factors.

4.3.1 Definition.

An integer a is **DIVISIBLE** *by an integer b if there exists an integer c such that $a = bc$. We write $b \mid a$ and say that b is a* **FACTOR** *of a.*

4.3.2 Example.

(i) $3 \mid 12$ *since* $12 = 4 \cdot 3$.

(ii) $5 \mid -25$ *since* $-25 = 5 \cdot (-5)$.

(iii) $1 \mid n$ *for any integer n since* $n = n \cdot 1$.

We adopt the convention that stating $a \mid b$ for some $a, b \in \mathbb{Z}$ implies $a \neq 0$.

4.3.3 Example (Transitivity of divisibility).

Let $a, b, c \in \mathbb{Z}$. If $a \mid b$ and $b \mid c$, then $a \mid c$.

Proof Sketch.

• Rewrite hypotheses and conclusion from notation to something useful: $a \mid b$ means $b = am$ for some $m \in \mathbb{Z}$, while $b \mid c$ means $c = bn$ for some $n \in \mathbb{Z}$. We want to show $a \mid c$, and so we need $c = ar$ for some $r \in \mathbb{Z}$.

• Try to connect our given information to our goal: we are assuming the equalities $b = am$ and $c = bn$ and we want to get an equality of the form $c = ar$. So we need to find r.

• Play with arithmetic: if $b = am$ and $c = bn$, we can substitute $c = bn = (am)n = a(mn)$. So it looks like taking $r = mn$ works.

Formal Proof. Suppose that $a \mid b$ and $b \mid c$. Then there are integers m and n such that $b = am$ and $c = bn$. Thus

$$c = bn = (am)n = a(mn).$$

Since mn is also an integer, we conclude $a \mid c$.

Proof Analysis. Suppose that $a \mid b$ and $b \mid c$. Then there are integers m and n such that $b = am$ and $c = bn$. [State hypotheses and deduce consequences. Throughout this proof, every time we use "integer" or "integers," we could have written notationally "$\in \mathbb{Z}$." This is purely a stylistic choice here.] Thus

$$c = bn = (am)n = a(mn).$$

[We did go through the algebra in a little more detail here than in other places, as it requires (1) substituting $b = am$ and (2) rearranging the parentheses. It seemed more straightforward just to do the algebra than to describe it in words.] Since mn is also an integer, we conclude $a \mid c$. [We explain how the calculation above gets us our desired conclusion. That is, we have used both our "math" (in deriving $c = a(mn)$) and our "words" (in this final sentence) to conclude the proof. A small amount of redundancy in one's explanations can be helpful.]

Our next definition formalizes the concept of a "common factor."

4.3.4 Definition.

Let $a, b \in \mathbb{Z}$. An integer $n \in \mathbb{Z}$ is a **COMMON FACTOR** *of a and b if $n \mid a$, $n \mid b$, and $n \neq 1$.*

Of course, $1 \mid n$ for any $n \in \mathbb{Z}$, so we do not allow 1 to "count" as a common factor to avoid redundancies later.

4.3.5 Example.

(i) *2 is a common factor of 8 and 12. So is 4.*

(ii) *2, 4, and 8 are common factors of 8 and 24.*

4.3.2. Prime and composite numbers.

Among the most studied integers are the primes.

4.3.6 Definition.

(i) *An integer $p \geq 2$ is* **PRIME** *if the only factors of p are 1 and p. We restrict "primality" to integers greater than or equal to 2 to avoid technical complications, which we discuss later, with integers less than or equal to 1.*

(ii) *An integer is* **COMPOSITE** *if it is not prime.*

This definition implies that any integer $n \geq 2$ is either prime or composite (not prime). It may be helpful to cast the definitions of prime and composite into symbolic language: $p \in \mathbb{N}$ is prime if and only if

$$\forall m \in \mathbb{N} : \big(m \mid p \Longrightarrow (m = 1) \vee (m = p)\big).$$

And so an integer n is composite if and only if

$$\sim\big[\forall m \in \mathbb{N} : \big(m \mid n \Longrightarrow (m = 1) \vee (m = n)\big)\big] \equiv \exists m \in \mathbb{N} : \sim\big(m \mid n \Longrightarrow (m = 1) \vee (m = n)\big).$$

It has been some time since our journey through symbolic logic, but

$$\sim(P \Longrightarrow (Q \vee R)) \equiv \sim\big(\sim(P \wedge \sim(Q \vee R))\big) \equiv P \wedge \sim(Q \vee R) \equiv P \wedge \sim Q \wedge \sim R.$$

Thus n is composite if and only if

$$\exists m \in \mathbb{N} : (m \mid n) \wedge (m \neq 1) \wedge (m \neq n).$$

In other words, n is composite if and only if n has a factor not equal to 1 or itself.

4.3.7 Example.

(i) *The integers 2, 3, 5, and 7 are prime.*

(ii) *The integers $4 = 2 \cdot 2$, $6 = 2 \cdot 3$, $8 = 2 \cdot 4 = 2 \cdot 2 \cdot 2$, and $9 = 3 \cdot 3$ are all composite.*

This is where we finished on Monday, September 20, 2021.

We will show that every integer greater than or equal to 2 has a prime factor; this is the first step in proving the result, long familiar to us, that every integer factors completely into a product of primes. To do so, we will need to bring to light a property of $\mathbb{N}$ that we have not yet needed.

4.3.8 Axiom ("Well-ordering" property of $\mathbb{N}$).

If A is a set of natural numbers and there exists[16] *$n \in A$, then A has a* **LEAST ELEMENT***: there exists $\ell \in A$ such that $\ell \leq b$ for all $b \in A$.*

4.3.9 Example.

(i) *Let A be the set of all ages in years of the people who came to class on the day this material was discussed in lecture. If at least one person came to class, then A has a least element, and so there is a person in class whose age is the youngest of all. (It is possible that more than one person is that youngest age, but there is only one youngest age value.)*

(ii) *We recast the well-ordering property using quantifiers: if A is a set of natural natural numbers and there exists $n \in A$, then*

$$\exists \ell \in A \ \forall b \in A : \ell \leq b.$$

4.3.10 Theorem.

Every positive integer greater than or equal to 2 is divisible by a prime number.

Proof Sketch.

- We think about the process of prime factorization and start with an easy number, say, 24. Perhaps we write $24 = 3 \cdot 8$, in which case we see that 3 is a prime factor, or perhaps we start with $24 = 4 \cdot 6$, in which case we break down $4 = 2 \cdot 2$ or $6 = 2 \cdot 3$, which give prime factors.

- Since 24 is rich with factors, we should also play with something simpler: how can we factor 5? Here we are much more restricted, since 5 is prime. But 5 is a factor of itself, and a prime factor at that.

- We think about an easy case: if $p \geq 2$ is prime, then p is divisible by the prime p.

- We think about the harder case: suppose n is not prime (i.e., n is composite). Then n factors into the product of two numbers, neither of which is 1 or n. So we write $n = ab$, where $a \neq 1$, $a \neq n$, $b \neq 1$, and $b \neq n$. If either a or b is prime, then we are done.

- Otherwise, if both a and b are not prime, then we could repeat this argument on both. So, $a = cd$, where $c \neq 1$, $c \neq a$, $d \neq 1$, and $d \neq a$. If either c or d is prime, then we are done, by the transitivity of divisibility. That is, if c is prime, then $c \mid a$, and since $a \mid n$, we have $c \mid n$.

- But what if c and d are both composite? Must we repeat this factorization process forever? Will it ever stop?

[16] By requiring the existence of some $n \in A$, we have guaranteed that A is **NONEMPTY**. The reason for this will become apparent when we agree later on the existence of an **EMPTY SET**, i.e., a set that contains no elements. Obviously we would want to rule out an empty A in the well-ordering property, as an empty set cannot contain any element, let alone a least element!

• We should be consoled by the observation that each time we factor, the factors "get smaller." That is, if $n = ab$ and $a \neq n$ and $b \neq n$, then $1 \leq a < n$ and $1 \leq b < n$. In other words, n should have a smallest factor.

• Now we wonder if this smallest factor is prime. If it is, then we are done; if it is not, then this smallest factor should factor into two more factors! But then it would not be the smallest factor after all, a contradiction.

• By the way, we should be careful about the phrase "smallest factor": of course 1 is always the smallest factor. So we should look for the smallest factor of n that is not 1.

• Finally, we try not to be too ambitious: we are not showing that every integer factors into a product of entirely prime numbers but rather than every integer (greater than or equal to 2) *has* a prime factor.

• We put all these observations together and (maybe with some help from on high) try to invoke the well-ordering principle. We want n to have a smallest factor not equal to 1, so we let A be the set of all integers b such that $b \neq 1$ and $b \mid n$. Then A has a least element ℓ. Since $\ell \in A$, we must have $\ell \mid n$, so if ℓ is prime then we are done.

• We want to rule out the case that ℓ could be composite, and one way to do this is to assume that ℓ *is* composite and then to see what goes wrong. In particular, we should expect that if ℓ is composite then ℓ has a smaller factor which is then a factor of n. But then n has a factor smaller than ℓ, which should give our contradiction.

• Let us make this precise. If ℓ is composite, then there exists c such that $c \mid \ell$, $c \neq 1$, and $m \neq \ell$. The only things available for us to contradict are properties of ℓ and its relation to n. Transitivity of divisibility implies $c \mid n$ and so $c \neq 1$ implies $c \in A$.

• But $c \mid \ell$, and $c \neq \ell$, so we expect $c < \ell$. We need to prove this, and we will. If this is the case, then we have $\ell \leq c < \ell$. That is, $\ell < \ell$, which is impossible, our contradiction.

• Finally, here is the proof of that claim that if $c \mid \ell$ and $c \neq \ell$, then $c < \ell$. Really we have c, $\ell \in \mathbb{N}$, and so $c > 0$ and $\ell > 0$. Since $c \mid \ell$, we have $\ell = cd$ for some $d \in \mathbb{Z}$. If $d \leq 0$, then $\ell \leq 0$, which is impossible. So $d > 0$ and therefore $d \geq 1$, since d is an integer. If $d = 1$, then $\ell = c \cdot 1 = c$, which is impossible. Thus $d > 1$ and so $\ell = cd > c \cdot 1 = c$.

This is where we finished on Wednesday, September 22, 2021.

Before giving the formal proof, we state a lemma, whose proof we gave quite rigorously at the end of the proof sketch above.

4.3.11 Lemma.

Let $c, \ell \in \mathbb{N}$ with $c \mid \ell$ and $c \neq \ell$. Then $c < \ell$.

Now we are ready for the formal proof of Theorem 4.3.10.

Formal Proof. First suppose that n is prime. Then $n \mid n$, so n is its own (and only) prime factor.

Now suppose n is composite, so there exists $m \in \mathbb{N}$ such that $m \mid n$, $m \neq 1$, and $m \neq n$. Let A be the set of all factors of n in $\mathbb{N}$ that are not equal to 1. That is, $b \in A$ if and only if $b \mid n$ and $b \neq 1$. In particular, $m \in A$, so A has a least element, which we denote by ℓ.

We claim that ℓ is a prime factor of n. First, since $\ell \in A$, we have $\ell \mid n$. Next, we show that a contradiction results if ℓ is composite. In this case, we can find $c \in \mathbb{N}$ such that $c \mid \ell$, $c \neq 1$, and $c \neq \ell$. By the transitivity of divisibility, since $c \mid \ell$ and $\ell \mid n$, we have $c \mid n$. Since we know $c \neq 1$, this implies $c \in A$. Thus $\ell \leq c$.

But we also know that $c \mid \ell$ and $c \neq \ell$ with both $c > 0$ and $\ell > 0$. Lemma 4.3.11 then implies $c < \ell$, and so we have the inequalities $c < \ell \leq c$. That is, $c < c$, which is impossible. This is our desired contradiction, and so ℓ cannot be composite and is therefore prime.

Proof Analysis. First suppose that n is prime. [We are giving the proof in two cases: n prime and n composite. If we had more to say here, we might have written out a header for this part of the proof, like "Case 1: n is prime."] Then $n \mid n$, so n is its own (and only) prime factor.

Now suppose n is composite, so there exists $m \in \mathbb{N}$ such that $m \mid n$, $m \neq 1$, and $m \neq n$. [Here is the second part of the proof. The change in focus is signaled by the word "Now."] Let A be the set of all factors of n in $\mathbb{N}$ that are not equal to 1. That is, $b \in A$ if and only if $b \mid n$ and $b \neq 1$. [We describe the set A in words and in symbols — the first description is more evocative, the second is more useful.] In particular, $m \in A$, so A has a least element, which we denote by ℓ.

We claim that ℓ is a prime factor of n. [Here we announce our goal for the rest of the proof: ℓ is the desired prime factor.] First, since $\ell \in A$, we have $\ell \mid n$. [We start with the easy part, proving $\ell \mid n$.] Next, we show that a contradiction results if ℓ is composite. [We provide direction for how the rest of the proof will go.] In this case, we can find $c \in \mathbb{N}$ such that $c \mid \ell$, $c \neq 1$, and $c \neq \ell$. By the transitivity of divisibility, since $c \mid \ell$ and $\ell \mid n$, we have $c \mid n$. Since we know $c \neq 1$, this implies $c \in A$. Thus $\ell \leq c$.

But we also know that $c \mid \ell$ and $c \neq \ell$ with both $c > 0$ and $\ell > 0$. [We start a new paragraph to emphasize the coming contradiction.] Lemma 4.3.11 then implies $c < \ell$, and so we have the inequalities $c < \ell \leq c$. That is, $c < c$, which is impossible. This is our desired contradiction, and so ℓ cannot be composite and is therefore prime. [It has been a long proof, so we remind the reader of where we are, and why we are done.]

4.3.3. An excursion into proofs by cases.

The proof of Theorem 4.3.10 was a "proof by cases." We wanted to show $P \Longrightarrow Q$, where P was the statement "$n \geq 2$ is an integer" and Q was the statement "n has a prime divisor." Let A be the statement "n is prime" and B be the statement "n is composite." Then $P \Longrightarrow (A \vee B)$ is true. Consequently, if we know that $(A \vee B) \Longrightarrow Q$ is true, then $P \Longrightarrow Q$ will be true:

$$\big[P \Longrightarrow \big((A \vee B) \Longrightarrow Q\big)\big] \Longrightarrow (P \Longrightarrow Q).$$

In turn, we have

$$(A \vee B) \Longrightarrow Q \equiv (A \Longrightarrow Q) \wedge (B \Longrightarrow Q).$$

The truth of $A \Longrightarrow Q$ was easy to establish (if n is prime, then n is its own prime divisor), while that of $B \Longrightarrow Q$ was harder. Depending on the situation, it may be advantageous to attempt a "proof by cases" if one can show that the original hypothesis P implies a number of straightforward subcases.

4.3.12 Example.

Here are some useful subcases in number theory.

(i) *If $n \in \mathbb{N}$, then $n = 1$, or $n \geq 2$ and n is prime, or $n \geq 2$ and n is composite.*

(ii) *If $n \in \mathbb{Z}$, then n is odd or n is even. (We still have yet to show that this is true!)*

(iii) *If $n \in \mathbb{Z}$, then $n > 0$, or $n = 0$, or $n < 0$.*

4.3.13 Example.

The product of any two consecutive integers is even.

Proof Sketch.

- We should introduce notation to capture what "consecutive integers" means. Integers n and m are consecutive if one immediately precedes (or succeeds) the other. So either $n = m + 1$ or $m = n + 1$.

- For simplicity, we fix one $n \in \mathbb{Z}$ and study the product $n(n + 1)$. We need to show that $n(n + 1)$ is even; equivalently, we need to prove the statement

$$\forall n \in \mathbb{Z} : 2 \mid n(n + 1).$$

This is where we finished on Friday, September 24, 2021.

- That is, we need to find $k \in \mathbb{Z}$ such that $n(n + 1) = 2k$. Somehow we have to get a factor of 2 out of n or $n + 1$.

• The key idea is to recall that parity alternates (if we believe that every integer is either even or odd, which, annoyingly, we *still* have not rigorously proved). So either n will be even or $n+1$ will be even, and that will give us our factor of 2.

• This suggests that we write our proof in two cases: the case where n is even and the case where n is odd. Those two cases will exhaust all possibilities of parity for an integer.

Formal Proof. Let $n \in \mathbb{Z}$; we show that $n(n+1)$ is even.

Case 1: n is even. Then there is $k \in \mathbb{Z}$ such that $n = 2k$ and so

$$n(n+1) = 2k(2k+1) = 2\big[k(2k+1)\big].$$

Since $k(2k+1) \in \mathbb{Z}$, we conclude that $n(n+1)$ is even.

Case 2: n is odd. Then there is $j \in \mathbb{Z}$ such that $n = 2j+1$, and so

$$n(n+1) = (2j+1)\big((2j+1)+1\big) = (2j+1)(2j+2) = 2\big[(2j+1)(j+1)\big].$$

Since $(2j+1)(j+1) \in \mathbb{Z}$, we see that $n(n+1)$ is again even.

Proof Analysis. Let $n \in \mathbb{Z}$; we show that $n(n+1)$ is even. [We have introduced notation and stated our goal in the language of our chosen notation. This is important, since the original claim just used "words" and no concrete mathematical notation.] We consider the cases of n even and n odd separately. [We tell the reader that this will be a "proof by cases" argument, and we state what those cases are.]

Case 1: n is even. [We have identified our first case, and we distinguish this case from the rest of the proof typographically by (1) starting an new paragraph and (2) introducing the header in italics. Of course, we could just start this paragraph with an actual sentence, like "First suppose n is even," but sometimes a header can make things easier to read.] Then there is $k \in \mathbb{Z}$ such that $n = 2k$ and so

$$n(n+1) = 2k(2k+1) = 2\big[k(2k+1)\big].$$

[This is a delicate calculation, so we give all the arithmetical and algebraic details.] Since $k(2k+1) \in \mathbb{Z}$, we conclude that $n(n+1)$ is even. [We signal that we have done enough work to conclude the desired result in this case.]

Case 2: n is odd. [Another case, another header.] Then there is $j \in \mathbb{Z}$ such that $n = 2j+1$, and so

$$n(n+1) = (2j+1)\big((2j+1)+1\big) = (2j+1)(2j+2) = 2\big[(2j+1)(j+1)\big].$$

[Another gory calculation!] Since $(2j+1)(j+1) \in \mathbb{Z}$, we see that $n(n+1)$ is again even.

4.3.4. The division algorithm.

If $a, b \in \mathbb{Z}$ and it is not the case that $b \mid a$, then we write $b \nmid a$. Recall that

$$b \mid a \equiv \exists c \in \mathbb{Z} : a = bc,$$

and so

$$b \nmid a \equiv \sim(\exists c \in \mathbb{Z} : a = bc) \equiv \forall c \in \mathbb{Z} : a \neq bc.$$

4.3.14 Example.

(i) $5 \nmid 48$ *but* $24 = 9(5) + 3$.

(ii) $7 \nmid 48$ *but* $48 = 6(7) + 6$.

(iii) $18 \nmid 48$ *but* $48 = 2(18) + 12$.

(iv) $30 \nmid 48$ *but* $48 = 30 + 18 = 1(30) + 18$.

In each case we started with a number n and tried to divide n by some d. We could not divide perfectly, because we got a quotient q and a remainder term r: $n = qd + r$. If we look carefully, we see that $r < d$ each time. This aligns with our intuition about division "with remainder," and here is the formal statement of that result.

4.3.15 Theorem (Division algorithm).

Let $n, d \in \mathbb{N}$. Then there exist $q, r \in \mathbb{Z}$ such that

$$n = qd + r.$$

Moreover, $q \geq 0$ and $0 \leq r < d$.

We will not prove this theorem, but we do take the time to cast it in symbolic language:

$$\forall n, d \in \mathbb{N}\ \exists q, r \in \mathbb{Z} : (n = qd + r) \wedge (0 \leq r < d) \wedge (q \geq 0).$$

For example, in dividing 48 by 5, we found $48 = 9(5) + 3$, so here $n = 48$, $d = 5$, $q = 9$, and $r = 3$.

Why is it important to specify $r < d$? Suppose $r \geq d$. Then $r = (r - d) + d$, where $r - d \geq 0$ and thus $r - d \in \mathbb{N}$. Then

$$n = qd + r = qd + d + (r - d) = (q + 1)d + (r - d),$$

and so this quotient q is not "optimal" — the remainder is not as small as possible. For example, we would not accept $48 = 8(5) + 8$ as the appropriate quotient-remainder form when dividing 48 by 5.

With the division algorithm we can give a very simple proof that every integer is either even or odd. (A rather harder proof can be accomplished using the well-ordering principle!)

4.3.16 Corollary.

Every integer is even or odd. That is, every integer has a parity.

Proof. We give the proof only for natural numbers and leave it as an exercise to establish the consequent parity of all elements of $\mathbb{Z}$. So, let $n \in \mathbb{N}$. Apply the division algorithm with $d = 2$ to see that $n = 2q + r$ for some q, $r \in \mathbb{N}$, where $0 \leq r < d$. That is, $0 \leq r < 2$. Since r is an integer, this forces $r = 0$ or $r = 1$. In the case $r = 0$ we have $n = 2q$, so n is even, while in the case $r = 1$, we have $n = 2q + 1$, so then n is odd. ■

5. Elementary Set Theory

5.1. Fundamentals.

We begin by (un)formalizing some concepts that we have used throughout our discussions of predicates and number theory.

5.1.1 Undefinition.

A **set** *is a group of collection of objects, called the* **elements** *of the set.*

If A is a set and x is an element of A, we write $x \in A$. We read the statement $x \in A$ as "x is an element of A" or "x is a member of A" or "x belongs to A." Do not confuse the membership symbol $\in$ with two common ways of writing the Greek letter epsilon, ϵ and ε. If x is not an element[17] of A, we write $x \notin A$. That is,

$$x \notin A \equiv \sim(x \in A).$$

We often denote a set by listing all of its elements between left and right curly braces ({ and }). For example, the set consisting of the numbers 1 and 2 is

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

We read the statement "$A = \{1, 2\}$" as "A is the set consisting of 1 and 2" or "A is the set whose elements are 1 and 2." We should avoid saying something like "A is the set *containing* 1 and 2," as $\{1, 2, 3\}$, $\mathbb{N}$, and $\mathbb{Z}$ are also sets containing 1 and 2.

When we define a set by listing its elements within curly braces, we adopt the convention that the order of the elements and repetition of the elements do not matter. Thus

$$\{1, 2\} = \{2, 1\} = \{1, 1, 2\}.$$

Later we will define mechanisms for expressing situations where order and/or repetition *do* matter.

We must be very careful to distinguish an element of a set from the set consisting precisely of that element. For example, 1 is a number, but $\{1\}$ is a set, specifically the set whose only element is the number 1. We have $1 \in \{1\}$ but $1 \neq \{1\}$. How can a number be the same as the set consisting of that number? More generally, we adopt the convention[18] that for any element x,

$$x \neq \{x\}.$$

[17] If x is not an element of A, then what is x? Surely x is *something*. Most "practical" applications of set theory operate with some universal set in mind — $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$ are popular for obvious reasons. The problem is that there is no universal set that contains everything; we will show in Section **??** that the existence of such a set contradicts a fundamental, plausible, and highly useful axiom that we shall shortly adopt. So, when we say "x is an element," we really mean that x is an element of another set, a set that we could mention if we really had to, but that we typically choose not to.

[18] This "convention" is really an axiom — and that axiom is really two axioms: if A is a set and $x \in A$, then (1) there exists a set $\{x\}$ whose unique element is x (a single-variable version of the **axiom of pairing**) and (2) this set $\{x\}$ and its sole element x are not the same, i.e., $x \neq \{x\}$ (the **axiom of regularity**).

This is where we finished on Monday, September 27, 2021.

5.1.1. The axiom of separation.

The disadvantage of the "curly braces" way of defining sets is that, typographically, it quickly becomes cumbersome. It can be much cleaner to define a set by some common property that all of its elements share — the "set of all people who came to class today" is a much slicker phrasing than enumerating all 30+ names.

Let $P(x)$ be a predicate with domain D. We denote by

$$\{x \in D \mid P(x)\}$$

the set of all[19] x in D for which it is the case that $P(x)$. (If this sounds weird, remember that $P(x)$ is a sentence!) Some authors use : instead of our |, i.e.,

$$\{x \in D \mid P(x)\} = \{x \in D : P(x)\}.$$

5.1.2 Example.

(i) *Let $D = \mathbb{R}$ and let $P(x)$ be the predicate "$x^2 = 4$." Then*

$$\{x \in D \mid P(x)\} = \{x \in \mathbb{R} \mid x^2 = 4\} = \{2, -2\} = \{-2, 2\}.$$

(ii) *Let $D = \mathbb{R}$ and let $P(x)$ be the predicate "$0 \leq x$ and $x \leq 1$." Then*

$$\{x \in D \mid P(x)\} = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\} = [0, 1].$$

(iii) *Let D be a set, let $P(x)$ be a predicate with domain D, and let $A = \{x \in D \mid P(x)\}$. Then if x is an element and $x \notin A$, then either $x \notin D$ or $\sim P(x)$.*

It is worthwhile questioning *why* we should be able to form this set. After all, when presented with a description of an object with certain properties, one of the most fundamental questions of mathematics is *Does that object exist?* We take the quick, but perhaps unsatisfying, route of assuming axiomatically that given a predicate $P(x)$ with domain D, we can always form the set $\{x \in D \mid P(x)\}$. We call this assumption the **AXIOM OF SEPARATION**, in the sense that we are "separating out" those elements x of D for which it is the case that $P(x)$.

For definiteness, we record here what we hope are familiar definitions of subintervals of $\mathbb{R}$. These intervals can make excellent, illustrative examples of the various abstract properties of sets that will follow. Recall that we abbreviate by $a \leq x \leq b$ the statement $(a \leq x) \wedge (x \leq b)$ and similarly for $<$ in place of $\leq$.

[19] When writing a set in this "set-builder notation," the use of x is flexible. We should think of x as a "dummy variable"; thus

$$\{x \in D \mid P(x)\} = \{t \in D \mid P(t)\} = \{\alpha \in D \mid P(\alpha)\}.$$

However, if we have specified an *element* $x \in D$, then we should not confuse x with some other $t \in D$ unless we are sure that x and t are the same (which begs the interesting question of how to define "the same," a question that we punt to Footnote 26).

5.1.3 Example.

Let $a, b \in \mathbb{R}$ with $a \leq b$. A **SUBINTERVAL** *of $\mathbb{R}$, or, more plainly, an* **INTERVAL** *in $\mathbb{R}$, is a set that has one of the following forms.*

(i) $[a, b] := \{x \in \mathbb{R} \mid a \leq x \leq b\}$. *We say that this is a* **CLOSED** *and* **BOUNDED** *interval.*

(ii) $[a, b) := \{x \in \mathbb{R} \mid a \leq x < b\}$.

(iii) $(a, b) := \{x \in \mathbb{R} \mid a < x < b\}$. *We say that this is a* **OPEN** *interval.*

(iv) $(a, b] := \{x \in \mathbb{R} \mid a < x \leq b\}$.

(v) $(-\infty, b) := \{x \in \mathbb{R} \mid x < b\}$.

(vi) $(-\infty, b] := \{x \in \mathbb{R} \mid x \leq b\}$.

(vii) $(a, \infty) := \{x \in \mathbb{R} \mid x > a\}$.

(viii) $[a, \infty) := \{x \in \mathbb{R} \mid x \geq a\}$.

(ix) $(-\infty, \infty) := \mathbb{R}$.

5.1.2. Subsets.

Often it is the case that one set under consideration is "contained" in another, "larger" set. For example, every natural number is an integer, and every integer is a real number.

5.1.4 Definition.

A set A is a **SUBSET** *of the set B if for each $x \in A$ it is the case that $x \in B$. We write $A \subseteq B$, read "A is a subset of B."*

Thus

$$A \subseteq B \iff (\forall x \in A : x \in B).$$

In turn, $A \subseteq B$ if and only if the statement[20]

$$x \in A \Longrightarrow x \in B$$

is always true.

If A is not a subset of B, we write $A \not\subseteq B$; thus

$$A \not\subseteq B \iff \sim(A \subseteq B) \iff \sim(\forall x \in A : x \in B) \iff \exists x \in A : x \notin B.$$

[20] This is really a predicate in the variable x. For a given x, it is false if $x \in A$ and $x \notin B$ and true otherwise. What is the domain of this predicate? A? B? All elements of A and B? Some yet larger set to which all the elements of A and B belong?

5.1.5 Example.

(i) $\{1\} \subseteq \{1,2\}$.

(ii) $\{1,2\} \not\subseteq \{1,3\}$.

(iii) $\mathbb{N} \subseteq \mathbb{Z}$.

(iv) $\mathbb{Z} \subseteq \mathbb{R}$.

(v) $\{n \in \mathbb{Z} \mid 4 \mid n\} \subseteq \{n \in \mathbb{Z} \mid 2 \mid n\}$.

Proof. All of the containment (or, in the case of (ii), noncontainment) claims should be obvious by inspection, except perhaps part (v). For that part, let

$$A = \{n \in \mathbb{Z} \mid 4 \mid n\} \quad \text{and} \quad B = \{n \in \mathbb{Z} \mid 2 \mid n\}.$$

We need to show that if $n \in A$, then $n \in B$. So, suppose $n \in A$. Then $n \in \mathbb{Z}$ and $4 \mid n$, so there is $m \in \mathbb{Z}$ such that $n = 4m = 2(2m)$. Since $2m \in \mathbb{Z}$, we conclude $2 \mid n$, and thus $n \in B$. ■

If A, B, and C are sets with $A \subseteq B$ and $B \subseteq C$, then we write

$$(A \subseteq B) \wedge (B \subseteq C) \equiv A \subseteq B \subseteq C,$$

much as we might write $1 \leq 2 \leq 3$. The $\subseteq$ condition is "transitive" in the following sense.

5.1.6 Theorem.

Let A, B, and C be sets with $A \subseteq B$ and $B \subseteq C$. Then $A \subseteq C$.

Proof. Here is a proof in words. Let $x \in A$. Then since $A \subseteq B$, we have $x \in B$. Then since $B \subseteq C$, we have $x \in C$.

Here is a proof in symbols:

$$(x \in A) \wedge (A \subseteq B) \Longrightarrow x \in B$$
$$(x \in B) \wedge (B \subseteq C) \Longrightarrow x \in C.$$ ■

Sometimes it is the case that $A \subseteq B$ but B contains some elements that are not elements of A. In this case we say that A is a **PROPER** subset of B and write $A \subset B$ or $A \subsetneq B$. That is,

$$A \subset B \equiv A \subsetneq B \equiv (A \subseteq B) \wedge (\exists x \in B : x \notin A).$$

We should be careful to distinguish the symbols $\not\subseteq$ and $\subsetneq$. We also point out that some authors do not use $\subset$ to denote a *proper* subset but instead in lieu of our $\subseteq$. (If $\subseteq$ is meant to resemble $\leq$ when comparing real numbers, then $\subset$ should remind us of $<$. Outside of these remarks, it is frankly unlikely that we will ever use the notation $\subset$.)

5.1.7 Example.

(i) $\{1\} \subset \{1, 2\}$.

(ii) $\mathbb{N} \subset \mathbb{Z}$.

(iii) $\mathbb{Z} \subsetneq \mathbb{R}$.

This is where we finished on Wednesday, September 29, 2021.

5.1.8 Theorem.

Let A be a set. Then $A \subseteq A$.

Proof. We need to show that if $x \in A$, then $x \in A$. This is obviously true: it is a tautology of the form $P \implies P$. ■

5.1.3. Set equality.

Two sets should be equal if they contain precisely the same elements.

5.1.9 Definition.

Let A and B be sets. Then $A = B$ if for all $x \in A$, it is the case that $x \in B$, and for all $x \in B$, it is the case that $x \in A$.

We (should) immediately see that $A = B$ if and only if the statement

$$x \in A \iff x \in B \tag{5.1.1}$$

is true and that

$$A = B \iff \big[(A \subseteq B) \wedge (B \subseteq A)\big], \tag{5.1.2}$$

Proving the separate "containments" $A \subseteq B$ and $B \subseteq A$ is usually exactly how we prove set equality.

Note that we will be making the symbol $=$ do a lot of work: $2^2 = 4$ and $\{4\} = \{4, 4\}$. Thus it is possible to speak of equality of elements of a set *and* equality of sets.

5.1.10 Example.

(i) $\big\{x \in \mathbb{R} \;\big|\; x^2 - 2x + 1 = 0\big\} = \{1\}$.

(ii) $\big\{1, \{1\}\big\} \neq \{1\}$.

(iii) $\{x \in \mathbb{R} \mid x \geq 0\} \neq \{x \in \mathbb{N} \mid x \geq 0\}$.

Proof. (i) First suppose $x \in \big\{x \in \mathbb{R} \;\big|\; x^2 - 2x + 1 = 0\big\}$. Then it is the case that $x^2 - 2x + 1 = 0$. We factor the left side of this equality to find $(x - 1)^2 = 0$, which forces $x = 1$. Thus $x \in \{1\}$.

Now suppose $x \in \{1\}$. Then $x = 1$, so $x \in \mathbb{R}$, and we check that $x^2 - 2x + 1 = 1^2 - 2(1) + 1 = 0$. Thus $x \in \{x \in \mathbb{R} \mid x^2 - 2x + 1 = 0\}$.

(ii) We do have $\{1\} \subseteq \{1, \{1\}\}$: if $x \in \{1\}$, then $x = 1$, and $1 \in \{1, \{1\}\}$. But $\{1\} \in \{1, \{1\}\}$ and $\{1\} \notin \{1\}$, although $\{1\} = \{1\}$.

(iii) We do have $\{x \in \mathbb{N} \mid x \geq 0\} \subseteq \mathbb{N} \subseteq \mathbb{R}$. But $1/2 \in \{x \in \mathbb{R} \mid x \geq 0\}$ and $1/2 \notin \mathbb{N}$, thus $1/2 \notin \{x \in \mathbb{N} \mid x \geq 0\}$. ■

5.1.4. The empty set.

Experience suggests that a set containing no elements at all should exist. For example, there are no real numbers x such that $x^2 < 0$, and so the set

$$A := \{x \in \mathbb{R} \mid x^2 < 0\}$$

should not contain anything. Likewise, the set

$$B := \{n \in \mathbb{N} \mid 2 \mid n \text{ and } 2 \nmid n\}$$

should not contain anything, either. Nonetheless, the axiom of separation ensures that both A and B are defined.

Even though A and B were defined via different predicates with different domains, remarkably they must be the same set! One verification of this is a fascinating exercise in submitting to our reality that the statement $P \implies Q$ is true whenever P is false, regardless of the truth value of Q. Another version of the proof relies on the equivalence of $P \implies Q$ and its contrapositive $\sim Q \implies \sim P$, as well as the fact that a statement of the form $R \implies S$ is true whenever S is true, regardless of the truth of R.

5.1.11 Theorem.

Suppose that A and B are sets with the property that for all elements x, it is the case that $x \notin A$ and $x \notin B$. Then $A = B$.

Proof. It suffices to show $A \subseteq B$ and $B \subseteq A$. We give three proofs that $A \subseteq B$; the proofs that $B \subseteq A$ will proceed in the same way, just swapping the roles of A and B.

(i) First we give a direct proof. We need to show that the statement "$x \in A \implies x \in B$" is always true. But the statement $x \in A$ is always false, and so the preceding if-then statement has a false hypothesis and therefore is true.

(ii) Now we give an indirect proof that does not rely on fundamental properties of $\implies$. Suppose that $A \not\subseteq B$. Then there exists $x \in A$ such that $x \notin B$. But the condition $x \in A$ contradicts our assumption that $x \notin A$, and so it must be the case that $A \subseteq B$.

(iii) Finally, we prove the contrapositive of the statement "$x \in A \implies x \in B$." The contrapositive is $x \notin B \implies x \notin A$. The statement $x \notin A$ is always true, and so the contrapositive is always true. (It happens to be the case that $x \notin B$ is always true, too.) ■

So, there exists only one set S with the property that $x \notin S$ for all elements x. We denote this **EMPTY SET** by $\varnothing$:

$$\forall x : x \notin \varnothing.$$

(Some texts use $\emptyset$ instead of $\varnothing$.)

5.1.12 Theorem.

Let A be a set. Then $\varnothing \subseteq A$.

Proof. We give three proofs in the style of the proof of Theorem 5.1.11.

(i) First we give a direct proof. We need to show that the statement $x \in \varnothing \implies x \in A$ is always true. But the statement $x \in \varnothing$ is always false, and so the preceding if-then statement has a false hypothesis and is therefore true.

(ii) Now we give an indirect proof. Suppose that $\varnothing \not\subseteq A$. Then there exists $x \in \varnothing$ such that $x \notin A$. But it cannot be the case that there exists $x \in \varnothing$, so we have reached a contradiction.

(iii) Finally we prove the contrapositive. We need to show that the statement $x \notin A \implies x \notin \varnothing$ is always true. The conclusion of this if-then statement is $x \notin \varnothing$, and that is always true; thus the contrapositive is true. (The hypothesis $x \notin A$ may or may not be true, but that is irrelevant here.) ■

5.1.13 Remark.

Now that we are convinced of the existence of the empty set, which contains nothing, should we try to establish the existence of a universal set $\mathcal{U}$, which contains everything? (This would lead, among other things, to the squicky realization $\mathcal{U} \in \mathcal{U}$, so perhaps this is not a good idea.) It turns out, at least according to the axiomatic model of set theory that we are more or less implicitly following, that the existence of a universal set is incompatible with our beloved, if subtle, axiom of separation; see the discussion of Russell's paradox in Appendix B. And so we do not start down this dark path of universality.

5.1.5. The power set.

5.1.14 Definition.

Let A be a set. The **POWER SET** *of A is the set of all subsets of A. We denote the power set of A by $\mathscr{P}(A)$.*

5.1.15 Theorem.

Let A be a set.

(i) $\varnothing \in \mathscr{P}(A)$.

(ii) $A \in \mathscr{P}(A)$.

(iii) $\{\varnothing, A\} \subseteq \mathscr{P}(A)$.

(iv) $\mathscr{P}(A) \neq \varnothing$.

Proof. **(i)** Theorem 5.1.12 tells us that $\varnothing \subseteq A$ for any set A, and so $\varnothing \in \mathscr{P}(A)$.

(ii) Theorem 5.1.8 tells us that $A \subseteq A$ for any set A, and so $A \in \mathscr{P}(A)$.

(iii) Parts (i) and (ii) above tell us that $\varnothing \in \mathscr{P}(A)$ and $A \in \mathscr{P}(A)$, and so $\{\varnothing, A\} \subseteq \mathscr{P}(A)$.

(iv) Part (i) above tells us that $\varnothing \in \mathscr{P}(A)$, so $\mathscr{P}(A)$ contains at least one element and therefore cannot be empty. ■

This is where we finished on Friday, October 1, 2021.

5.1.16 Example.

Determine all the elements of $\mathscr{P}(\{1\})$.

Solution. Theorem 5.1.12 tells us that $\varnothing \subseteq \{1\}$, while Theorem 5.1.8 tells us that $\{1\} \subseteq \{1\}$. Thus

$$\big\{\varnothing, \{1\}\big\} \subseteq \mathscr{P}(\{1\}),$$

and it appears that this is actually a set equality.

Let us check to see if we have missed any elements of $\mathscr{P}(\{1\})$. Let $A \in \mathscr{P}(\{1\})$ and suppose $A \neq \varnothing$ and $A \neq \{1\}$. Since $A \in \mathscr{P}(\{1\})$, we have $A \subseteq \{1\}$. We know that $A \neq \varnothing$, so there exists $x \in A$; then $x \in \{1\}$. Thus $x = 1$, and so $1 \in A$. Hence $\{1\} \subseteq A$. Since we already had $A \subseteq \{1\}$, we conclude $A = \{1\}$. ▲

5.1.17 Lemma.

Let A be a set such that $A \subseteq \varnothing$. Then $A = \varnothing$.

Proof. Theorem 5.1.12 tells us that $\varnothing \subseteq A$. We are assuming that $A \subseteq \varnothing$, and so by the "subset characterization" of set equality in (5.1.2) we must have $A = \varnothing$. ■

5.1.18 Example.

Show that $\mathscr{P}(\varnothing) = \{\varnothing\}$.

Solution. Let $A \in \mathscr{P}(\varnothing)$. Then $A \subseteq \varnothing$. Lemma 5.1.17 tells us that $A = \varnothing$, and so $A \in \{\varnothing\}$. Hence $\mathscr{P}(\varnothing) \subseteq \{\varnothing\}$.

Conversely, let $A \in \{\varnothing\}$. Then $A = \varnothing$, and Theorem 5.1.15 tells us that $\varnothing \in \mathscr{P}(\varnothing)$. Hence $\{\varnothing\} \subseteq \mathscr{P}(\varnothing)$. ▲

5.1.19 Remark.

If A "has n elements" for some $n \in \mathbb{N}$, then $\mathscr{P}(A)$ "has 2^n elements." We will state this precisely and prove it later. For example, $\{1\}$ has 1 element, and $\mathscr{P}(\{1\}) = \{\varnothing, \{1\}\}$ has $2 = 2^1$ elements; $\varnothing$ has 0 elements, and $\mathscr{P}(\varnothing) = \{\varnothing\}$ has $1 = 2^0$ element. For this reason an occasional alternative notation for $\mathscr{P}(A)$ is 2^A.

5.2. Set operations.

There are several "algebraic" operations that we can perform on groups of sets to construct new sets from old ones. These operations have some "moral" similarities with the properties of logical connectives discussed in Section 2.2.2 and indeed are often proved by recourse to those properties.

5.2.1. Unions.

Unions allow us to "put two sets together and obtain a (maybe larger) set."

5.2.1 Definition.

Let A and B be sets. The UNION *of A and B is the set $A \cup B$ consisting of all elements that belong to either A or B:*

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

We immediately reinterpret the definition of $A \cup B$ as

$$x \in A \cup B \iff (x \in A) \vee (x \in B).$$

That is, $A \cup B$ if and only if the statement above is true. We might pronounce the symbol $A \cup B$ as "A union B."

Note that by the equivalence $P \vee Q \equiv Q \vee P$, we have

$$(x \in A) \vee (x \in B) \iff (x \in B) \vee (x \in A),$$

and thus $A \cup B = B \cup A$.

5.2.2 Example.

Verify each purported equality.

(i) $\{1\} \cup \{2\} = \{1, 2\}$.

(ii) $\{1, 2, 3\} \cup \{3, 4\} = \{1, 2, 3, 4\}$.

(iii) $\mathbb{N} \cup \mathbb{Z} = \mathbb{Z}$.

(iv) $[0, 1] = [0, 1) \cup \{1\}$.

Solution. **(i)** This is a direct calculation.

(ii) This is a direct calculation, observing that 3 is an element in both sets and need not be repeated more than once in the final union.

(iii) *Proof in words.* If $n \in \mathbb{Z}$ then certainly $n \in \mathbb{N} \cup \mathbb{Z}$ by definition of $\cup$, so $\mathbb{Z} \subseteq \mathbb{N} \cup \mathbb{Z}$; conversely, if $n \in \mathbb{N} \cup \mathbb{Z}$, then either $n \in \mathbb{N}$ or $n \in \mathbb{Z}$. If $n \in \mathbb{N}$, then since $\mathbb{N} \subseteq \mathbb{Z}$ we have $n \in \mathbb{Z}$, while if $n \in \mathbb{Z}$ there is nothing to prove. Thus $\mathbb{N} \cup \mathbb{Z} \subseteq \mathbb{Z}$.

Proof in symbols[21]. We have

$$\begin{aligned} n \in \mathbb{N} \cup \mathbb{Z} &\implies (n \in \mathbb{N}) \vee (n \in \mathbb{Z}) \\ &\implies (n \in \mathbb{Z}) \vee (n \in \mathbb{Z}) \text{ since } \mathbb{N} \subseteq \mathbb{Z} \\ &\implies n \in \mathbb{Z} \text{ since } P \vee P \equiv P. \end{aligned}$$

Thus $\mathbb{N} \cup \mathbb{Z} \subseteq \mathbb{Z}$. Next,

$$n \in \mathbb{Z} \implies (n \in \mathbb{N}) \vee (n \in \mathbb{Z})$$

is always true, since $P \implies (Q \vee P)$ is a tautology (exercise: check this). Thus $\mathbb{Z} \subseteq \mathbb{N} \cup \mathbb{Z}$.

(iv) *Proof in symbols.* We have

$$\begin{aligned} x \in [0,1] &\iff 0 \leq x \leq 1 \\ &\iff 0 \leq x \wedge x \leq 1 \\ &\iff 0 \leq x \wedge \big((x < 1) \vee (x = 1)\big) \\ &\iff \big((0 \leq x) \wedge (x < 1)\big) \vee \big((0 \leq x) \wedge (x = 1)\big) \\ &\iff (0 \leq x < 1) \vee (x = 1) \\ &\iff \big(x \in [0,1)\big) \vee \big(x \in \{1\}\big) \\ &\iff x \in [0,1) \cup \{1\}. \end{aligned}$$

Thus $[0,1] = [0,1) \cup \{1\}$ using the definition (5.1.1) of set equality.
Proof in words. First suppose $x \in [0,1]$. Then $0 \leq x \leq 1$; in particular, either $x < 1$ or $x = 1$. If $x < 1$, then $0 \leq x < 1$, and so $x \in [0,1)$, hence $x \in [0,1) \cup \{1\}$. If $x = 1$, then certainly $x \in [0,1) \cup \{1\}$. We conclude $[0,1] \subseteq [0,1) \cup \{1\}$.

Now suppose $x \in [0,1) \cup \{1\}$. Then either $x \in [0,1)$ or $x \in \{1\}$. If $x \in [0,1)$, then $0 \leq x < 1$, so $0 \leq x \leq 1$, and therefore $x \in [0,1]$. If $x \in \{1\}$, then $x = 1$, and so, again, $x \in [0,1]$. We conclude $[0,1) \cup \{1\} \subseteq [0,1]$. ▲

This is where we finished on Monday, October 4, 2021.

[21] Proofs like this, where it is so clear and straightforward to operate at a purely symbolic level, are one of the rare times that it is socially and rhetorically acceptable to mix logical connectives directly into an argument.

5.2.3 Remark.

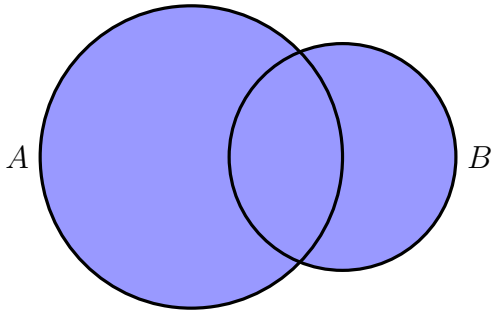
The observant reader will note that in our definition of union as

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\},$$

we omitted "on the left" a specification of a set to which x belongs. Usually when we have defined sets via predicates and the axiom of separation, we have written something like $x \in D$ for some set D. Implicitly we are thinking that A and B are both subsets of a larger "universal" (though not universal in the sense of Russell's paradox) set, and so x would belong to that universal set.

5.2.4 Example.

A **Venn diagram** *can provide a useful illustration of many algebraic operations on sets. Here is a Venn diagram for the union of sets A and B.*



$A \cup B$

While a Venn diagram is not by itself a proof of a particular identity for an algebraic operation, it can nonetheless suggest that the identity is true in the first place. And expecting the truth of a statement is the first step toward proving it.

5.2.5 Theorem.

Let A and B be sets.

(i) $A \subseteq A \cup B$.

(ii) $A \cup \varnothing = A$.

(iii) *If B is a set and $A \subseteq B$, then $A \cup B = B$.*

Proof. **(i)** Let $x \in A$. Then it is the case that $x \in A$ or $x \in B$, so $x \in A \cup B$. In symbols,

$$x \in A \Longrightarrow (x \in A) \vee (x \in B).$$

The truth of this if-then statement is guaranteed by the tautology $P \Longrightarrow (P \vee Q)$.

(ii) Let $x \in A \cup \varnothing$. Then either $x \in A$ or $x \in \varnothing$. We know that $x \notin \varnothing$, so it must be the case that $x \in A$. Thus $A \cup \varnothing \subseteq A$.

Conversely, part (i) ensures $A \subseteq A \cup \varnothing$.

(iii) Let $x \in A \cup B$. Then either $x \in A$ or $x \in B$. If $x \in A$, then since $A \subseteq B$ we have $x \in B$. Thus $A \cup B \subseteq B$.

Conversely, part (i) ensures $B \subseteq A \cup B$. ■

5.2.2. Intersections.

Intersections enable us to see what elements two sets have in common.

5.2.6 Definition.

Let A and B be sets. The **INTERSECTION** *of A and B is the set of all elements that belong to both A and B:*

$$A \cap B := \{x \mid x \in A \text{ and } x \in B\}.$$

We immediately write

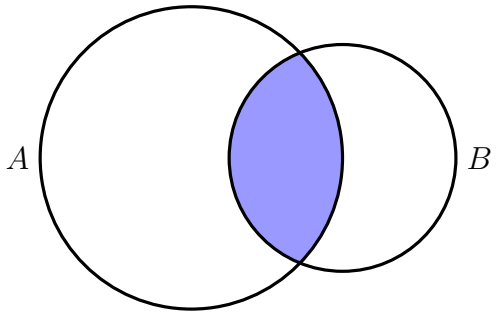
$$x \in A \cap B \iff (x \in A) \wedge (x \in B).$$

By the equivalence $P \wedge Q \equiv Q \wedge P$, we have

$$x \in A \cap B \iff (x \in A) \wedge (x \in B) \iff (x \in B) \wedge (x \in A) \iff x \in B \cap A,$$

and thus $A \cap B = B \cap A$. We might pronounce the symbol $A \cap B$ as "A intersect B."

Here is a Venn diagram for the intersection of A and B.



$A \cap B$

5.2.7 Example.

Verify each purported equality.

(i) $\{1, 2, 3\} \cap \{3, 4\} = \{3\}$.

(ii) $\{1, 2\} \cap \{3\} = \varnothing$.

(iii) $\mathbb{N} \cap \mathbb{Z} = \mathbb{N}$.

(iv) $[1, 2] = [0, 2] \cap [1, 3]$.

Solution. **(i)** The only element in both $\{1,2,3\}$ and $\{3,4\}$ is 3.

(ii) The sets $\{1,2\}$ and $\{3\}$ have no elements in common.

(iii) Let $n \in \mathbb{N} \cap \mathbb{Z}$. Then $n \in \mathbb{N}$ and $n \in \mathbb{Z}$, so $\mathbb{N} \cap \mathbb{Z} \subseteq \mathbb{N}$. Conversely, let $n \in \mathbb{N}$. Then since $\mathbb{N} \subseteq \mathbb{Z}$ we also have $n \in \mathbb{Z}$, and so $n \in \mathbb{N} \cap \mathbb{Z}$. Thus $\mathbb{N} \subseteq \mathbb{N} \cap \mathbb{Z}$.

(iv) It may help to draw a picture.

Full proof in words. First let $x \in [1,2]$. Then $1 \leq x \leq 2$. Since $0 \leq 1$, we have $0 \leq x$, and so $0 \leq x \leq 2$. Thus $x \in [0,2]$. Next, since $2 \leq 3$, we have $x \leq 3$, and so $1 \leq x \leq 3$. Thus $x \in [1,3]$. Hence $x \in [0,2] \cap [1,3]$, and so $[1,2] \subseteq [0,2] \cap [1,3]$.

Now let $x \in [0,2] \cap [1,3]$. Then $x \in [0,2]$ and $x \in [1,3]$. Hence $0 \leq x \leq 2$ and $1 \leq x \leq 3$. In particular, $x \leq 2$ and $1 \leq x$, and so $1 \leq x \leq 2$. Thus $x \in [1,2]$. We conclude $[0,2] \cap [1,3] \subseteq [1,2]$.

Proof of $(\supseteq)$ *in symbols.* Here is a proof of the inclusion $[0,2] \cap [1,3] \subseteq [1,2]$ in symbols and almost no words:

$$\begin{aligned}
x \in [0,2] \cap [1,3] &\iff \big(x \in [0,2]\big) \wedge \big(x \in [1,3]\big) \\
&\iff (0 \leq x \leq 2) \wedge (1 \leq x \leq 3) \\
&\iff \big((0 \leq x) \wedge (x \leq 2)\big) \wedge \big((1 \leq x) \wedge (x \leq 3)\big) \\
&\iff (0 \leq x) \wedge (x \leq 2) \wedge (1 \leq x) \wedge (x \leq 3) \\
&\implies (1 \leq x) \wedge (x \leq 2) \\
&\iff 1 \leq x \leq 2 \\
&\iff x \in [1,2].
\end{aligned}$$

Here we used the tautology $(P \wedge Q \wedge R \wedge S) \implies (R \wedge Q)$. ▲

5.2.8 Theorem.

Let A and B be sets.

(i) $A \cap B \subseteq A$.

(ii) *If $A \subseteq B$, then $A \cap B = A$.*

(iii) $A \cap \varnothing = \varnothing$.

Proof. **(i)** Let $x \in A \cap B$. Then $x \in A$ and $x \in B$. In particular, $x \in A$. (Here we are using the tautology $(P \wedge Q) \implies P$ to conclude $[(x \in A) \wedge (x \in B)] \implies (x \in A)$.)

(ii) Part (i) tells us that $A \cap B \subseteq A$, so we just need to show $A \subseteq A \cap B$. Let $x \in A$. Since $A \subseteq B$, we also have $x \in B$. Thus $x \in A \cap B$.

This is where we finished on Wednesday, October 6 (Section 54).

(iii) We prove this by contradiction. Suppose there exists $x \in A \cap \varnothing$. Then $x \in A$ and $x \in \varnothing$. But the result $x \in \varnothing$ is impossible, so we were wrong to assume the existence of $x \in A \cap \varnothing$. By the uniqueness of the empty set (Theorem 5.1.11), we have $A \cap \varnothing = \varnothing$. ■

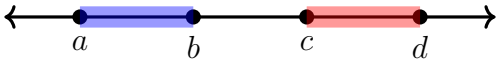
5.2.9 Definition.

Sets A and B are **DISJOINT** *if $A \cap B = \varnothing$.*

5.2.10 Example.

Let a, b, c, $d \in \mathbb{R}$ such that $a \leq b < c \leq d$. Show that $[a, b]$ and $[c, d]$ are disjoint.

Solution. Here is a picture to help illustrate the situation.



We need to show $[a, b] \cap [c, d] = \varnothing$, and we do this by contradiction. Assume there exists $x \in [a, b] \cap [c, d]$. Then $x \in [a, b]$ and $x \in [c, d]$, so $a \leq x \leq b$ and $c \leq x \leq d$. But this means $x \leq b < c \leq x$, so $x < x$, a contradiction. That is, $[a, b] \cap [c, d]$ cannot contain any elements, and so $[a, b] \cap [c, d] \subseteq \varnothing$. ▲

This is where we finished on Wednesday, October 6 (Section 53).

5.2.3. Set-theoretic differences.

The set-theoretic difference[22] gives us a mechanism for excluding elements of one set from another set.

5.2.11 Definition.

Let A and B be sets. The **SET-THEORETIC DIFFERENCE OF A AND B** *or the* **COMPLEMENT OF B IN A,** *is the set of all elements of A that are not elements of B. It is denoted*

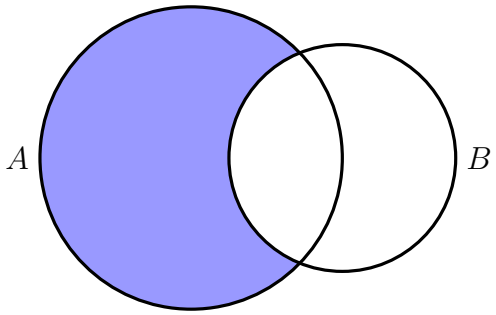
$$A - B = A \setminus B = \{x \in A \mid x \notin B\}.$$

Thus

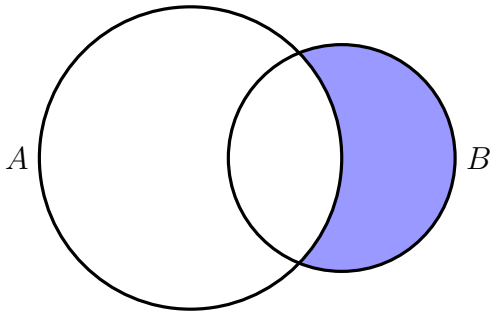
$$x \in A - B \iff (x \in A) \wedge (x \notin B) \iff (x \in A) \wedge \sim(x \in B).$$

We might pronounce the symbol $A - B$ as "A set-minus B" (this may not exactly be standard). Here are Venn diagrams for $A - B$ and $B - A$.

[22] As opposed to the difference, say, of real numbers.



$A - B$



$B - A$

We should be very careful in that *order matters* with the complement: $A - B$ need not equal $B - A$. In contrast, we always have $A \cup B = B \cup A$ and $A \cap B = B \cap A$.

5.2.12 Remark.

The notation $A - B$ clearly calls to mind a "difference," i.e., that we are "subtracting" or "deleting" the elements of B from A, but it is not wholly standard. In particular, if A and B are sets of real numbers, one might want to consider the set of genuine differences between elements of A and elements of B, i.e.,

$$\{x - y \mid x \in A,\ y \in B\}.$$

The notation $A \setminus B$ avoids this confusion with subtraction. Be careful not to write A/B like a quotient or fraction; the symbol A/B has a wholly different meaning in abstract algebra that we will not encounter in this course.

5.2.13 Example.

Verify each purported equality.

(i) $\{1, 2, 3\} - \{1, 2\} = \{3\}$.

(ii) $\{1, 2\} - \{1, 2, 3\} = \varnothing$.

(iii) $\mathbb{Z} - (-\infty, 0] = \mathbb{N}$.

(iv) $[0, 1] - \{1\} = [0, 1)$.

Solution. **(i)** This is a direct calculation: the only element of $\{1, 2, 3\}$ that is not in $\{1, 2\}$ is 3.

(ii) This is another direct calculation: there are no elements of $\{1, 2, 3\}$ that are not in $\{1, 2\}$.

(iii) This uses the definition of $\mathbb{N}$ as $\mathbb{N} := \{n \in \mathbb{Z} \mid n \geq 1\}$. Thus if $n \in \mathbb{Z} - (-\infty, 0]$, we have $n \in \mathbb{Z}$ and $n \notin (-\infty, 0]$, hence $n > 0$. But since $n \in \mathbb{Z}$, it must be the case that $n \geq 1$, hence $n \in \mathbb{N}$. This shows $\mathbb{Z} - (-\infty, 0] \subseteq \mathbb{N}$.

Conversely, if $n \in \mathbb{N}$, we have $n \in \mathbb{Z}$ and $n \geq 1$, hence $n \notin (-\infty, 0]$. Thus $n \in \mathbb{Z} - (-\infty, 0]$, and so $\mathbb{N} \subseteq \mathbb{Z} - (-\infty, 0]$.

(iv) *Proof in (mostly) symbols.* In part (iv) of Example 5.2.2, we showed $[0,1] = [0,1) \cup \{1\}$. Thus

$$x \in [0,1] \iff \big(x \in [0,1)\big) \vee (x = 1).$$

Consequently,

$$\begin{aligned} x \in [0,1] - \{1\} &\iff \big(x \in [0,1]\big) \wedge \big(x \notin \{1\}\big) \\ &\iff \big[\big(x \in [0,1)\big) \vee (x = 1)\big] \wedge (x \neq 1) \\ &\iff \big[\big(x \in [0,1)\big) \wedge (x \neq 1)\big] \vee \big((x = 1) \wedge (x \neq 1)\big). \end{aligned}$$

We have

$$\big(x \in [0,1)\big) \wedge (x \neq 1) \iff x \in [0,1).$$

This is true because if $x \in [0,1)$ and $x \neq 1$, then clearly $x \in [0,1)$. Conversely, if $x \in [0,1)$, then $x < 1$, so $x \neq 1$.

Next, the statement[23] $(x = 1) \wedge (x \neq 1)$ is always false. Consequently, the statement $x \in [0,1] \setminus \{1\}$ is equivalent to the statement $P \vee Q$, where P is the statement $x \in [0,1)$ and Q is the false statement $(x = 1) \wedge (x \neq 1)$. Hence the truth value of $P \vee Q$ is equivalent[24] to the truth value of P. That is,

$$\big[\big(x \in [0,1)\big) \wedge (x \neq 1)\big] \vee \big((x = 1) \wedge (x \neq 1)\big) \iff x \in [0,1),$$

and so

$$x \in [0,1] - \{1\} \iff x \in [0,1).$$

Proof in (more) words. Suppose $x \in [0,1] - \{1\}$. Then $x \in [0,1]$ and $x \notin \{1\}$, hence $x \neq 1$. That is, $0 \leq x \leq 1$ and $x \neq 1$, so $0 \leq x < 1$. Thus $x \in [0,1)$, and so $[0,1] - \{1\} \subseteq [0,1)$.

Conversely, suppose $x \in [0,1)$. Then $0 \leq x < 1$, so $0 \leq x \leq 1$ and $x \in [0,1]$. But $x < 1$, so $x \neq 1$, and therefore $x \notin \{1\}$. Hence $x \in [0,1] - \{1\}$, and so $[0,1) \subseteq [0,1] - \{1\}$. ▲

5.2.14 Theorem.

Let A and B be sets.

(i) $A - B \subseteq A$.

(ii) $A - A = \varnothing$.

(iii) $A - \varnothing = A$.

Proof. (i) Let $x \in A - B$. Then $x \in A$ and $x \notin B$. In particular, $x \in A$, and so $A - B \subseteq A$.

[23] If we assume that x is one particular, fixed number, then this sentence is really a statement and not a predicate.

[24] Recall the truth table definition of $\vee$ and check that if Q is false, then $P \vee Q$ is true if and only if P is true, while $P \vee Q$ is false if and only if Q is false.

(ii) Let $x \in A - A$. Then $x \in A$ and $x \notin A$, a contradiction. That is, we were wrong to assume $x \in A - A$, and so $A - A = \varnothing$.

(iii) Let $x \in A - \varnothing$. Then $x \in A$ and $x \notin \varnothing$. In particular, $x \in A$, and so $A - \varnothing \subseteq A$. (The statement $x \notin \varnothing$ is true but irrelevant here.) ■

5.2.15 Remark.

Suppose that A is a subset of an "underlying universal" set $\mathcal{U}$ (although not universal in the sense of Russell's paradox!). Then the notation $A^c := \mathcal{U} - A$ is sometimes used; the c obviously means "complement." For example, if we are considering the real numbers and its subsets, and $A \subseteq \mathbb{R}$, some authors might write A^c instead of $\mathbb{R} - A$. We will never do this.

5.2.4. Interactions of algebraic operations on sets.

The three fundamental operations $\cup$, $\cap$, and $-$ interact with each other in a variety of useful ways. We will state and prove a small number of identities, and if we need others, we will state and prove them as the need arises. Many of the proofs rely on the following parallels of set-theoretic algebraic operations and logical connectives.

$\cup$	$\vee$	$x \in A \cup B \iff (x \in A) \vee (x \in B)$
$\cap$	$\wedge$	$x \in A \cap B \iff (x \in A) \wedge (x \in B)$
$-$	$\sim$	$x \in A - B \iff (x \in A) \wedge \sim(x \in B)$

5.2.16 Example.

Let A, B, and C be sets. Then

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

This is where we finished on Friday, October 8, 2021 (Section 54).

Proof. We have

$$\begin{aligned} x \in A \cap (B \cup C) &\iff (x \in A) \wedge (x \in B \cup C) \\ &\iff (x \in A) \wedge \big((x \in B) \vee (x \in C)\big) \\ &\iff \big((x \in A) \wedge (x \in B)\big) \vee \big((x \in A) \wedge (x \in C)\big) \\ &\iff (x \in A \cap B) \vee (x \in A \cap C) \\ &\iff x \in (A \cap B) \cup (A \cap C). \end{aligned}$$

Here we have used the identity

$$P \wedge (Q \vee R) \equiv (P \wedge Q) \vee (P \wedge R)$$

from our study of basic logic. ■

5.2.17 Example.

Let A and B be sets. Then

$$(A - B) \cap B = \varnothing.$$

Proof. Assume to the contrary that $(A - B) \cap B \neq \varnothing$. Then there exists $x \in (A - B) \cap B$. But

$$\begin{aligned} x \in (A - B) \cap B &\iff (x \in A - B) \wedge (x \in B) \\ &\iff \big((x \in A) \wedge (x \notin B)\big) \wedge (x \in B) \\ &\iff (x \in A) \wedge \big((x \notin B) \wedge (x \in B)\big). \end{aligned}$$

Thus we reach the contradiction $x \in B$ and $x \notin B$. ■

5.2.18 Example.

Let A and B be sets.

(i) *Show that $\mathscr{P}(A - B) \subseteq \mathscr{P}(A)$.*

(ii) *Why did we not say $\mathscr{P}(A - B) \subseteq \mathscr{P}(A) - \mathscr{P}(B)$?*

Solution. (i) We have

$$\begin{aligned} C \in \mathscr{P}(A - B) &\iff C \subseteq A - B \\ &\iff (x \in C \implies x \in A - B) \\ &\iff \big(x \in C \implies (x \in A \wedge x \notin B)\big) \\ &\implies (x \in C \implies x \in A) \\ &\implies C \subseteq A \\ &\implies C \in \mathscr{P}(A). \end{aligned}$$

Another way to prove this is to establish first the auxiliary claim $\mathscr{P}(D) \subseteq \mathscr{P}(E)$ for any sets D and E with $D \subseteq E$ and then to take $D = A - B$ and $E = A$.

This is where we finished on Friday, October 8, 2021 (Section 53).

(ii) In the proof above, we saw that if $C \in \mathscr{P}(A - B)$, then it is the case that *if* $x \in C$, then $x \in A$ and $x \notin B$. This gave us $C \subseteq A$. This also tells us that *if there exists* $x \in C$, then $x \notin B$, and so *in this case* we do have $C \not\subseteq B$, thus $C \notin \mathscr{P}(B)$. But what if there did not exist $x \in C$ in the first place? That is, what if $C = \varnothing$?

Recall that if D is any set, then $\varnothing \in \mathscr{P}(D)$. Thus $\varnothing \in \mathscr{P}(A - B)$, $\varnothing \in \mathscr{P}(A)$, and $\varnothing \in \mathscr{P}(B)$. Consequently, $\varnothing \notin \mathscr{P}(A) - \mathscr{P}(B)$, and so $\mathscr{P}(A - B)$ cannot be a subset of $\mathscr{P}(A) - \mathscr{P}(B)$, for otherwise we would have $\varnothing \in \mathscr{P}(A) - \mathscr{P}(B)$. ▲

5.2.19 Remark.

Observe that

$$\mathscr{P}(\varnothing - \varnothing) = \mathscr{P}(\varnothing) = \{\varnothing\} \quad \text{and} \quad \mathscr{P}(\varnothing) - \mathscr{P}(\varnothing) = \{\varnothing\} - \{\varnothing\} = \varnothing.$$

Then taking $A = B = \varnothing$ in part (ii) of Example 5.2.18 implies $\{\varnothing\} \not\subseteq \varnothing$. We knew this before, since $\varnothing \in \{\varnothing\}$ and therefore $\{\varnothing\} \neq \varnothing$, but it is nice to have another reminder of that fact.

This is where we finished on Monday, October 11, 2021 (Section 54).

5.2.20 Example.

Let A, B, and C be sets. Then

$$A - (B \cup C) = (A - B) \cap (A - C).$$

Proof. We have

$$\begin{aligned}
x \in A - (B \cup C) &\iff (x \in A) \wedge (x \notin B \cup C) \\
&\iff (x \in A) \wedge \sim(x \in B \cup C) \\
&\iff (x \in A) \wedge \sim\big((x \in B) \vee (x \in C)\big) \\
&\iff (x \in A) \wedge \big(\sim(x \in B) \wedge \sim(x \in C)\big) \\
&\iff (x \in A) \wedge \big((x \notin B) \wedge (x \notin C)\big) \\
&\iff (x \in A) \wedge (x \in A) \wedge \big((x \notin B) \wedge (x \notin C)\big) \\
&\iff \big((x \in A) \wedge (x \notin B)\big) \wedge \big((x \in A) \wedge (x \notin C)\big) \\
&\iff (x \in A - B) \wedge (x \in A - C) \\
&\iff x \in (A - B) \cap (A - C).
\end{aligned}$$

■

5.3. Ordered pairs and Cartesian products.

5.3.1. The ordered pair.

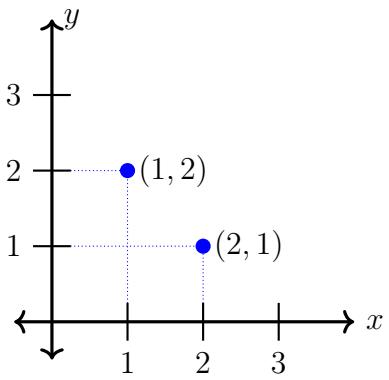
Let A be the set whose elements are precisely the integers 1 and 2. Then $A = \{1, 2\}$. Of course, we can also write $A = \{2, 1\}$, and so our set notation cannot reflect any "ordering" or "choice" of the elements of A.

Now consider the symbol $(1, 2)$ as the point[25] in the xy-plane whose x-coordinate is 1 and whose y-coordinate is 2. The "components" 1 and 2 completely determine this point, in the sense that if $x \neq 1$ or $y \neq 2$, we expect the point (x, y) to be something different. That is, for x, $y \in \mathbb{R}$, we expect

$$(x, y) = (1, 2) \iff (x = 1) \wedge (y = 2).$$

[25] That is, here $(1, 2)$ is not the open interval with endpoints 1 and 2, i.e., it is not the set $\{x \in \mathbb{R} \mid 1 < x < 2\}$.

In particular, $(1,2)$ is not the same point as $(2,1)$!



Also, whatever $(1,2)$ really is, we cannot define it as $(1,2) = \{1,2\}$ and expect to have this defining property; if $\{x,y\} = \{1,2\}$, there is no reason to expect $x = 1$ and $y = 2$.

More generally, if A and B are sets, and if $a \in A$ and $b \in B$, we want to define the symbol (a,b) to have the following property: for $c \in A$ and $d \in B$, we have

$$(a,b) = (c,d) \iff (a = c) \wedge (b = d).$$

This is a clear[26] characteristic of (a,b), but it is *not* a definition of (a,b): it does not tell us what (a,b) is in terms of more fundamental concepts (i.e., sets). Happily, there is such a definition of the ordered pair.

5.3.1 Definition.

Let A and B be a sets and let $x \in A$ and $y \in B$. The **ORDERED PAIR** *whose* **FIRST COMPONENT** *is x and whose* **SECOND COMPONENT** *is y is the set*

$$(x,y) := \big\{\{x\},\{x,y\}\big\}.$$

[26] Actually, this is not so clear, and it begs an interesting, if technical, question, which we set down here. If A is a set and a, $c \in A$, what does $a = c$ mean? It *should* mean that a and c are "the same," but what, concretely, does "the same" mean? We have a (perhaps vague) understanding of equality of numbers, but for other "mathematical objects," the right meaning of $=$ may not be so clear. After all, we had to define $A = B$ for sets A and B in terms of the behavior of their elements:

$$A = B \iff (x \in A \iff x \in B).$$

This might lead us to expect, given x, $y \in A$ (and here we are switching to the letters x and y for our elements, instead of a and c, purely for pointless aesthetic reasons)

$$x = y \iff \{x\} = \{y\},$$

and this suggests *defining* $x = y$ as the condition $\{x\} = \{y\}$. In other words, if A is a set and x, $y \in A$ with $\{x\} = \{y\}$, then we write $x = y$. We can check the equality $\{x\} = \{y\}$ in terms of the more "primitive" $\in$:

$$\{x\} = \{y\} \iff \big(z \in \{x\} \iff z \in \{y\}\big).$$

Thus if we have some abstract notion of what $\in$ means, we have a useful notion of equality for elements of an arbitrary set!

5.3.2 Theorem.

Let A and B be sets and let a, $c \in A$ and b, $d \in B$. Then $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$.

Proof. ($\Longleftarrow$) Suppose $a = c$ and $b = d$. Then

$$(a, b) = \big\{\{a\}, \{a, b\}\big\} = \big\{\{c\}, \{c, d\}\big\} = (c, d).$$

($\Longrightarrow$) Suppose $(a, b) = (c, d)$; then

$$\big\{\{a\}, \{a, b\}\big\} = \big\{\{c\}, \{c, d\}\big\}.$$

We consider two cases: $a = b$ and $a \neq b$.

Case 1: $a = b$. Hence $\{a, b\} = \{a\}$, and so

$$\big\{\{a\}\big\} = \big\{\{a\}, \{a\}\big\} = \big\{\{a\}, \{a, a\}\big\} = \big\{\{a\}, \{a, b\}\big\} = \big\{\{c\}, \{c, d\}\big\}.$$

Consequently, $\{c\} \in \big\{\{a\}\big\}$, and therefore $\{c\} = \{a\}$, which is to say $c = a$. Then

$$\big\{\{a\}\big\} = \big\{\{c\}, \{c, d\}\big\} = \big\{\{a\}, \{a, d\}\big\},$$

and so $\{a, d\} \in \big\{\{a\}\big\}$. Thus $\{a, d\} = \{a\}$. Hence $d \in \{a\}$, and so $d = a$.

We conclude

$$a = b = c = d,$$

so in particular $a = c$ and $b = d$.

Case 2: $a \neq b$. We have $\{c\} \in \big\{\{a\}, \{a, b\}\big\}$, and so either $\{c\} = \{a\}$ or $\{c\} = \{a, b\}$. If $\{c\} = \{a, b\}$, then a, $b \in \{c\}$, and so $a = c = b$. Thus $a = b$, a contradiction, and so $\{c\} \neq \{a, b\}$. The only possibility left is that $\{c\} = \{a\}$, thus $a = c$.

This implies

$$\big\{\{a\}, \{a, b\}\big\} = \big\{\{c\}, \{c, d\}\big\} = \big\{\{a\}, \{a, d\}\big\}.$$

Hence $\{a, b\} \in \big\{\{a\}, \{a, d\}\big\}$, and so either $\{a, b\} = \{a\}$ or $\{a, b\} = \{a, d\}$. We cannot have $\{a, b\} = \{a\}$, for then $b \in \{a\}$, which would imply $a = b$.

So, $\{a, b\} = \{a, d\}$, and therefore $b \in \{a, d\}$. Hence $b = a$ or $b = d$; we know $b \neq a$, so $b = d$. ■

This is where we finished on Monday, October 10, 2021 (Section 54).

5.3.3 Example.

Let A be a set and $x \in A$. Then

$$(x, x) = \big\{\{x\}, \{x, x\}\big\} = \big\{\{x\}, \{x\}\big\} = \big\{\{x\}\big\}.$$

5.3.4 Example.

We have

$$(1,2) = \big\{\{1\},\{1,2\}\big\}$$

but

$$(2,1) = \big\{\{2\},\{2,1\}\big\} = \big\{\{2\},\{1,2\}\big\}.$$

Although $\{1,2\} \in (1,2) \cap (2,1)$*, we have* $\{1\} \in (1,2)$ *and* $\{1\} \notin (2,1)$*. Likewise,* $\{2\} \in (2,1)$ *but* $\{2\} \notin (1,2)$*. Thus* $(1,2) \neq (2,1)$*.*

This is where we finished on Wednesday, October 11, 2021 (Section 54).

5.3.2. The Cartesian product.

5.3.5 Definition.

Let A *and* B *be sets. The* **Cartesian product** *of* A *and* B *is the set of all ordered pairs whose first component is an element of* A *and whose second component is an element of* B*:*

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

Thus

$$w \in A \times B \iff \big(\exists x \in A\ \exists y \in B : w = (x,y)\big).$$

Consequently, when proving statements about a Cartesian product $A \times B$, instead of starting with something like "Assume $w \in A \times B$. Then there exist $x \in A$ and $y \in B$ such that $w = (x,y)\ldots$," we will usually begin our proofs "Let $(x,y) \in A \times B$. Then $x \in A$ and $y \in B$ and so $\ldots$."

5.3.6 Example.

Let $A = \{1\}$ *and* $B = \{2,3\}$*. Determine all elements of* $A \times B$*.*

Solution. We have

$$A \times B = \{1\} \times \{2,3\} = \{(x,y) \mid x \in \{1\}, y \in \{2,3\}\} = \{(1,2),(1,3)\}.$$ ▲

5.3.7 Example.

Suppose that $P(x,y)$ *is a predicate with* $x \in D$ *and* $y \in E$ *for some sets* D *and* E*. (For example,* $P(x,y)$ *could be "*$x^2 + y^2 = 1$*" for* $x,\ y \in \mathbb{R}$*.) Previously we used inelegant phrasing like "the domain of* P *is the set of all* $x \in D$ *and* $y \in E$*." Now we see that the domain of* P *really is the set of all ordered pairs* $(x,y) \in D \times E$*. When we encounter repeated quantifiers, we can now compress, if we should choose, our notation:*

$$\forall x \in D\ \forall y \in E : P(x,y) \equiv \forall (x,y) \in D \times E : P(x,y).$$

For completeness, we quickly define a concept now that we will later study in greater depth.

5.3.8 Definition.

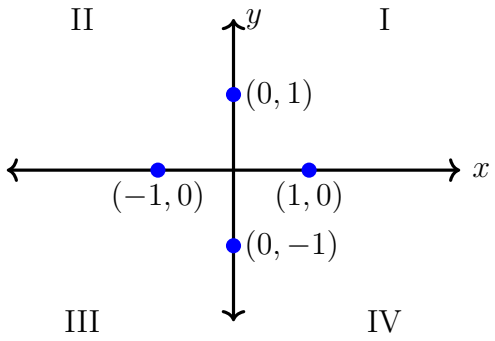
Let A and B be sets. A **RELATION FROM** A **TO** B *is a subset $R \subseteq A \times B$.*

Ideally, the word "relation" captures the idea that elements of $R \subseteq A \times B$ are ordered pairs $(x, y) \in A \times B$, where x and y are "related" by virtue of the property $(x, y) \in R$. The prepositions "from" and "to" are important in the definition above; they indicate not just "direction" but which factor comes first in the Cartesian product ("from" A "to" B, so A is the first factor and B is the second).

5.3.9 Example.

The **CARTESIAN PLANE** *or* ***xy*-PLANE** *is the set of all ordered pairs of real numbers. Thus the plane is the set $\mathbb{R} \times \mathbb{R}$. Sometimes we abbreviate this to $\mathbb{R}^2 := \mathbb{R} \times \mathbb{R}$. (More generally, if A is a set, we might write $A^2 := A \times A$.)*

We might recall from prior algebra the division of the plane into four "quadrants," I, II, III, and IV.



The "counterclockwise" unfurling of this division is related to the usual parametrization of the unit circle, which starts at the point $(1, 0)$ on the x-axis and proceeds successively (and circularly) through $(0, 1)$, $(-1, 0)$, and $(0, -1)$.

Assuming that we do not include the axes in any quadrant, we would say that Quadrant I is the set

$$\mathcal{Q}_{\mathrm{I}} := \left\{ (x, y) \in \mathbb{R}^2 \;\middle|\; x > 0,\ y > 0 \right\},$$

and $\mathcal{Q}_{\mathrm{I}}$ is a relation on $\mathbb{R}^2$.

This is where we finished on Wednesday, October 11, 2021 (Section 53).

There are any number of identities relating Cartesian products to the prior set-algebraic operations of unions, intersections, and complements. We state and prove just two; we will conjure others up as necessity dictates.

5.3.10 Example.

Let A, B, and C be sets. Then

$$(A \cap B) \times C = (A \times C) \cap (B \times C).$$

Proof. We need to show that the statement

$$w \in (A \cap B) \times C \iff w \in (A \times C) \cap (B \times C)$$

is true. Any such w will have the form $w = (x, y)$ for elements x and y of some sets (probably A, B, and/or C). So, instead of trying to prove the statement above, we jump to proving

$$(x, y) \in (A \cap B) \times C \iff (x, y) \in (A \times C) \cap (B \times C).$$

We have

$$\begin{aligned}
(x, y) \in (A \cap B) \times C &\iff (x \in A \cap B) \wedge (y \in C) \\
&\iff \big((x \in A) \wedge (x \in B)\big) \wedge (y \in C) \\
&\iff (x \in A) \wedge (x \in B) \wedge (y \in C) \wedge (y \in C) \\
&\iff \big((x \in A) \wedge (y \in C)\big) \wedge \big((x \in B) \wedge (y \in C)\big) \\
&\iff \big((x, y) \in A \times C\big) \wedge \big((x, y) \in B \times C\big) \\
&\iff (x, y) \in (A \times C) \cap (B \times C).
\end{aligned}$$

■

5.3.11 Example.

Let A, B, C, and D be sets with $A \subseteq C$ and $B \subseteq D$. Then $A \times B \subseteq C \times D$.

Proof. Let $(x, y) \in A \times B$. Then $x \in A$ and $y \in B$. Since $A \subseteq C$, we have $x \in C$, and since $B \subseteq D$, we have $y \in D$. Thus $(x, y) \in C \times D$. ■

Our prior geometric experience may motivate us to study "n-tuples," like ordered *triples* or *quadruples*. For example, if A, B, and C are sets and $x \in A$, $y \in B$, and $z \in C$, then we expect that the symbol (x, y, z) has the property

$$(x_1, y_1, z_1) = (x_2, y_2, z_2) \iff x_1 = x_2, y_1 = y_2, z_1 = z_2.$$

We might try to define $(x, y, z) := ((x, y), z)$, i.e., as the ordered pair whose first component is the ordered pair (x, y) and whose second component is the element z. Another, equally plausible definition would be $(x, y, z) := (x, (y, z))$, i.e., the ordered pair whose first component is the element x and whose second component is the ordered pair (y, z). So which definition should we use? And what will happen to an ordered *quadruple* — what horrible labyrinth of nested parentheses awaits?

A better idea is *not* to use ordered pairs to define higher-order tuples recursively but instead to use *functions* — a most glorious abstraction to which we now turn.

6. Functions

Intuitively, we say that one quantity is a "function" of another quantity if the value of the second quantity completely determines the value of the first. Using the words "quantity" and "value" may be too restrictive and call to mind only situations involving numbers. Certainly most of the functions in our mathematical education have involved numbers, but functions need not use any numbers.

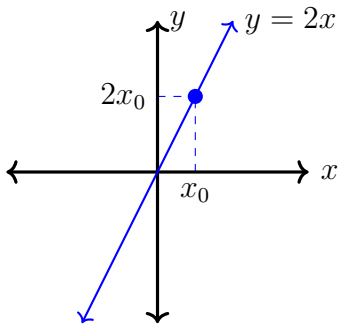
6.0.1 Example.

(i) *Consider the set of all people who showed up to class today and the set of all desks (or seats, or chairs) in the classroom. Now consider the association between a person who came to class and the desk in which that person sat throughout the class. Since people typically do not change seats in the middle of class, we can safely assume that each person sat in exactly one desk throughout the class. Thus we have a clear association between the set of all people who came to class and the set of all desks in the classroom: we pair each person with their desk. Also, no two people share a desk, so no desk gets paired with more than one person. Note that we need not pair every desk with a person; it is very possible that there are more desks in the classroom than people who showed up.*

(ii) *Consider the set of all students enrolled at Kennesaw State University and the set of all integers between 0 and 9 (inclusive), i.e., the set* $\mathbb{Z} \cap [0, 9]$. *Pair each student with the last digit of their student ID number. Then each student is associated with exactly one number. It is all but certain that every integer between 0 and 9 is paired with at least one person, probably many people.*

(iii) *We can pair every integer* $n \in \mathbb{Z}$ *with an even integer via the association of* n *and* $2n$. *This looks formulaic, and our intuition suggests we write this as* $f(n) = 2n$. *Since* $2n \in \mathbb{Z}$ *for all* $n \in \mathbb{Z}$, *we say that* f *is a function from* $\mathbb{Z}$ *to* $\mathbb{Z}$. *Every integer in* $\mathbb{Z}$ *is paired with a unique even integer: if* $f(n) = f(m)$, *then* $2n = 2m$, *and so* $n = m$. *Conversely, every even integer* $a \in \mathbb{Z}$ *can be written as* $a = 2n$ *for some* $n \in \mathbb{Z}$, *and so for each* $a \in \mathbb{Z}$ *there is* $n \in \mathbb{Z}$ *such that* $a = f(n)$.

(iv) *Let* $f(x) = 2x$. *Then* f *is a function from* $\mathbb{R}$ *to* $\mathbb{R}$. *If we are told to "graph* f*" or "graph* $y = 2x$*," then we produce the following picture.*



We note that the graph of f *is the set* $\{(x, 2x) \mid x \in \mathbb{R}\} \subseteq \mathbb{R} \times \mathbb{R}$.

Moving strictly beyond numbers, then, a function is a pairing or or association of elements of one set with elements of another. The pairing needs to be unique in that every element of the first set is matched with precisely one element of the second set. Often, though not always, there is a "formula" that explains how to determine the element of the second set in terms of an element of the first set. Since the concept of function fundamentally reduces to pairing, we can use ordered pairs to give the "true" definition of a function.

6.1. Fundamentals.

6.1.1. The true definition of a function.

6.1.1 Definition.

Let A and B be sets. A **FUNCTION** *f* **FROM** *A* **TO** *B is a set $f \subseteq A \times B$ with the property that for all $x \in A$ there exists a unique $y \in B$ such that $(x, y) \in f$. We abbreviate the statement "f is a function from A to B"*[27] *by $f \colon A \to B$.*

If $(x, y) \in f$, we write $y = f(x)$, pronounced "y equals f of x." Common synonyms for "function" include **MAP** *and* **MAPPING**. *The set A is the* **DOMAIN** *of A, while the set B is the* **CODOMAIN** *of A.*

This is where we finished on Monday, October 18, 2021.

6.1.2 Example.

Let $A = B = \{1\}$. Then $f = \{(1, 1)\}$ is a function from A to B.

In symbols, $f \subseteq A \times B$ is a function from A to B if

$$\forall x \in A \; \exists! y \in B : (x, y) \in f.$$

Equivalently, $f \subseteq A \times B$ is a function from A to B if

$$\big(\forall x \in A \; \exists y \in B : (x, y) \in f\big) \wedge \big(\forall x \in A \; \forall y_1, y_2 \in B \colon (x, y_1), (x, y_2) \in f \implies y_1 = y_2\big). \tag{6.1.1}$$

Since any function $f \colon A \to B$ is a subset of $A \times B$, such an f is a relation from A to B in the sense of Definition 5.3.8. We say that a relation $f \subseteq A \times B$ is a **WELL-DEFINED FUNCTION** from A to B if f satisfies the property (6.1.1).

[27] Occasionally we will use the redundant phrasing "Let $f \colon A \to B$ be a function" or "Consider the function $f \colon A \to B \ldots$" for emphasis. That is, sometimes we will treat the symbol $f \colon A \to B$ as a noun, not a statement.

6.1.3 Example.

Let $A = \{1, 2, 3\}$ and $B = \{4, 5, 6, 7\}$. Which of the following relations from A to B are well-defined functions from A to B?

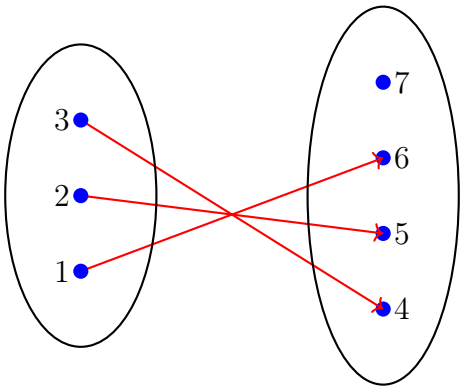
(i) $\{(1, 6), (2, 5), (3, 4)\}$

(ii) $\{(1, 4), (2, 4), (3, 6)\}$

(iii) $\{(1, 4), (2, 6)\}$

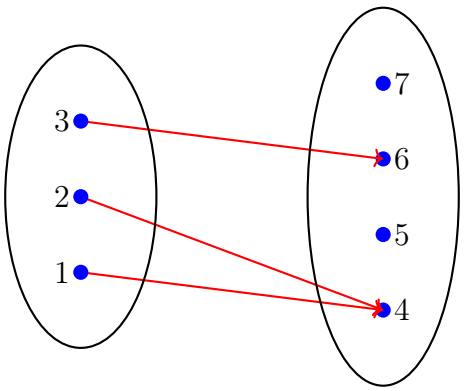
(iv) $\{(1, 5), (2, 4), (2, 5), (3, 7)\}$

Solution. **(i)** It may help to draw some "arrow diagrams." We put all the points of A in one "blob" and all the points of B in another and draw arrows connecting the points as the relations specify.



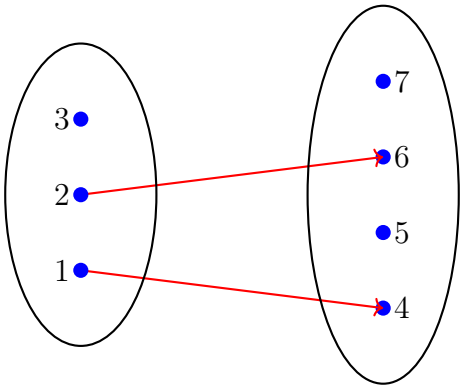
We see that each element of A is paired with exactly one element of B, so this is a function. As it happens, no element of B is paired with more than one element of A. Also, one element of B, 7, is not paired with any element of A, These two observations, however, do not affect the relation's status as a function.

(ii) We draw an arrow diagram.



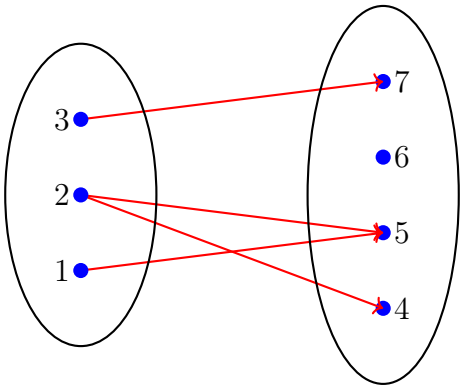
Again, each element of A is paired with exactly one element of B, and so this relation is a function. This time, though, one element of B is paired with more than one element of A.

(iii) We draw the arrow diagram.



This time we note that 3 is not paired with any element of B; that is, the relation contains no ordered pair of the form $(3, y)$ for some $y \in B$. And so this is not a function.

(iv) Here is the arrow diagram.



We see that 2 is paired with both 4 and 5; that is, the relation contains the pairs $(2, 4)$ and $(2, 5)$, and so this relation cannot be a function. ▲

6.1.4 Definition.

Let A and B be sets. The set of all functions from A to B is the set B^A.

That is,

$$B^A = \{R \subseteq A \times B \mid \forall x \in A \; \exists! y \in B : (x, y) \in R\}.$$

We will later show that if A "has m elements" and if B "has n elements" (the property of having a certain number of elements not yet being defined), then B^A "has n^m elements."

6.1.5 Example.

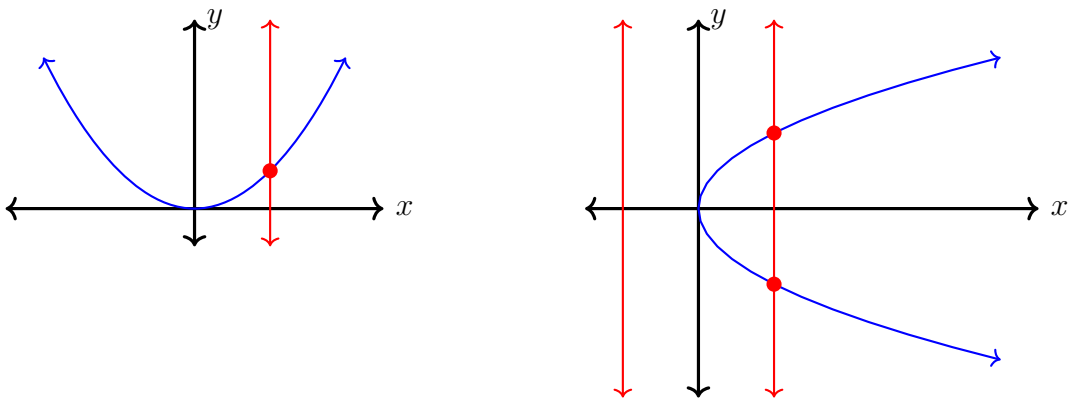
Let $A \subseteq \mathbb{R}$ and let $f \in \mathbb{R}^A$. The set

$$\{(x, f(x)) \mid x \in A\}$$

is often called the **GRAPH** *of f. Of course, we now know that a function's graph is the*

same as that function!

More generally, if S is a relation from A to $\mathbb{R}$, then S is a function if and only if S "passes the **VERTICAL LINE TEST***": any vertical line in the plane intersects (the graph of) S at most once. The relation graphed on the left is a function; the one on the right is not.*



Why "at most" once? If there is one intersection, there cannot be another; otherwise, two points on the graph of S would have the same x-coordinate but different y-coordinates, and then S would not be a function. But there need not be any intersection of a given vertical line with the graph of S; if that line has the form $x = c$, perhaps $c \notin A$.

Here is a more formal statement of the vertical line test: a relation S from $A \subseteq \mathbb{R}$ to $\mathbb{R}$ is a function from A to $\mathbb{R}$ if and only if, for each $c \in \mathbb{R}$, there exists at most one $d \in \mathbb{R}$ such that

$$(c, d) \in \{(c, y) \mid y \in \mathbb{R}\} \cap S.$$

Defining functions as sets of ordered pairs, even when compressed in set-builder notation, can be clunky. When possible, we will define functions via formulas (which is how we usually think of them anyway). For example, instead of saying "Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = x^2$," we might just say "Define the function $f : \mathbb{R} \to \mathbb{R} : x \mapsto x^2$." That is, $f = \{(x, x^2) \mid x \in \mathbb{R}\}$. The letters here are really arbitrary; we could say instead "Consider $g : \mathbb{R} \to \mathbb{R} : t \mapsto t^2$," and then we would have

$$g = \{(t, t^2) \mid t \in \mathbb{R}\} = \{(x, x^2) \mid x \in \mathbb{R}\} = f.$$

The same issue came up with sets defined by predicates, and we blithely kicked that to Footnote 19.

6.1.6 Remark.

We now have three "right-pointing" arrows at our disposal: $\Longrightarrow$, $\to$, and $\mapsto$. Confuse them not.

We have used the symbol $=$ in a variety of contexts: equality of real numbers, equality of sets, and equality of elements[28] in a set. What does it mean for functions to be equal?

[28] Recall the noodlescratching in Footnote 26.

6.1.7 Example.

Let $f\colon A \to B$ and $g\colon C \to D$. What does $f = g$ mean?

Solution. Since f and g are sets, $f = g$ means

$$w \in f \iff w \in g.$$

Let us try to be more descriptive. Since $f \subseteq A \times B$, we have $w \in f$ if and only if $w = (x, y)$ for some $(x, y) \in A \times B$. So, if $f = g$, then for each $(x, y) \in f$ we have $(x, y) \in g$.

We know in particular that if $x \in A$, then there is $y \in B$ such that $(x, y) \in f$, and so $(x, y) \in g$. Thus for all $x \in A$ we can find $y \in B$ such that $(x, y) \in g$. But since $g \subseteq C \times D$, we then have $x \in C$. That is, for every $x \in A$ it is the case that $x \in C$, and so $A \subseteq C$. An identical argument, which we leave to the reader, shows that $C \subseteq A$.

Thus if $f = g$, we have $A = C$. We cannot say anything particular about the relationship between B and D, however; put $B = \{n/2 \mid n \in \mathbb{Z}\}$ and consider $f\colon \mathbb{N} \to B\colon n \mapsto n/2$ and $g\colon \mathbb{N} \to \mathbb{R}\colon n \mapsto n/2$. But using the definition of set equality, the result $A = C$, and the definition of a function, we have

$$\forall x \in A\, \forall y \in B : (x, y) \in f \implies (x, y) \in g \quad \text{and} \quad \forall x \in A\, \forall z \in D : (x, z) \in g \implies (x, z) \in f.$$

Let us rewrite this in more familiar and palatable language. Take $x \in A$ and recall that, for $y \in B$, we write $y = f(x)$ if $(x, y) \in f$. Then $(x, f(x)) \in f$, and so $(x, f(x)) \in g$. But we write $z = g(x)$ whenever $(x, z) \in g$, and so $f(x) = g(x)$. Since $x \in A$ was arbitrary, we conclude

$$f = g \iff A = C \quad \text{and} \quad \forall x \in A : f(x) = g(x).$$

Of course, this is exactly what function equality should mean: the functions have the same domains and they agree at every point in those domains. But this was *not* precisely obvious from the set-theoretic definition of function equality. ▲

This is where we finished on Wednesday, October 20, 2021 (Section 53).

We may also use "piecewise" notation to (attempt to) define functions.

6.1.8 Example.

Explain why the formula

$$f(x) := \begin{cases} x, & 0 \le x \le 1 \\ 3 - x, & 1 \le x \le 2 \end{cases}$$

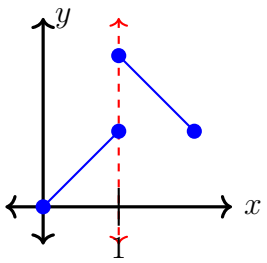
does not define a function $f\colon [0, 2] \to \mathbb{R}$. (Is the symbol $=$ even appropriate here?) How can you change this formula so that it does define a function on $[0, 2]$?

Solution. This formula really defines a set of ordered pairs of the form

$$f = \left\{(x, y) \in [0, 2] \times \mathbb{R} \;\middle|\; \left[(0 \le x \le 1) \wedge (y = x)\right] \vee \left[(1 \le x \le 2) \wedge (y = 3 - x)\right]\right\}.$$

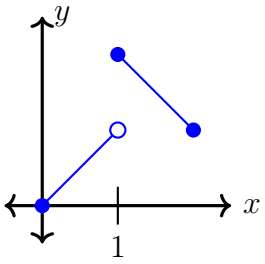
If f is a function, then it should satisfy two properties. First we need to show that for every $x \in [0,2]$ there exists $y \in \mathbb{R}$ such that $(x,y) \in f$. This is easy: just take y to be the number $f(x)$. In particular, $f(x)$ "makes sense" or "is defined" for each $x \in [0,2]$.

Next we need to show that for every pair of pairs (x,y_1), $(x,y_2) \in f$, it is the case that $y_1 = y_2$. This works out[29] unless $x = 1$. The problem is that $(1,1) \in f$ and $(1,2) \in f$, so f is not a function. Specifically, f fails the vertical line test at $x = 1$.



There are at least two ways that we could modify the formula f. One possibility is to break up the domain:

$$g(x) := \begin{cases} x, & 0 \le x < 1 \\ 3 - x, & 1 \le x \le 2. \end{cases}$$

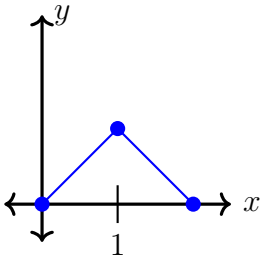


That is,

$$g = \left\{ (x,y) \in [0,2] \times \mathbb{R} \;\middle|\; \left[(0 \le x < 1) \wedge (y = x) \right] \vee \left[(1 \le x \le 2) \wedge (y = 3 - x) \right] \right\}.$$

Another is to tweak the "formula":

$$h(x) := \begin{cases} x, & 0 \le x \le 1 \\ 2 - x, & 1 \le x \le 2 \end{cases}$$



Observe that $1 = 2 - 1$, and so the pieces "match up." More precisely, here That is,

$$h = \left\{ (x,y) \in [0,2] \times \mathbb{R} \;\middle|\; \left[(0 \le x \le 1) \wedge (y = x) \right] \vee \left[(1 \le x \le 2) \wedge (y = 2 - x) \right] \right\}.$$

We leave it as exercises to verify the property that for $k = 1, 2$ it is the case that if $x \in [0,2]$, then there exists a unique $y \in \mathbb{R}$ such that $(x,y) \in f_k$. ▲

[29] Specifically, let $x \in [0,2] - \{1\}$. Then either $0 \le x < 1$ or $1 < x \le 2$. If $0 \le x < 1$ and (x,y_1), $(x,y_2) \in f$, then $y_1 = x$ and $y_2 = x$, so $y_1 = y_2$. If $1 < x \le 2$ and (x,y_1), $(x,y_2) \in f$, then $y_1 = 3 - x$ and $y_2 = 3 - x$, so $y_1 = y_2$ again.

This is where we finished on Wednesday, October 20, 2021 (Section 53).

6.1.2. Images, ranges, and pre-images.

6.1.9 Definition.

Let $f\colon A \to B$.

(i) *Let $C \subseteq A$. The* **IMAGE OF C UNDER f** *is the set*[30]

$$f(C) := \{f(x) \mid x \in C\} = \{y \in B \mid \exists x \in C : y = f(x)\} = \{y \in B \mid \exists x \in C : (x, y) \in f\}.$$

(ii) *The* **RANGE** *of f is the image of A under f, i.e., the set $f(A)$ as defined in part (i) above (with $C = A$).*

(iii) *Let $D \subseteq B$. The* **PRE-IMAGE OF D UNDER f** *is the set*

$$\begin{aligned} f^{-1}(D) := \{x \in A \mid f(x) \in D\} &= \{x \in A \mid \exists y \in D : y = f(x)\} \\ &= \{x \in A \mid \exists y \in D : (x, y) \in f\}. \end{aligned}$$

6.1.10 Remark.

We have not yet defined what the inverse of a function is, so the notation f^{-1} for a preimage is purely symbolic (even though we may mispronounce it as "f-inverse of D" from time to time).

6.1.11 Example.

Let $A = \{1, 2, 3\}$ and $B = \{4, 5, 6, 7\}$. Let $f = \{(1, 4), (2, 4), (3, 5)\}$, so $f\colon A \to B$ is a function. Calculate each of the following.

(i) $f(\{1\})$

(ii) $f(\{1, 2\})$

(iii) $f(\{1, 2, 3\})$

(iv) $f^{-1}(\{4\})$

(v) $f^{-1}(\{6\})$

(vi) $f^{-1}(\{5, 6\})$

(vii) $f^{-1}(\varnothing)$

[30] The second and third equalities are just additional ways of expressing what the image is. They follow so directly from the definition $f(C) := \{f(x) \mid x \in C\}$ that we do not write out their proofs explicitly.

Solution. **(i)** $f(\{1\}) = \{f(x) \mid x \in \{1\}\} = \{f(x) \mid x = 1\} = \{f(1)\} = \{4\}$.

(ii) $f(\{1,2\}) = \{f(x) \mid x \in \{1,2\}\} = \{f(1), f(2)\} = \{4,4\} = \{4\}$.

(iii) $f(\{1,2,3\}) = \{f(x) \mid x \in \{1,2,3\}\} = \{f(1), f(2), f(3)\} = \{4,4,5\} = \{4,5\}$.

(iv) $f^{-1}(\{4\}) = \{x \in A \mid f(x) = 4\} = \{1,2\}$.

(v) $f^{-1}(\{6\}) = \{x \in A \mid f(x) = 6\} = \varnothing$, since there exists no $x \in A$ such that $(x,6) \in f$.

(vi) $f^{-1}(\{5,6\}) = \{x \in A \mid f(x) \in \{5,6\}\} = \{3\}$.

(vii) $f^{-1}(\varnothing) = \{x \in A \mid f(x) \in \varnothing\} = \varnothing$, since it is impossible for $f(x) \in \varnothing$ for any x. ▲

6.1.12 Example.

Define $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2$. *Calculate each of the following.*

(i) $f([-1,1])$

(ii) $f^{-1}(\{1\})$

(iii) $f^{-1}(\{0\})$

Solution. **(i)** We have $f([-1,1]) = \{f(x) \mid x \in [-1,1]\} = \left\{x^2 \;\middle|\; -1 \le x \le 1\right\} = [0,1]$. Here is a rigorous proof of the last equality. First, suppose $-1 \le x \le 1$. Since $-1 \le x$, then $x^2 \le 1$, by properties of inequalities; since $x \le 1$, we also have $x^2 \le 1$, again by those delightful properties of inequalities. Thus $\left\{x^2 \;\middle|\; -1 \le x \le 1\right\} \subseteq [0,1]$. Now let $y \in [0,1]$. Assuming the existence of the square root function, we know $0 \le \sqrt{y} \le 1$ and $(\sqrt{y})^2 = 1$. Put $x = \sqrt{y}$ to get $-1 \le x \le 1$ and $x^2 = y$.

(ii) We have $f^{-1}(\{1\}) = \left\{x \in \mathbb{R} \;\middle|\; x^2 = 1\right\}$. We know $x^2 = 1$ if and only if $x = \pm 1$; specifically, if $x^2 = 1$, then $0 = x^2 - 1 = (x-1)(x+1)$, and so $x = \pm 1$, while it is easy to calculate $(\pm 1)^2 = 1$.

(iii) We have $f^{-1}(\{0\}) = \left\{x \in \mathbb{R} \;\middle|\; x^2 = 0\right\}$. We know that $x^2 = 0$ if and only if $x = 0$, and so $f^{-1}(\{0\}) = \{0\}$. ▲

This is where we finished on Friday, October 22, 2021 (Section 54).

6.1.3. Restrictions.

6.1.13 Theorem.

Let $f\colon A \to B$ *and let* $E \subseteq A$. *Let*

$$f\big|_E := \{(x,y) \in f \mid x \in E\}.$$

Then $f\big|_E\colon E \to B$, *and we call the function* $f\big|_E$ *the* **RESTRICTION OF** f **TO** E.

Proof. We need to show that (1) $f\big|_E \subseteq E \times B$ and (2) for all $x \in E$ there exists a unique $y \in B$ such that $(x, y) \in f\big|_E$.

The subset condition $f\big|_E \subseteq E \times B$ follows directly from the definition of $f\big|_E$: if $(x, y) \in f\big|_E$, then $(x, y) \in f$ with $x \in E$. Since $f \subseteq A \times B$, we have $y \in B$ whenever $(x, y) \in f$.

Now let $x \in E$. Then since $E \subseteq A$, we have $x \in A$, so there exists $y \in B$ such that $(x, y) \in f$. Since $x \in A$, we have $(x, y) \in f\big|_E$ by definition of $f\big|_E$.

Finally, suppose there exist y_1, $y_2 \in B$ such that (x, y_1), $(x, y_2) \in f\big|_E$. Then (x, y_1), $(x, y_2) \in f$ as well, and so $y_1 = y_2$. ■

6.1.14 Example.

Let $A = \{1, 2, 3\}$ and $B = \{4, 5, 6, 7\}$. Define $f\colon A \to B$ by $f := \{(1, 4), (2, 5), (3, 6)\}$. Then

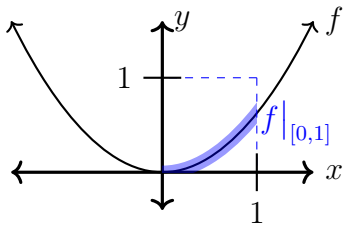
$$f\big|_{\{1\}} = \{(1, 4)\}, \qquad f\big|_{\{2,3\}} = \{(2, 5), (3, 6)\}, \qquad \text{and} \qquad f\big|_{\{1,3\}} = \{(1, 4), (3, 6)\}.$$

This is where we finished on Friday, October 22, 2021 (Section 53).

6.1.15 Example.

Define $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2$. Illustrate with a graph the distinction between f and $f\big|_{[0,1]}$.

Solution. We overlay the graph of f (in black) on the graph of $f\big|_{[0,1]}$ (in thicker blue).



The graph of $f\big|_{[0,1]}$ is but a snippet[31] of the larger picture of f. ▲

6.1.16 Remark.

Let $f\colon A \to B$ and $x \in A$. Then

$$f(\{x\}) = \{f(x)\} \qquad \text{and} \qquad f\big|_{\{x\}} = \{(x, f(x))\}.$$

In particular, $f(\{x\})$, $\{f(x)\}$, and $f\big|_{\{x\}}$ are sets; $f(\{x\}) \neq f(x)$; and $f(\{x\}) \neq f\big|_{\{x\}}$. We have $f(f^{-1}(\{f(x)\})) = \{f(x)\}$ and $\{x\} \subseteq f^{-1}(\{f(x)\})$, but $f^{-1}(\{f(x)\})$ may contain more elements than just x.

[31] Marley's ghost: "The dealings of my trade were but a drop of water in the comprehensive ocean of my business!"

Although varying the domain of a function genuinely changes that function, changing the codomain does not.

6.1.17 Example.

Let A, B, and C be sets with $B \subseteq C$. Suppose that $f\colon A \to B$. Then $f\colon A \to C$.

Proof. We know (1) $f \subseteq A \times B$ and (2) for all $x \in A$ there exists a unique $y \in B$ such that $(x, y) \in f$. We need to show (3) $f \subseteq A \times C$ and (4) for all $x \in A$ there exists a unique $y \in C$ such that $(x, y) \in f$.

First, since $B \subseteq C$, we have $A \times B \subseteq A \times C$ by Example 5.3.11. Then since $f \subseteq A \times B$, we have $f \subseteq A \times C$.

Next, let $x \in A$ and choose $y \in B$ such that $(x, y) \in f$. Since $B \subseteq C$, we also have $y \in C$. This gives the existence of $y \in C$ such that $(x, y) \in f$.

For uniqueness, suppose there exist y_1, $y_2 \in C$ such that (x, y_1), $(x, y_2) \in f$. Since $f \subseteq A \times B$, we find y_1, $y_2 \in B$, and since $f\colon A \to B$ is a function, we must have $y_1 = y_2$. ■

6.1.18 Example.

If $f\colon \mathbb{N} \to \mathbb{N}$, then $f\colon \mathbb{N} \to \mathbb{Z}$, $f\colon \mathbb{N} \to [0, \infty)$, $f\colon \mathbb{N} \to [1, \infty)$, and $f\colon \mathbb{N} \to \mathbb{R}$ are also functions.

6.1.19 Example.

The language that we use regarding codomains matters. For example, we can think of $f\colon \mathbb{N} \to \mathbb{N}\colon n \mapsto 2n$ as a function $f\colon \mathbb{N} \to \mathbb{Z}$ or $f\colon \mathbb{N} \to \mathbb{R}$. But the statements

$$\forall y \in \mathbb{Z}\ \exists x \in \mathbb{N} : y = f(x) \quad \text{and} \quad \forall y \in \mathbb{R}\ \exists x \in \mathbb{N} : y = f(x)$$

are both false. It is only when we define $\mathcal{E} := \{x \in \mathbb{N} \mid 2 \mid x\}$ that we have $f\colon \mathbb{N} \to \mathcal{E}$ and the statement

$$\forall y \in \mathcal{E}\ \exists x \in \mathbb{N} : y = f(x)$$

is true. The same function can have different codomains and different properties relative to those different codomains.

6.1.4. Sequences.

Functions defined on $\mathbb{N}$ tend to play a large role in many areas of mathematics, especially the more theoretical underpinnings of calculus.

6.1.20 Definition.

A **SEQUENCE** *in a set A is a function $f\colon \mathbb{N} \to A$. If $a_k := f(k)$ for $k \in \mathbb{N}$, one often writes $f = (a_k)$. Technically, then,*

$$(a_k) = \{(k, a_k) \mid k \in \mathbb{N}\} = \{(k, f(k)) \mid k \in \mathbb{N}\},$$

whereas the range of $f = (a_k)$ is

$$f(\mathbb{N}) = \{a_k \mid k \in \mathbb{N}\}.$$

The elements a_k (equivalently, the values $f(k)$) are the **TERMS** *of the sequence (a_k).*

6.1.21 Example.

For $k \in \mathbb{N}$ define $a_k := 2^k$. Then (a_k) is a sequence in $\mathbb{N}$, and thus a sequence in $\mathbb{R}$. We have

$$(a_k) = \{(k, 2^k) \mid k \in \mathbb{N}\} \quad \textit{but} \quad \{a_k \mid k \in \mathbb{N}\} = \{2^k \mid k \in \mathbb{N}\}.$$

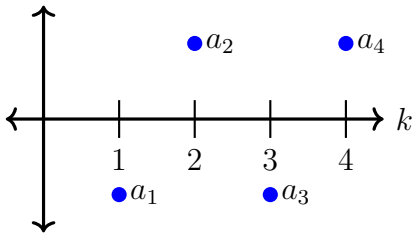
More generally, a sequence in A may be a function $f\colon [n_0, \infty) \cap \mathbb{Z} \to A$ for some $n_0 \in \mathbb{Z}$; when we study function composition and bijections we will see how to express such a sequence as a function on $\mathbb{N}$. There are good mathematical reasons to consider "doubly infinite" sequences in a set A as functions from $\mathbb{Z}$ to A, but we will not do so in this course. We can think of a sequence as an "infinite ordered list," which may allow repetition of elements on that list.

6.1.22 Example.

Define a sequence (a_k) in $\mathbb{R}$ by $a_k := (-1)^k$. Then

$$a_k = \begin{cases} 1, & 2 \mid k \\ -1, & 2 \nmid k. \end{cases}$$

The set of all terms of (a_k) is $\{1, -1\}$. Here is a graph of (a_k) for some small values of k.



This is where we finished on Monday, October 25, 2021 (Section 54).

Occasionally we will work with a "finite" sequence, and for this we need some new notation.

6.1.23 Definition.

For $n \in \mathbb{N}$ put

$$\mathcal{F}_n := [1, n] \cap \mathbb{N} = \{k \in \mathbb{N} \mid 1 \leq k \leq n\}.$$

For example, $\mathcal{F}_3 = \{1, 2, 3\}$. The notation $\mathcal{F}_n$ is by no means standard. We might

be tempted to euphemistically express $\mathcal{F}_n = \{1, \dots, n\}$, except the elision elicited by the ellipsis can be egregiously erroneous, e.g., do we really have $\{1, \dots, 4\} = \{1, 2, 3, 4\}$ or $\{1, \dots, 4\} = \{1, 2, 4\}$?

6.1.24 Definition.

Let A be a set. A **FINITE SEQUENCE** *in A is a function $f\colon \mathcal{F}_n \to A$ for some $n \in \mathbb{N}$. We might call a map $f\colon \mathcal{F}_n \to A$ a* **SEQUENCE OF LENGTH** *n* **IN** *A.*

6.1.25 Example.

Let A and B be sets. Show that we can "identify" the Cartesian product $A \times B$ with a certain subset of the set of all sequences of length 2 in $A \cup B$.

Solution. Recall that $A \times B = \{(x, y) \mid x \in A, y \in B\}$. Given $(x, y) \in A \times B$, define

$$f := \{(1, x), (2, y)\}. \tag{6.1.2}$$

Then $f\colon \{1, 2\} \mapsto A \cup B$ and $(x, y) = (f(1), f(2))$. Moreover, $f(1) = x \in A$ and $f(2) = y \in B$. Thus $f \in \left\{g \in (A \cup B)^{\mathcal{F}_2} \;\middle|\; g(1) \in A, g(2) \in B\right\} =: \mathcal{X}$.

Conversely, let $g \in \mathcal{X}$. Then $g(1) \in A$ and $g(2) \in B$, so $(g(1), g(2)) \in A \times B$. Although $\mathcal{X}$ and $A \times B$ are not the same set — technically $\mathcal{X} \subseteq \mathcal{F}_2 \times (A \cup B)$ — it turns out that the "natural" association (6.1.2) "completely" identifies elements of $\mathcal{X}$ with elements of $A \times B$. We will discuss this "complete" identification in more rigorous language later. ▲

Despite everyday language treating "sequence" and "series" as synonyms, they are quite distinct terms in mathematics, and we will never substitute "series" for "sequence." Thus one's experiences with this course may best be described as *A Finite Sequence of Unfortunate Events*.

This is where we finished on Monday, October 25, 2021 (Section 53).

This is where we finished on Wednesday, October 27, 2021 (Section 54), as well.

6.1.5. Function composition.

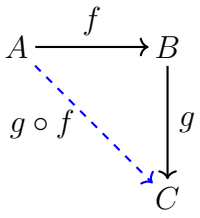
There are many ways to build new functions out of existing ones; we likely have significant life experience with adding or multiplying functions from $\mathbb{R}$ to $\mathbb{R}$. Whether or not algebraic operations are defined on the sets under consideration, we can always build new functions out of old ones via function composition.

6.1.26 Theorem.

Let $f\colon A \to B$ and $g\colon B \to C$. Define

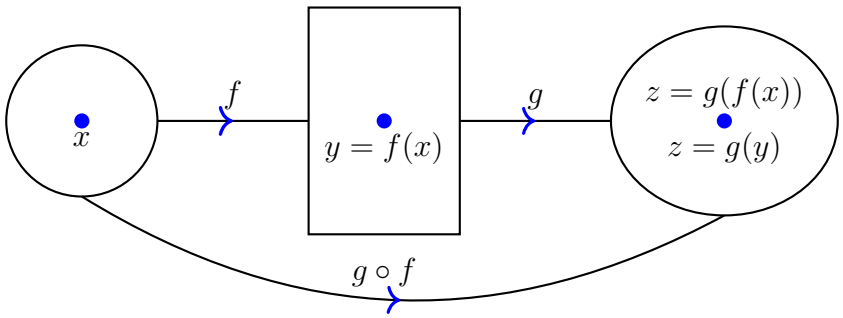
$$g \circ f := \left\{\big(x, g(f(x))\big) \;\middle|\; x \in A\right\}. \tag{6.1.3}$$

Then $g \circ f \colon A \to C$, and we call $g \circ f$ the **COMPOSITION OF g WITH f.**



Proof. For each $x \in A$ we have $f(x) \in B$, and so $g(f(x)) \in C$. Thus for each $x \in A$ there is $z \in C$ such that $(x, z) \in g \circ f$.

Now suppose (x, z_1), $(x, z_2) \in g$. Then $z_1 = g(f(x))$ and $z_2 = g(f(x))$, so $z_1 = z_2$. We have therefore verified that $g \circ f$ is a relation on $A \times C$ that satisfies the well-definedness property (6.1.1), and so $g \circ f$ is a function from A to C. ■



We should be mindful of the prepositions "of" and "with": $g \circ f$ is not, for us, the composition **of** f **with** g, which would be $f \circ g$. And that composition need not even be defined for $f \colon A \to B$ and $g \colon B \to C$! When composing g with f, it is most important that the image of f be a subset of the domain of g, i.e., $f(A) \subseteq B$.

6.1.27 Example.

(i) *Define $f \colon \mathbb{R} \to \mathbb{R} \colon x \mapsto x + 1$ and $g \colon \mathbb{R} \to \mathbb{R} \colon x \mapsto x^2$. Then*

$$(g \circ f)(x) = g(f(x)) = g(x + 1) = (x + 1)^2 = x^2 + 2x + 1$$

and

$$(f \circ g)(x) = f(g(x)) = f(x^2) = x^2 + 1.$$

Note that $(f \circ g)(x) = (g \circ f)(x)$ if and only if $x = 0$. We could also have written

$$g(f(x)) = (f(x))^2 = (x + 1)^2 \quad \text{and} \quad f(g(x)) = g(x) + 1 = x^2 + 1.$$

(ii) *Define $f \colon \mathbb{R} \to \mathbb{R} \colon x \mapsto x^2$ and $g \colon \mathbb{R} \to \mathbb{R} \colon x \mapsto x^3$. Then*

$$(g \circ f)(x) = g(x^2) = (x^2)^3 = x^6.$$

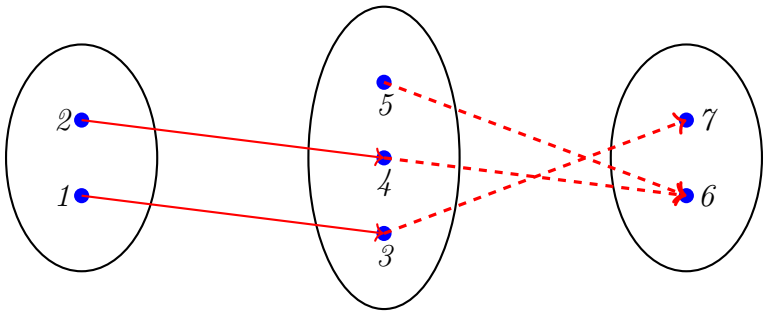
It also happens to be the case that $(f \circ g)(x) = x^6$.

(iii) *Let* $A = \{1, 2\}$, $B = \{3, 4, 5\}$, *and* $C = \{6, 7\}$. *Define relations*

$$f = \{(1, 3), (2, 4)\} \quad \text{and} \quad g = \{(3, 7), (4, 6), (5, 6)\}.$$

Then $f\colon A \to B$ *and* $g\colon B \to C$ *and*

$$g \circ f = \{(1, 7), (2, 6)\}.$$



6.1.28 Example.

Define functions f, g, $h\colon \mathbb{R} \to \mathbb{R}$ *by*

$$f(x) = x^2, \qquad g(x) = x^3, \qquad \text{and} \qquad h(x) = x^4.$$

The compositions $g \circ f$ *and* $h \circ g$ *are both defined, and we have*

$$(g \circ f)(x) = g(f(x)) = g(x^2) = (x^2)^3 = x^6$$

and

$$(h \circ g)(x) = h(g(x)) = h(x^3) = (x^3)^4 = x^{12}.$$

Note that $h \circ g \neq g \circ f$ *(not that we expect this equality at all).*

The compositions $h \circ (g \circ f)$ *and* $(h \circ g) \circ f$ *are both defined, and we have*

$$\big(h \circ (g \circ f)\big)(x) = h\big(g \circ f(x)\big) = h(x^6) = (x^6)^4 = x^{24},$$

and

$$\big((h \circ g) \circ f\big)(x) = (h \circ g)(f(x)) = (h \circ g)(x^2) = (x^2)^{12} = x^{24}.$$

We conclude the function equality $h \circ (g \circ f) = (h \circ g) \circ f$, *which is likely what we expect from prior experience.*

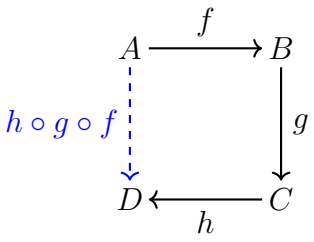
The previous example generalizes substantially to all triples of functions whose compositions are defined.

6.1.29 Theorem (Associativity of function composition).

Let $f\colon A \to B$, $g\colon B \to C$, and $h\colon C \to D$. Then

$$(h \circ g) \circ f = h \circ (g \circ f).$$

Because of this equality we just write $h \circ g \circ f$ without the parentheses.



Proof. Technically we must establish

$$w \in (h \circ g) \circ f \iff w \in h \circ (g \circ f).$$

Using (6.1.3), we have

$$(h \circ g) \circ f = \{(x, (h \circ g)(f(x))) \mid x \in A\} = \{(x, h(g(f(x)))) \mid x \in A\}$$

and

$$h \circ (g \circ f) = \{(x, h((g \circ f)(x))) \mid x \in A\} = \{(x, h(g(f(x)))) \mid x \in A\}.$$

At the level of ordered pairs, then, we conclude that the sets $(h \circ g) \circ f$ and $h \circ (g \circ f)$ are equal. ■

Consequently, while the *order* in which we arrange functions when composing them matters very much, we may "pairwise" evaluate the function composition however we please. This is exactly like the associativity of the logical connectives $\wedge$ and $\vee$, the set operations $\cup$ and $\cap$, and addition:

$$(P \wedge Q) \wedge R \equiv P \wedge (Q \wedge R), \qquad (A \cup B) \cup C = A \cup (B \cup C), \qquad (1+2)+3 = 1+(2+3), \ldots$$

Note, though, that function composition is not commutative ($f \circ g \neq g \circ f$, usually), unlike the commutativity of the operations above ($P \wedge Q \equiv Q \wedge P$, $A \cup B = B \cup A$, $1+2 = 2+1$).

6.2. Injections, surjections, and bijections.

Let A and B be sets and let $f \subseteq A \times B$. The following situations do not automatically prevent f from being a function from A to B.

1. There exist $x_1, x_2 \in A$ such that $x_1 \neq x_2$ and yet $f(x_1) = f(x_2)$. In other words, f maps more than one element of A to the same element of B.

2. There exists $y \in B$ such that $y \neq f(x)$ for all $x \in A$. In other words, no element of A is mapped to y.

These situations are, perhaps, less than "ideal" — in the first, f is "inefficient," while in the second, f is "deficient." Functions for which one or both of these situations do *not* occur have special names, and special properties.

6.2.1. Injections.

6.2.1 Definition.

A function $f\colon A \to B$ is **INJECTIVE (ON A)** *or* **ONE-TO-ONE (ON A)** *if for all x_1, $x_2 \in A$, whenever $f(x_1) = f(x_2)$, we have $x_1 = x_2$. An injective function is an* **INJECTION**.

In symbols, $f\colon A \to B$ is injective if

$$\forall x_1, x_2 \in A : f(x_1) = f(x_2) \implies x_1 = x_2.$$

Taking the contrapositive, we see that $f\colon A \to B$ is injective if

$$\forall x_1, x_2 \in A : x_1 \neq x_2 \implies f(x_1) \neq f(x_2).$$

Informally, we might say that a function is injective if "the outputs are equal if and only if[32] the inputs are equal," or, equivalently, "the outputs are different if and only if the inputs are different."

6.2.2 Example.

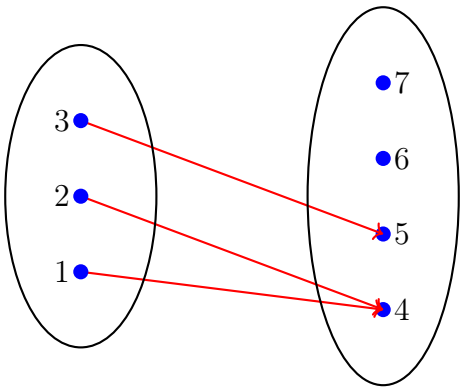
Determine if the following functions are injective.

(i) $f := \{(1,4),(2,4),(3,5)\}$ *as a function from* $A = \{1,2,3\}$ *to* $B = \{4,5,6,7\}$

(ii) $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2$

(iii) $f\colon [0,\infty) \to \mathbb{R}\colon x \mapsto x^2$

Solution. (i) Since $(1,4)$, $(2,4) \in f$, we have $f(1) = f(2)$ but of course $1 \neq 2$. Hence $f\colon A \to B$ is not injective. Here is the arrow diagram.



[32] If we swap the hypothesis and the conclusion in the if-then statement that defines injectivity, then we get the statement

$$\forall x_1, x_2 \in A : x_1 = x_2 \implies f(x_1) = f(x_2).$$

This is trivially true from the definition of a function.

(ii) We have $f(1) = 1 = f(-1)$, so $f\colon \mathbb{R} \to \mathbb{R}$ is not injective.

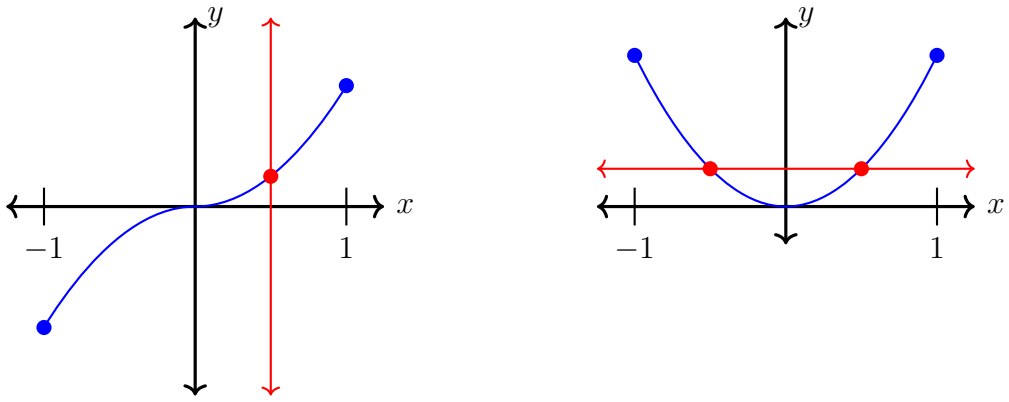
(iii) Suppose $f(x_1) = f(x_2)$ for some $x_1, x_2 \in [0, \infty)$. Then $x_1^2 = x_2^2$. Taking the square root, we find[33] $\sqrt{x_1^2} = \sqrt{x_2^2}$, and thus $|x_1| = |x_2|$. But $x_1, x_2 \geq 0$, so $|x_1| = x_1$ and $|x_2| = x_2$. Thus $x_1 = x_2$, so $f\colon [0, \infty) \to \mathbb{R}$ is injective. ▲

6.2.3 Remark.

When discussing matters of injectivity, it is critical to be clear about the domain of the function involved. Parts (ii) and (iii) of Example 6.2.2 show that a function $f\colon A \to B$ may not be injective but $f\big|_E\colon E \to B$ could be injective.

6.2.4 Example.

Let $E \subseteq \mathbb{R}$. A function $f\colon E \to \mathbb{R}$ must pass the following **HORIZONTAL LINE TEST** *to be injective: for each $d \in \mathbb{R}$ the graph of $y = d$ intersects the graph of f at most once. The function graphed on the left is injective from $[-1, 1]$ to $\mathbb{R}$; the one on the right is not.*



That is, $f\colon E \to \mathbb{R}$ is injective if for each $d \in \mathbb{R}$ there exists at most one $c \in E$ such that

$$(c, d) \in \{(x, d) \mid x \in \mathbb{R}\} \cap f.$$

Contrast this with the vertical line test from Example 6.1.5.

The composition of injections is an injection.

6.2.5 Theorem.

Suppose that $f\colon A \to B$ and $g\colon B \to C$ are injective. Then $g \circ f\colon A \to C$ is also injective.

Proof. Suppose that for some $x_1, x_2 \in A$ we have $(g \circ f)(x_1) = (g \circ f)(x_2)$. That is, $g(f(x_1)) = g(f(x_2))$. We must show that $x_1 = x_2$.

Since g is injective, we must have $f(x_1) = f(x_2)$. (If it helps to parse the parentheses,

[33] Here we are using the identity $\sqrt{x^2} = |x|$, valid for all $x \in \mathbb{R}$. The absolute value really is necessary: $\sqrt{(-2)^2} = \sqrt{4} = 2 = |-2|$, and $\sqrt{(-2)^2} \neq -2$.

write $y_1 = f(x_1)$ and $y_2 = f(x_2)$ to find $g(y_1) = g(y_2)$, and thus $y_1 = y_2$.) And since f is injective we must have $x_1 = x_2$. ■

This is where we finished on Wednesday, October 27, 2021 (Section 53).

6.2.2. Surjections.

6.2.6 Definition.

A function $f\colon A \to B$ is **SURJECTIVE** *or, more simply,* **ONTO** (B) *if for each $y \in B$ there exists $x \in A$ such that $y = f(x)$. For extra emphasis, sometimes we say that $f\colon A \to B$ is* **SURJECTIVE ONTO** B *or* **SURJECTIVE FROM** A **(ON)TO** B*. A surjective function is a* **SURJECTION**.

In symbols, $f\colon A \to B$ is surjective if

$$\forall y \in B\ \exists x \in A : y = f(x).$$

We make no requirement of uniqueness of this x. Discussions of surjections are one of the rarer times that we must mention a function's codomain; we must be able to solve the equation $y = f(x)$ for each y in the codomain.

6.2.7 Example.

Determine if the following functions are surjective.

(i) $f = \{(1,4),(2,5),(3,5)\}$ *as a function from* $A = \{(1,2,3\}$ *to* $B = \{4,5,6,7\}$

(ii) $f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2$

(iii) $f\colon \mathbb{R} \to [0,\infty)\colon x \mapsto x^2$

Solution. (i) There is no $x \in A$ such that $(x,6) \in f$ or $(x,7) \in f$, and so f is not surjective. However, $f\colon A \to \{4,5\}$ is surjective.

(ii) There exists no $x \in \mathbb{R}$ such that $x^2 = -1$, and so $f\colon \mathbb{R} \to \mathbb{R}$ is not surjective.

(iii) Let $y \in [0,\infty)$. Then $\sqrt{y} \in [0,\infty) \subseteq \mathbb{R}$, and $f(\sqrt{y}) = y$. Thus $f\colon \mathbb{R} \to [0,\infty)$ is surjective. ▲

6.2.8 Remark.

Changing the codomain of a function can alter whether or not that function is surjective; contrast parts (ii) and part (iii) of Example 6.2.7.

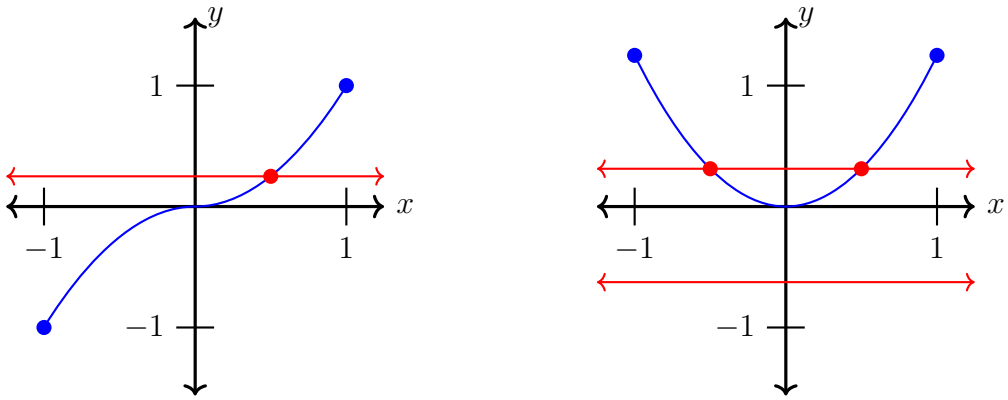
6.2.9 Example.

A function $f\colon A \to B$ is surjective if and only if $f(A) = B$.

Proof. For any $E \subseteq A$ it is the case that $f(E) \subseteq B$, so in particular $f(A) \subseteq B$. Now let $y \in B$. By definition of surjectivity, there is $x \in A$ such that $y = f(x)$. Hence $y \in f(A)$, and so $f(A) \subseteq B$. ■

6.2.10 Example.

Let A, $B \subseteq \mathbb{R}$. A function $f\colon A \to B$ must pass another version of the **HORIZONTAL LINE TEST** *to be surjective from A to B: for each $d \in B$ the graph of $y = d$ intersects the graph of f at least once. The function graphed on the left is surjective from $[-1, -1]$ to $[-1, 1]$; the one on the right is not.*



That is, $f\colon A \to B$ is surjective for each $d \in B$ we have

$$\{(x, d) \mid x \in A\} \cap f \neq \varnothing.$$

Contrast this with the horizontal line test (for injectivity) from Example 6.2.4.

The composition of surjections is a surjection.

6.2.11 Theorem.

Suppose that $f\colon A \to B$ and $g\colon B \to C$ are surjective. Then $g \circ f\colon A \to C$ is also surjective.

Proof. Let $z \in C$. We must find $x \in A$ such that $z = (g \circ f)(x) = g(f(x))$.

We work backwards. First, since $g\colon B \to C$ is surjective, there is $y \in B$ such that $g(y) = z$. Next, since $f\colon A \to B$ is surjective, there is $x \in A$ such that $f(x) = y$. Thus

$$g(f(x)) = g(y) = z,$$

as desired. ■

This is where we finished on Friday, October 29, 2021 (Section 54).

6.2.3. Bijections.

A bijection is the best of both worlds.

6.2.12 Definition.

A function $f\colon A \to B$ is **BIJECTIVE** *if f is injective on A and surjective from A to B. A bijective function is a* **BIJECTION**.

6.2.13 Example.

Let a, $b \in \mathbb{R}$ with $a < b$. Then the map

$$f\colon [0,1] \to [a,b]\colon t \mapsto (1-t)a + tb$$

is bijective. In particular,

$$[a,b] = \{(1-t)a + tb \mid 0 \le t \le 1\}. \tag{6.2.1}$$

Proof. Throughout this proof we will use the fact that since $a < b$ we have $b - a > 0$ and in particular $b - a \neq 0$. (Note that if $a = b$, then $[a,b] = \{a\}$. But then $f(0) = f(1) = a$, and so f is not injective, hence not a bijection.) We will not mention why this is true again.

Also, we are using t as the independent variable of f; this

We first discuss why we expect the set equality (6.2.1) to be true. We have

$$\begin{aligned} a \le (1-t)a + tb \le b &\iff a \le a - ta + tb \le b \\ &\iff 0 \le -ta + tb \le b - a \\ &\iff 0 \le t(b-a) \le b - a. \end{aligned}$$

Thus $b - a \neq 0$ implies

$$0 \le t \le 1 \iff 0 \le t(b-a) \le b - a.$$

So, if we start with $0 \le t \le 1$, then the chain of if-and-only-if statements above forces $a \le (1-t)a + tb \le b$.

This is where we finished on Friday, October 29, 2021 (Section 54).

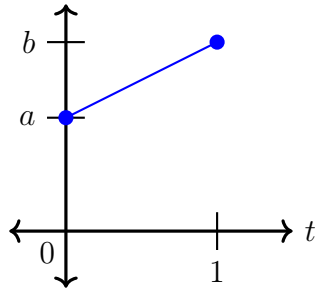
This, however, merely shows the *set inclusion*

$$\{(1-t)a + tb \mid 0 \le t \le 1\} \subseteq [a,b].$$

We will prove the more powerful *set equality* in (6.2.1). This will be a consequence of the bijectivity of f (specifically, the surjectivity of f and Example 6.2.9), which we now establish.

Here is a graph of f, which also suggests (via the horizontal line tests) but does not prove

its bijectivity.



For injectivity, let t_1, $t_2 \in [0,1]$. Then

$$\begin{aligned} f(t_1) = f(t_2) &\iff (1-t_1)a + t_1 b = (1-t_2)a + t_2 b \\ &\iff a - t_1 a + t_1 b = a - t_2 a + t_2 b \\ &\iff -t_1 a + t_1 b = -t_2 a + t_2 b \\ &\iff t_1(b-a) = t_2(b-a). \end{aligned}$$

Since $b - a \neq 0$, we may divide in the last equality to find

$$t_1(b-a) = t_2(b-a) \iff t_1 = t_2.$$

Thus $f(t_1) = f(t_2)$ implies $t_1 = t_2$.

For surjectivity, let $x \in [a,b]$. We need to find $t \in [0,1]$ such that $x = f(t)$. In other words, we must solve the equation $x = f(t)$ for t given x. We have

$$\begin{aligned} x = f(t) &\iff (1-t)a + tb = x \\ &\iff a - ta + tb = x \\ &\iff a + t(b-a) = x \\ &\iff t(b-a) = x - a \\ &\iff t = \frac{x-a}{b-a}. \end{aligned}$$

Here we were able to divide by $b - a$ since $b - a \neq 0$. Since all of our algebraic steps above are reversible, the if-and-only-if statements are justified. Thus we have shown

$$f\left(\frac{x-a}{b-a}\right) = x.$$

We do need to check, though, that $(x-a)/(b-a) \in [0,1]$. We accomplish this with another sequence of if-and-only-if manipulations:

$$\begin{aligned} 0 \leq \frac{x-a}{b-a} \leq 1 &\iff 0 \leq x - a \leq b - a \\ &\iff a \leq x \leq b. \end{aligned}$$

Here we used $b - a > 0$ to preserve the directions of the inequalities. Thus if $x \in [a,b]$, then $(x-a)/(b-a) \in [0,1]$. ■

6.2.14 Remark.

We think of $f(t) = (1-t)a + tb$ as a "parametrization" of the interval $[a, b]$. We start at "time" $t = 0$ (this is why we are using t for the independent variable of f) with $f(0) = a$ and move right until at time $t = 1$ we have $f(1) = b$. Along the way, we pass through each number x in $[a, b]$ precisely once.

The following result gives us a "test" for bijectivity that may be easier to apply than checking injectivity and surjectivity separately.

6.2.15 Theorem.

A function $f\colon A \to B$ is bijective if and only if for each $y \in B$ there exists a unique $x \in A$ such that $y = f(x)$.

Proof. ($\Longrightarrow$) First suppose that f is bijective and let $y \in B$. Since f is surjective, there exists $x \in A$ such that $y = f(x)$. And since f is injective, if there is $w \in A$ such that $y = f(w)$ as well, then $f(x) = f(w)$ and so $w = x$. Thus this x is unique.

($\Longleftarrow$) Suppose that for each $y \in B$ there exists a unique $x \in A$ such that $y = f(x)$. The "existence" part of this assumption guarantees that f is surjective. For injectivity, suppose $f(x_1) = f(x_2)$ for some x_1, $x_2 \in A$. Put $y = f(x_1)$; the uniqueness part of the assumption then forces $x_1 = x_2$. ■

In symbols, a function $f\colon A \to B$ is a bijection if and only if

$$\forall y \in B\ \exists! x \in A : y = f(x).$$

This should feel quite similar to the definition of a function — except there is a swap of both quantifiers and variables going on. Recall that a relation $f \subseteq A \times B$ is a function from A to B if and only if

$$\forall x \in A\ \exists! y \in B : (x, y) \in f$$

Thus a relation $f \subseteq A \times B$ is a bijection from A to B if and only if

$$\big(\forall x \in A\ \exists! y \in B : (x, y) \in f\big) \wedge \big(\forall y \in B\ \exists! x \in A : (x, y) \in f\big).$$

6.2.16 Example.

Show that

$$f\colon [0, \infty) \to [0, 1)\colon x \mapsto \frac{x}{1+x}$$

is bijective.

Proof. Observe that since $-1 \notin [0, \infty)$, f is defined on $[0, \infty)$. We need to show that for each $y \in [0, 1)$ there exists a unique $x \in [0, \infty)$ such that $y = f(x)$. That is, we need to solve

$$y = \frac{x}{1+x}$$

uniquely for x given y, and we must also check that our x satisfies $0 \leq x$.

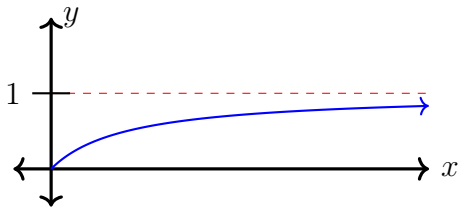
By standard properties of algebra, we have

$$\begin{aligned} y = \frac{x}{1+x} &\iff (1+x)y = x \\ &\iff y + xy = x \\ &\iff y = x - xy \\ &\iff y = x(1-y) \\ &\iff \frac{y}{1-y} = x. \end{aligned}$$

Note that every step above was reversible, so the if-and-only-if statements are justified. In particular, the algebra above shows that for all $y \in [0,1)$, we have $y = f(x)$ if and only if $x = y/(1-y)$.

Now, given $0 \leq y < 1$, we have $1 - y > 0$, and so $y/(1-y) \geq 0$. That is, $x = y/(1-y) \in [0,\infty)$. And so we have shown that given $y \in [0,1)$, there is $x \in [0,\infty)$ such that $y = f(x)$ and moreover that x must equal $y/(1-y)$, which is to say that this x is unique.

Here is a graph of f.



We see that $f(0) = 0$ and, for x "large," the values of $f(x)$ approach 1. In calculus terms, f is continuous on $[0,\infty)$ and $\lim_{x\to\infty} f(x) = 1$, which means that $f([0,\infty)) = [0,1)$. ■

6.2.17 Remark.

Typically in applying the "test for bijectivity" to a function $f\colon A \to B$, one starts with an arbitrary $y \in B$ and attempts to solve the equation $y = f(x)$. Ideally, one finds a "formula for x in terms of y":

$$y = f(x) \iff x = g(y)$$

for some function $g\colon B \to A$. The $\iff$ above is precisely the $\exists!$ quantifier in the test for bijectivity. The direction $\Longrightarrow$ is uniqueness: if $y = f(x)$, then the only possibility for x is $x = g(y)$. The direction $\Longleftarrow$ is existence: taking $x = g(y)$, one has $f(g(y)) = y$.

Since composition preserves injectivity (Theorem 6.2.5) and surjectivity (Theorem 6.2.11), composition also preserves bijectivity.

6.2.18 Theorem.

Let $f\colon A \to B$ and $g\colon B \to C$ be bijections. Then $g \circ f\colon A \to C$ is a bijection.

6.3. Inverses.

Ideally, an inverse of a process should undo that process and return us to where we started. Subtraction is the inverse of addition in that if we start with $x \in \mathbb{R}$, add $y \in \mathbb{R}$, and then subtract $y \in \mathbb{R}$, we are left with x:

$$(x + y) - y = x + (y - y) = x + 0 = x.$$

The inverse of procedures performed successively should undo those procedures in the reverse order. If we get dressed by putting on our socks and then our shoes, then we get undressed by taking off our shoes first and then our socks.

6.3.1 Example.

Prior life experience suggests that inverting a function $y = f(x)$ amounts to "solving for x in terms of y." Illustrate this with $f(x) = 2x + 1$.

Solution. Suppose we want to solve $y = 2x + 1$ for x given y. We first subtract 1 to find $y - 1 = 2x$. Then we divide by 2 to get $(y - 1)/2 = x$.

Let us analyze this at the level of function notation. Put

$$f(x) = 2x + 1 \quad \text{and} \quad g(y) = \frac{y - 1}{2}.$$

Then $g \circ f$ should amount to "doing f first, and then g." But "doing g" should amount to "undoing f," and so we expect $(g \circ f)(x) = x$. We check

$$(g \circ f)(x) = g(f(x)) = \frac{f(x) - 1}{2} = \frac{1}{2}\big((2x + 1) - 1\big) = \frac{1}{2}(2x) = x.$$

It also turns out that

$$(f \circ g)(y) = f(g(y)) = 2g(y) + 1 = 2\left(\frac{y - 1}{2}\right) + 1 = (y - 1) + 1 = y.$$

That is, we have established the identities

$$(g \circ f)(x) = x \text{ for all } x \in \mathbb{R} \quad \text{and} \quad (f \circ g)(y) = y \text{ for all } y \in \mathbb{R}. \tag{6.3.1}$$

▲

Whatever "inverse" means, the functions f and g in Example 6.3.1 must be inverses of each other. The identities (6.3.1) capture the true meaning of inverse at the level of functions.

This is where we finished on Monday, November 1, 2021 (Section 54).

6.3.2 Definition.

Let $f\colon A \to B$. A function $g\colon B \to A$ is an **INVERSE** *of f if*

$$(g \circ f)(x) = x \text{ for all } x \in A \quad \text{and} \quad (f \circ g)(y) = y \text{ for all } y \in B.$$

Interchanging A and B throughout the definition above, it follows immediately that $g\colon B \to A$ is an inverse of $f\colon A \to B$ if and only if f is an inverse of g.

6.3.3 Remark.

If we are thinking about an inverse of $f\colon A \to B$, we are fond of using the symbol x for an element of A (and thus the independent variable of f) and y for an element of B (and therefore the independent variable of the inverse). Of course, when working with formulas for functions, letters really are arbitrary ($f\colon \mathbb{R} \to \mathbb{R}\colon x \mapsto x^2$ is the same as $g\colon \mathbb{R} \to \mathbb{R}\colon t \mapsto t^2$) but consistent use of particular letters can illustrate important differences and patterns.

6.3.4 Example.

Let $A = \{1, 2, 3\}$, $B = \{4, 5, 6\}$, and $f = \{(1, 4), (2, 5), (3, 6)\}$. Find an inverse $g\colon B \to A$ for the function $f\colon A \to B$.

Solution. Since $f(1) = 4$, $f(2) = 5$, and $f(3) = 6$, if we want to "undo" f, we should take $g(4) = 1$, $g(5) = 2$, and $g(6) = 3$. That is, $g = \{(4, 1), (5, 2), (6, 3)\}$.

We examine this at the level of composition. We have

$$\begin{aligned} g \circ f &= \big\{\big(x, g(f(x))\big) \mid x \in A\big\} = \big\{\big(1, g(f(1))\big), \big(2, g(f(2))\big), \big(3, g(f(3))\big)\big\} \\ &= \{(1, g(4)), (2, g(5)), (3, g(6))\} = \{(1, 1), (2, 2), (3, 3)\} \end{aligned}$$

and

$$\begin{aligned} f \circ g &= \big\{\big(y, f(g(y))\big) \mid y \in B\big\} = \big\{\big(4, f(g(4))\big), \big(5, f(g(5))\big), \big(6, f(g(6))\big)\big\} \\ &= \{(4, f(1)), (5, f(2)), (6, f(3))\big\} = \{(4, 4), (5, 5), (6, 6)\}. \end{aligned}$$

We see that, indeed,

$$(g \circ f)(x) = x \text{ for all } x \in A \quad \text{and} \quad (f \circ g)(y) = y \text{ for all } y \in B. \qquad \blacktriangle$$

This is where we finished on Monday, November 1, 2021.

The notion of inverse invites us to consider questions that one can pose about almost any mathematical object — when does it exist, and is it unique? The following theorem gives conditions under which an inverse exists and assures us that the inverse, if it exists, is unique.

6.3.5 Theorem.

A function $f\colon A \to B$ has an inverse $g\colon B \to A$ if and only if $f\colon A \to B$ is bijective. If f is bijective, then the following also hold.

(i) *If $g_1,\ g_2\colon B \to A$ are inverses of f, then $g_1 = g_2$. That is, the inverse is unique, and we denote it by f^{-1}. In particular,*

$$f^{-1} = \{(y,x) \in B \times A \mid (x,y) \in f\} = \{(f(x),x) \mid x \in A\}.$$

(ii) *The inverse $f^{-1}\colon B \to A$ is also bijective and $(f^{-1})^{-1} = f$.*

Proof. ($\Longrightarrow$) Suppose that $f\colon A \to B$ has an inverse $g\colon B \to A$. First we show that f is injective. If $f(x_1) = f(x_2)$ for some $x_1,\ x_2 \in A$, then $g(f(x_1)) = g(f(x_2))$ since g is a function. But $g(f(x_1)) = x_1$ and $g(f(x_2)) = x_2$ by definition of inverse, thus $x_1 = x_2$.

Now we show that f is surjective. Let $y \in B$. Since $g(y) \in A$, we find $y = f(g(y))$ by definition of inverse.

($\Longleftarrow$) Suppose that $f\colon A \to B$ is bijective. The method of Example 6.3.4 suggests that we attempt to define an inverse of f as

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

We first need to show that g is a function from B to A. Certainly $g \subseteq B \times A$ by definition.

Now let $y \in B$. We need to find $x \in A$ such that $(y,x) \in g$. Since $f\colon A \to B$ is bijective, and therefore surjective, there is $x \in A$ such that $y = f(x)$, and so $(x,y) \in f$. Thus $(y,x) \in g$. That is, for all $y \in B$, there exists $x \in A$ such that $(y,x) \in g$.

Now we need to check the uniqueness of this x. That is, suppose $(y,x_1),\ (y,x_2) \in g$. We need to show $x_1 = x_2$. Since $(y,x_1) \in g$, we have $(x_1,y) \in f$, and so $y = f(x_1)$. Likewise, we obtain $y = f(x_2)$. Thus $f(x_1) = f(x_2)$. Since $f\colon A \to B$ is bijective, and therefore injective, we have $x_1 = x_2$.

We have therefore verified that $g\colon B \to A$ is a function. Last, we need to check that

$$(g \circ f)(x) = x \text{ for all } x \in A \quad \text{and} \quad (f \circ g)(y) = y \text{ for all } y \in B.$$

Fix $x \in A$. Then $(x, f(x)) \in f$, and so $(f(x),x) \in g$ by definition of g. That is, $g(f(x)) = x$. Similarly, for $y \in B$, we know $(y, g(y)) \in g$, so $(g(y),y) \in f$, again by definition of g. Hence $f(g(y)) = y$.

Now we prove the other two parts of the theorem.

(i) Suppose that $g_1,\ g_2\colon B \to A$ are inverses of f. Since f has an inverse, we know that $f\colon A \to B$ is bijective. We must show that $g_1(y) = g_2(y)$ for all $y \in B$.

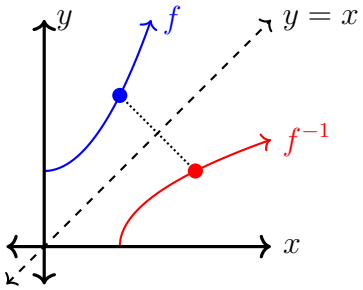
So, fix $y \in B$ and calculate $f(g_1(y)) = y$ and $f(g_2(y)) = y$ by definition of inverse. But $f\colon A \to B$ is bijective, and we have just established $f(g_1(y)) = f(g_2(y))$. Hence $g_1(y) = g_2(y)$. We now feel free to write f^{-1} for *the* inverse of f.

(ii) From the discussion after Definition 6.3.2, we know that because $f^{-1}\colon B \to A$ is the inverse of $f\colon A \to B$, the function $f\colon A \to B$ is the inverse of $f^{-1}\colon B \to A$. That is,

$f = (f^{-1})^{-1}$. Moreover, because f^{-1} has an inverse, it must be the case that $f^{-1}\colon B \to A$ is bijective. ■

6.3.6 Example.

Defining the inverse of a bijection $f\colon A \to B$ as $f^{-1} := \{(y,x) \in B \times A \mid (x,y) \in f\}$ suggests a sort of symmetry between the graphs of f and f^{-1}, if A and B are both subsets of $\mathbb{R}$. Namely, the graphs are symmetric about the line $y = x$.



6.3.7 Remark.

One of the (several) conclusions of Theorem 6.3.5 is the uniqueness of the inverse of a bijection $f\colon A \to B$. Using Definition 6.3.2, we see that if $g\colon B \to A$ is any function such that

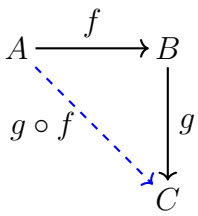
$$(g \circ f)(x) = x \text{ for all } x \in A \quad \text{and} \quad (f \circ g)(y) = y \text{ for all } y \in B,$$

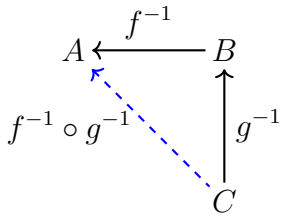
then $g = f^{-1}$.

6.3.8 Theorem.

Let $f\colon A \to B$ and $g\colon B \to C$ be bijections. Then $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Proof. We know that $g \circ f$ has an inverse in the first place because $g \circ f$ is bijective (Theorem 6.2.18) and bijections are invertible. The following diagrams suggest why the equality should be true.





Informally, if we "do f first and then g," we should undo our work by "first undoing g and then undoing f."

We abbreviate our candidate for the inverse of $g \circ f$ as $h := f^{-1} \circ g^{-1}$. By Remark 6.3.7, we will have $h = (g \circ f)^{-1}$ if

$$\big(h \circ (g \circ f)\big)(x) = x \text{ for all } x \in A \quad \text{and} \quad \big((g \circ f) \circ h\big)(y) = y \text{ for all } y \in C.$$

This amounts to nothing more than careful calculations; we do just one and leave the other as an exercise:

$$\begin{aligned}\big(h \circ (g \circ f)\big)(x) &= h((g \circ f)(x)) \\ &= h(g(f(x))) \\ &= (f^{-1} \circ g^{-1})(g(f(x))) \\ &= f^{-1}(g^{-1}(g(f(x)))) \\ &= f^{-1}(f(x)) \\ &= x.\end{aligned}$$

■

As an application of inverses, we discuss the perhaps startling fact that there exists a bijection between any two "nondegenerate" subintervals of $\mathbb{R}$.

6.3.9 Definition.

Recall the definition of a subinterval of $\mathbb{R}$ from Example 5.1.3. We say that a subinterval $I \subseteq \mathbb{R}$ is **NONDEGENERATE** *if I contains at least two distinct numbers.*

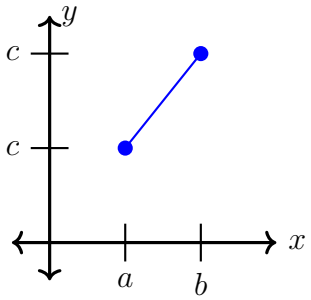
Thus intervals of the form $[a, a) = \varnothing$ and $[a, a] = \{a\}$, for $a \in \mathbb{R}$, are both degenerate. It turns out (although we will not prove this) that if I is a nondegenerate interval and c, $d \in I$ with $c \leq d$, then $[c, d] \subseteq I$. In other words, a nondegenerate interval contains all the points between any two points that it already contains.

6.3.10 Theorem.

Let I, $J \subseteq \mathbb{R}$ be nondegenerate intervals. Then there exists a bijection $f : I \to J$.

Proof. Per Example 5.1.3, there are nine possible kinds of intervals in $\mathbb{R}$. Altogether, there are a whopping 81 possible pairings, although some of these are redundant (i.e., if we can find a bijection from $[a, b]$ to (c, d), then the inverse of that bijection is a bijection from (c, d) to $[a, b]$). This, incidentally, makes for an interesting counting problem that we will study later.

In any case, we will not give a full proof. Rather, we will look at the particular case $I = [a, b]$ and $J = [c, d]$ with $a < b$ and $c < d$. Graphically, one bijection from I to J should be the line that passes through the points (a, c) and (b, d) in the plane.

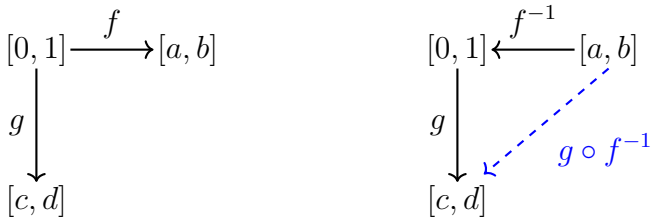


It is not at all difficult to show that this line is a bijection from $[a, b]$ to $[c, d]$, but we can also use Example 6.2.13 and function composition to get the bijection more abstractly.

Specifically, define

$$f\colon [0,1] \to [a,b]\colon t \mapsto (1-t)a + tb \quad \text{and} \quad g\colon [0,1] \to [c,d]\colon t \mapsto (1-t)c + td.$$

Then $f\colon [0,1] \to [a,b]$ and $g\colon [0,1] \to [c,d]$ are bijections. Hence $f^{-1}\colon [a,b] \to [0,1]$ is a bijection, and therefore $g \circ f^{-1}\colon [a,b] \to [c,d]$ is a bijection.



We leave the calculation of an exact formula for $g \circ f^{-1}$ as an exercise, but, as we intuited above, really $g \circ f^{-1}$ is just the line that passes through the points (a, c) and (b, d). ■

This is where we finished on Wednesday, November 3, 2021.

7. Induction and Recursion

7.1. The principle(s) of mathematical induction.

We first recall the well-ordering principle of $\mathbb{N}$ from Axiom 4.3.8. For fun, we restate it solely using quantifiers:

$$\forall A \in \mathscr{P}(\mathbb{N}) - \{\varnothing\}\ \exists \ell \in A\ \forall n \in A : \ell \leq n.$$

7.1.1 Theorem (Principle of mathematical induction).

Let $P(n)$ be a predicate with domain $\mathbb{N}$ such that the following hold.

(i) *$P(1)$ is true.*

(ii) *For each $n \in \mathbb{N}$, if $P(n)$ is true, then $P(n+1)$ is also true.*

Then $P(n)$ is true for all $n \in \mathbb{N}$.

Proof Sketch. Since $P(1)$ is true, part (ii) tells us that $P(2) \equiv P(1+1)$ is true[a] And so $P(3) \equiv P(2+1)$ is true...

[a] We write $P(2) \equiv P(1+1)$, not $P(2) = P(1+1)$. For us, statements can be equivalent ($\equiv$) but not equal ($=$).

Proof. Suppose, by way of contradiction, that $P(m)$ is false for some $m \in \mathbb{N}$. Let

$$A = \{k \in \mathbb{N} \mid P(k) \text{ is false}\} = \{k \in \mathbb{N} \mid \sim P(k)\}.$$

By definition, $A \subseteq \mathbb{N}$. Also, $m \in A$, so $A \neq \varnothing$.

The well-ordering principle then guarantees that A has a least element ℓ. That is, $\ell \in A$ and $\ell \leq k$ for all $k \in A$. In particular, since $\ell \in A$, we know that $P(\ell)$ is false. Finally, we claim that $\ell \geq 2$. The only other possibility is to have $\ell = 1$. But then $P(1) \equiv P(\ell)$ is false, which contradicts the assumption that $P(1)$ is true.

Now, observe that $\ell - 1 \notin A$, for then ℓ would not be the least element of A. Additionally, $\ell - 1 \in \mathbb{N}$ since $\ell \geq 2$. Thus the predicate $P(\ell - 1)$ is defined and has a truth value. And since $\ell - 1 \notin A$, it must be the case that $P(\ell - 1)$ is true, and so $P(\ell) \equiv P((\ell - 1) + 1)$ is also true, a contradiction. ■

7.1.2 Remark.

Induction is typically used to prove a statement of the form "$\forall n \in \mathbb{N} : P(n)$," where $P(n)$ is some predicate. There are two steps to an induction proof, and these steps correspond to verifying the hypotheses of Theorem 7.1.1.

1. *We verify that $P(1)$ is true. This is the* **base case**.

2. *We assume that $P(n)$ is true for an arbitrary $n \in \mathbb{N}$. This is the* **induction (in-**

DUCTIVE) HYPOTHESIS. *Then we must show that $P(n+1)$ is true.*

We emphasize that the induction hypothesis does not assume that $P(n)$ is true for all n and just stop there. That would be assuming what we want to prove, which is wholly unjustified and morally suspect! Rather, we are using the assumed truth of $P(n)$ to obtain the truth of $P(n+1)$.

We usually start the proof by saying something like "We will induct on n. We begin with the base case $n = 1$." After proving $P(1)$, typically we phrase the induction hypothesis by saying "Suppose that $P(n)$ is true for some $n \geq 1$." Then we go forth and show that $P(n+1)$ must be true.

We can summarize the principle of mathematical induction using quantifiers and if-then notation:

$$\big[P(1) \wedge \big(\forall n \in \mathbb{N} : P(n) \implies P(n+1)\big)\big] \implies \big(\forall n \in \mathbb{N} : P(n)\big).$$

7.1.3 Example.

Show that $0 < 2^n$ for all $n \in \mathbb{N}$.

Proof Sketch.

- Is this really worth our time? The number 2^n is just 2 multiplied by itself n times, and each factor of 2 is positive, so the whole product is positive.
- But what does "multiplied by itself n times" mean? How do we know that a product of n positive factors is positive? Did we ever prove that? Remember that one of our course goals is to make precise what has long been vague in our language.
- In any case, we want to prove a statement of the form $\forall n \in \mathbb{N} : P(n)$, where $P(n)$ is the predicate "$0 < 2^n$." This seems naturally suited to an induction argument.
- We check the base case at $n = 1$: $2^1 = 2 > 0$. So $P(1)$ is true.
- Now we want to show that if $P(n)$ is true, then $P(n+1)$ is also true. Assume that $0 < 2^n$. Why does this imply $0 < 2^{n+1}$?
- We calculate $2^{n+1} = 2(2^n)$. Since $0 < 2^n$ by the induction hypothesis, we have $2 \cdot 0 < 2(2^n)$ by properties of inequalities, and thus $0 < 2(2^n)$.

Formal Proof. We prove this by induction on n with the base case of $n = 1$. First, we see that $2^1 = 2 > 0$, so the base case is true. Next we assume that $0 < 2^n$ for some $n \geq 1$. Then $2^{n+1} = 2(2^n) > 2 \cdot 0 = 0$ by the induction hypothesis and properties of inequalities.

7.1.4 Example.

Suppose that $f\colon \mathbb{N} \to \mathbb{N}$ satisfies $f(k_1) < f(k_2)$ whenever k_1, $k_2 \in \mathbb{N}$ with $k_1 < k_2$. (Such a function is **STRICTLY INCREASING**.*) Show that $k \leq f(k)$ for all $k \in \mathbb{N}$.*

Proof Sketch.

- First, here are the ideas behind why one might think this statement is true in the first place. Since $f(k) \in \mathbb{N}$ for each k, we have $1 \leq f(k)$. In particular, $1 \leq f(1)$.
- And since f is strictly increasing, we have $f(1) < f(2)$. Thus $f(2) > 1$. But since $f(2) \in \mathbb{N}$, this strict inequality forces $f(2) \geq 2$.
- And so on.
- It looks like we sussed out the base case of $k = 1$ above. If we assume that $k \leq f(k)$, does this force $k + 1 \leq f(k + 1)$?
- Let us review the work we did to show $2 \leq f(2)$: we relied on the assumption that f is strictly increasing. This generalizes to give us $f(k + 1) > f(k) \geq k$. Hence $f(k + 1) > k$, and since $f(k + 1) \in \mathbb{N}$ this must mean $f(k + 1) \geq k + 1$.

Formal Proof. We induct on k. Since $f(k) \in \mathbb{N}$ for each k, we have $1 \leq f(k)$. In particular, $1 \leq f(1)$. Now assume that $k \leq f(k)$ for some $k \geq 1$. Since f is strictly increasing, we have $f(k + 1) > f(k) \geq k$. That is, $f(k + 1) > k$. But $f(k + 1) \in \mathbb{N}$, and so it must be the case that $f(k + 1) \geq k + 1$.

7.1.5 Remark.

It is essential that both the domain codomain of the function in Example 7.1.4 be $\mathbb{N}$. If we take $g(k) = \sqrt{k}$ for $k \in \mathbb{N}$, then $g(k) < k$ for $k \in \mathbb{N} - \{1\}$. If we take $h(x) = x/(1 + x)$ for $x \in [0, \infty)$, then $h(x) < x$ for $x > 0$. Nonetheless, both g and h are strictly increasing on their domains.

7.1.6 Example.

Show that $3 \mid (10^k - 1)$ for each $k \in \mathbb{N}$.

Proof Sketch.

- First we think about why this should be true. We compute $10^k - 1$ for some small values of k.

k	$10^k - 1$
0	0
1	9
2	99
3	999

These early results certainly appear to be divisible by 3.

• We worked out the base case $k = 1$ in the table above. We make the induction hypothesis that for some $k \geq 1$ we have $3 \mid (10^k - 1)$. Now we need to show $3 \mid (10^{k+1} - 1)$.

• Somehow we want to relate the quantity $10^{k+1} - 1$ to $10^k - 1$. Of course $10^{k+1} = 10 \cdot 10^k$, and perhaps this suggests that we try to factor 10 out of $10^{k+1} - 1$. But we cannot do this easily, because the term 1 is not divisible by 10.

• The right, and maybe unapparent, idea is to "add zero":

$$10^{k+1}-1 = 10^{k+1}-1+0 = 10^{k+1}-1+(10-10) = \big(10^{k+1}-10\big)+(10-1) = 10\big(10^k-1\big)+9.$$

• This reveals a factor of $10^k - 1$, and the induction hypothesis guarantees $3 \mid (10^k - 1)$, thus $3 \mid 10(10^k - 1)$. And the term 9 is certainly divisible by 3.

• Another way to see why we want to add zero is literally to draw the zero in: if instead of $10^{k+1} - 1 = 10(10^k) - 1$ we had $10(10^k) - 10$, then we would factor $10(10^k) - 10 = 10(10^k - 1)$, and this is divisible by 3 by the induction hypothesis. To counterbalance this 0 we need to add 9: $10(10^k - 1) + 9 = 10(10^k) - 1$.

Formal Proof. We induct on k. For $k = 1$ we have $10^1 - 1 = 9$, which is certainly divisible by 3. Now assume that $3 \mid (10^k - 1)$ for some $k \geq 1$. Then

$$10^{k+1} - 1 = 10\big(10^k - 1\big) + 9.$$

Certainly $3 \mid 9$ and $3 \mid 10\big(10^k - 1\big)$ by the induction hypothesis. It follows that $3 \mid \big[10\big(10^k - 1\big) + 9\big]$.

As an exercise, we invite the reader to prove that $m \mid \big((m+1)^n - 1\big)$ for any $m \in \mathbb{Z}$ and $n \in \mathbb{N}$.

This is where we finished on Friday, November 5, 2021.

The following result is a slight generalization of the principle of induction that allows for the base case to be an arbitrary integer (not just a natural number) and then the predicate to be defined for all integers greater than or equal to that base.

7.1.7 Lemma.

Let $m \in \mathbb{Z}$ and $S \subseteq [m, \infty) \cap \mathbb{Z}$. Suppose that if $n \in S$, then $n + 1 \in S$. Then $S = [m, \infty) \cap \mathbb{Z}$.

Proof Sketch.

• Since we assume $S \subseteq [m, \infty) \cap \mathbb{Z}$, we only need to show $[m, \infty) \cap \mathbb{Z} \subseteq S$. The elements of $[m, \infty) \cap \mathbb{Z}$ are m, $m + 1$, $m + 2, \ldots$. We know $m \in S$, and so $m + 1 \in S$, and so $m + 2 \in S, \ldots$.

• This has the feel of an argument amenable to induction, but we need to capture everything in a predicate defined on $\mathbb{N}$. What is this predicate?

• We just said that the elements of $[m, \infty) \cap \mathbb{Z}$ are m, $m + 1$, $m + 2, \ldots$, and so if we exclude m, we expect that the elements of $(m, \infty) \cap \mathbb{Z}$ are $m + 1$, $m + 2, \ldots$. These look very much like the integers $m + k$ for $k \in \mathbb{N}$, and so we expect that $(m, \infty) \cap \mathbb{Z} = \{m + k \mid k \in \mathbb{N}\}$.

• If this set equality is true, then it suffices to show that $m + k \in S$ for all $k \in \mathbb{N}$. That is, we want to prove the quantified statement $\forall k \in \mathbb{N} : m + k \in S$, and the predicate here is "$m + k \in S$." Call it $P(k)$ for future reference.

• This is exactly the sort of argument for which one uses induction. The case $k = 1$ follows from having $m \in S$ and thus $m + 1 \in S$. If we know $P(k)$, i.e., $m + k \in S$, then the hypothesis on S implies $(m + k) + 1 \in S$, and so $m + (k + 1) \in S$. That is, we have established $P(k + 1)$.

• We did not need the parentheses above at all — it would suffice to say "Since $m + k \in S$, we have $m + k + 1 \in S$" — but we feel the juggling of parentheses illustrates that having $(m + k) + 1 \in S$ arises from the hypothesis on S, while saying $m + (k + 1) \in S$ emphasizes that we have proved $P(k + 1)$.

• Last, we should show that $(m, \infty) \cap \mathbb{Z} = \{m + k \mid k \in \mathbb{N}\}$. If $n \in (m, \infty) \cap \mathbb{Z}$, then $n > m$ and $n \in \mathbb{Z}$. We want to write $n = m + k$ for some $k \in \mathbb{N}$; the only possibility is $k = n - m$. And, indeed, $k \in \mathbb{N}$ because n, $m \in \mathbb{Z}$ and thus $n - m \in \mathbb{Z}$, while $n > m$ implies $n - m > 0$. Conversely, if $k \in \mathbb{N}$, then $m + k \in \mathbb{Z}$ and $m + k > m$, thus $m + k \in (m, \infty)$.

Formal Proof. Since we assume $S \subseteq [m, \infty) \cap \mathbb{Z}$, we just have to show $[m, \infty) \cap \mathbb{Z} \subseteq S$. And since we assume $m \in S$, we really only must show $(m, \infty) \cap \mathbb{Z} \subseteq S$. We prove this in two steps.

1. First we prove the auxiliary identity $(m, \infty) \cap \mathbb{Z} = \{m + k \mid k \in \mathbb{N}\}$. First take $n \in (m, \infty) \cap \mathbb{Z}$. Then $n \in \mathbb{Z}$ and $n \in (m, \infty)$, so $n > m$. Then $n - m > 0$ and, since

both n, $m \in \mathbb{Z}$, we have $n - m \in \mathbb{Z}$. That is, $n - m \in \mathbb{N}$. Moreover, $n = n + (n - m)$, and so $n \in \{m + k \mid k \in \mathbb{N}\}$.

Conversely, let $n \in \{m + k \mid k \in \mathbb{N}\}$. Then $n = m + k$ for some $k \in \mathbb{N}$, and so $n \in \mathbb{Z}$ since m, $k \in \mathbb{Z}$. Next, since $k \in \mathbb{N}$, we have $k \geq 1$, and so $n = m + k > m$. Thus $n \in (m, \infty)$.

2. Now we prove that $\{m + k \mid k \in \mathbb{N}\} \subseteq S$. That is, we show that for all $k \in \mathbb{N}$, we have $m + k \in S$. We do this by induction on k starting with the base case of $k = 1$.

By hypothesis on S, we have $m \in S$ and thus $m + 1 \in S$. Now assume that $m + k \in S$ for some $k \geq 1$. Again by hypothesis on S, we have $(m + k) + 1 \in S$, which is to say $m + (k + 1) \in S$.

7.1.8 Theorem (Principle of mathematical induction with arbitrary base).

Let $m \in \mathbb{Z}$ and let $P(n)$ be a predicate with domain $[m, \infty) \cap \mathbb{Z}$. Suppose the following.

(i) *$P(m)$ is true.*

(ii) *For each $n \in [m, \infty) \cap \mathbb{Z}$, if $P(n)$ is true, then $P(n + 1)$ is true.*

Then $P(n)$ is true for each $n \in [m, \infty) \cap \mathbb{Z}$. In symbols,

$$\big[P(m) \wedge \big(\forall n \in [m, \infty) \cap \mathbb{Z} : P(n) \implies P(n + 1)\big)\big] \implies \big(\forall n \in [m, \infty) \cap \mathbb{Z} : P(n)\big).$$

Proof. Let $S = \{n \in [m, \infty) \cap \mathbb{Z} \mid P(n)\}$. We want to show that $S = [m, \infty) \cap \mathbb{Z}$. By hypothesis, we have $m \in S$ and we know that if $n \in S$, then $n + 1 \in S$. Thus S satisfies the conditions of Lemma 7.1.7, and so $S = [m, \infty) \cap \mathbb{Z}$. ■

7.1.9 Remark.

To prove a statement of the form "$\forall n \in [m, \infty) \cap \mathbb{Z} : P(n)$" by (generalized) induction, we need to perform two steps.

1. *We verify the base case $P(m)$.*

2. *We make the induction hypothesis that $P(n)$ is true for an arbitrary integer $n \geq m$. Then we show that $P(n + 1)$ is true.*

We usually start a "generalized" proof by induction by saying something like "We induct on n with the base case $n = m$." Then after proving $P(m)$, we introduce the induction hypothesis by saying something like "Now assume that $P(n)$ is true for an arbitrary $n \geq m$." Then we prove $P(n+1)$. The fact that $n+1 > n \geq m$ and thus $n+1 > m$ may be important.

7.1.10 Example.

Prove that if $n \geq 3$, then $2n + 1 < 2^n$.

Proof Sketch.

• If $n = 1$, then $2(1) + 1 = 3 > 2 = 2^1$. And if $n = 2$, then $2(2) + 1 = 5 > 4 = 2^2$. So this statement is definitely false for $n < 3$.

• We check the base case: $2(3) + 1 = 7$, while $2^3 = 8$, and $7 < 8$.

• We make the induction hypothesis that $2n + 1 < 2^n$ for some $n \geq 3$. We want to show $2(n + 1) + 1 < 2^{n+1}$.

• We can relate our goal to the induction hypothesis by both sides of the desired inequality to resemble more closely the sides of the induction hypothesis:

$$2(n + 1) + 1 = 2n + 2 + 1 = (2n + 1) + 2 \quad \text{and} \quad 2^{n+1} = 2(2^n).$$

So now we want to show

$$(2n + 1) + 2 < 2(2^n).$$

• Perhaps the "trick" here is to rewrite further $2(2^n) = 2^n + 2^n$. Then

$$2(n + 1) + 1 < 2^{n+1} \iff (2n + 1) + 2 < 2^n + 2^n.$$

We know $2n + 1 < 2^n$ by the induction hypothesis. And hopefully it is obvious that $2 \leq 2^n$ for all n; this can be proved by "ordinary" induction. Thus the induction hypothesis gives

$$2n + 1 < 2^n \implies (2n + 1) + 2 < 2^n + 2 \leq 2^n + 2^n \implies (2n + 1) + 2 < 2^{n+1},$$

as desired.

Formal Proof. We induct on n, starting with the base case $n = 3$. We have

$$2(3) + 1 = 7 < 8 = 2^3.$$

Now assume that $2n + 1 < 2^n$ for some $n \geq 3$. Then

$$2(n + 1) + 1 = (2n + 1) + 2 < 2^n + 2$$

by the induction hypothesis. Since $2 \leq 2^n$ for all n, we find

$$2^n + 2 \leq 2^n + 2^n = 2(2^n) = 2^{n+1}.$$

Thus $2(n + 1) + 1 < 2^{n+1}$.

7.2. Applications of induction to recursive processes.

In the broadest sense, a process (whatever "process") means is **RECURSIVE** if the kth step of that process depends on some or all of the preceding $k-1$ steps. For example, if by "process" we mean a sequence (a_k), we could define

$$a_1 := 1 \quad \text{and} \quad a_k := a_{k-1} + 1,\ k \geq 2,$$

to see that $a_2 = 2$, $a_3 = 3$, $a_4 = 4$, and more generally $a_k = k$. We might write this process piecewise via

$$a_k := \begin{cases} 1, & k = 1 \\ a_{k-1} + 1, & k \geq 2. \end{cases}$$

This specific process only called upon the $(k-1)$st step in defining the kth step, but there is no reason to be that restrictive. For example, put

$$a_k := \begin{cases} 1, & k = 1 \\ 2, & k = 2 \\ a_{k-1} - a_{k-2}, & k \geq 3 \end{cases}$$

to see that

$$a_3 = a_2 - a_1 = 1, \qquad a_4 = a_3 - a_2 = 1 - 2 = -1, \qquad \ldots$$

7.2.1 Example.

Define a sequence (a_k) in $\mathbb{R}$ recursively by

$$a_k := \begin{cases} \sqrt{6}, & k = 1 \\ \sqrt{a_{k-1} + 6}, & k \geq 2. \end{cases}$$

(i) *Show that $a_k < 3$ for all k.*

(ii) *Show that $a_k < a_{k+1}$ for all k.*

Together, the inequalities above show that (a_k) is a bounded, strictly increasing sequence, a type of sequence that has very nice behavior in calculus.

Proof Sketch.

- We calculate a few terms of this sequence:

$$a_1 = \sqrt{6}, \qquad a_2 = \sqrt{a_1 + 6} = \sqrt{\sqrt{6} + 6}, \quad \text{and} \quad a_3 = \sqrt{a_2 + 6} = \sqrt{\sqrt{\sqrt{6} + 6} + 6}.$$

- We need to know that the square root function is strictly increasing: if $0 \leq x_1 < x_2$, then $\sqrt{x_1} < \sqrt{x_2}$. Then since $6 < 9$, we have $a_1 = \sqrt{6} < \sqrt{9} = 3$. And for a_2 we use $a_1 < 3$ to find $a_1 + 6 < 9$, thus $a_2 = \sqrt{a_1 + 6} < \sqrt{9} = 3$. And so on.

• Next, since $\sqrt{6} > 0$, we have $6 < \sqrt{6} + 6$ and so $a_1 = \sqrt{6} < \sqrt{\sqrt{6} + 6} = a_2$. Thus $a_1 < a_2$. And

$$a_2 = \sqrt{\sqrt{6} + 6} < \sqrt{(\sqrt{6} + 6) + 6} = a_3.$$

• So, both inequalities appear to be true (based on a very, very small sample of the terms of the sequence). Each time, we had to use very carefully the recursive definition of the sequence, properties of the square root, and maybe something about a_k to show the desired result for a_{k+1}.

Formal Proof.

(i) We induct on k with the base case at $k = 1$. We have

$$a_1 = \sqrt{6} < \sqrt{9} = 3$$

since $6 < 9$ and the square root function is strictly increasing. Now suppose that $a_k < 3$ for some $k \geq 1$. Then

$$a_{k+1} = \sqrt{a_k + 6} < \sqrt{3 + 6} = \sqrt{9} = 3$$

by the induction hypothesis and properties of the square root.

This is where we finished on Monday, November 8, 2021.

(ii) We induct on k with the base case at $k = 1$. We calculate

$$a_2 = \sqrt{a_1 + 6} = \sqrt{\sqrt{6} + 6} > \sqrt{6}$$

since $6 < \sqrt{6} + 6$ and the square root function is strictly increasing. Now suppose that $a_k < a_{k+1}$ for some $k \geq 1$. We must show that $a_{k+1} < a_{(k+1)+1}$, i.e., that $a_{k+1} < a_{k+2}$, where $a_{k+2} = \sqrt{a_{k+1} + 6}$. And so we must show that $\sqrt{a_{k+1} + 6} > a_{k+1}$. But

$$a_{k+1} = \sqrt{a_k + 6} < \sqrt{a_{k+1} + 6}$$

by the induction hypothesis and properties of the square root.

7.2.1. Recursion and finite sums.

One of the most important recursive processes in arithmetic is addition.

7.2.2 Definition.

Let (a_k) be a sequence in $\mathbb{R}$. For $n \in \mathbb{N}$, we define

$$\sum_{k=1}^{n} a_k := \begin{cases} a_1, & n = 1 \\ \left(\sum_{k=1}^{n-1} a_k \right) + a_n, & n \geq 2. \end{cases}$$

We also write $\sum_{k=1}^{n} a_k$ in lieu of the larger notation above.

We might euphemistically write

$$\sum_{k=1}^{n} a_k = a_1 + \cdots + a_n,$$

but the symbols $+\cdots+$ are meaningless without the rigor of Definition 7.2.2. The **INDEX** k is just a "dummy variable" in the sense that

$$\sum_{k=1}^{n} a_k = \sum_{j=1}^{n} a_j = \sum_{\ell=1}^{n} a_\ell.$$

7.2.3 Example.

Let $a_k = k^2$ for $k \in \mathbb{N}$. Then

$$\sum_{k=1}^{3} a_k = \left(\sum_{k=1}^{2} a_k \right) + a_3 = \left(\sum_{k=1}^{1} a_k \right) + a_2 + a_3 = a_1 + a_2 + a_3 = 1^2 + 2^2 + 3^2 = 1 + 4 + 9 = 14.$$

This is where we finished on Wednesday, November 10, 2021 (Section 53).

7.2.4 Example.

Let $n \in \mathbb{N}$. Prove the identity

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Proof Sketch.

• First, observe that since $n \in \mathbb{N}$, either $2 \mid n$ or $2 \mid (n+1)$, so $n(n+1)/2 \in \mathbb{N}$. (It should not be the case that a bunch of integers adds up to a noninteger!)

• To convince ourselves that this identity is legitimate, it may be helpful to do a few small calculations. These are, of course, not proofs, but they should reassure us:

$$\sum_{k=1}^{1} k = 1 \quad \text{and} \quad \frac{1(1+1)}{2} = \frac{2}{2} = 1,$$

$$\sum_{k=1}^{2} k = 1+2 = 3 \quad \text{and} \quad \frac{2(2+1)}{2} = \frac{2(3)}{2} = 3,$$

and

$$\sum_{k=1}^{3} k = 1+2+3 = 6 \quad \text{and} \quad \frac{3(3+1)}{2} = \frac{12}{2} = 6.$$

• We want to prove a statement of the form "$\forall n \in \mathbb{N} : P(n)$," where $P(n)$ is the predicate "$\sum_{k=1}^{n} k = n(n+1)/2$." This calls for induction.

• We did the base case $n = 1$ above in our motivating calculations. Assume that the statement is true for an arbitrary $n \geq 1$. That is, assume that

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

for an arbitrary $n \geq 1$.

• Now we must show

$$\sum_{k=1}^{n+1} k = \frac{(n+1)[(n+1)+1]}{2}.$$

First, we rewrite

$$\frac{(n+1)[(n+1)+1]}{2}. = \frac{(n+1)(n+2)}{2}.$$

Next, we use the recursive definition of the sum to find

$$\sum_{k=1}^{n+1} k = \left(\sum_{k=1}^{n} k\right) + (n+1).$$

We use the induction hypothesis to rewrite this further:

$$\left(\sum_{k=1}^{n} k\right) + (n+1) = \frac{n(n+1)}{2} + (n+1).$$

We get a common denominator:

$$\frac{n(n+1)}{2} + (n+1) = \frac{n(n+1)}{2} + \frac{2(n+1)}{2} = \frac{n(n+1)+2(n+1)}{2} = \frac{(n+1)(n+2)}{2},$$

and this is what we wanted.

This is where we finished on Wednesday, November 10, 2021 (Section 54).

Formal Proof. We induct on n. For $n = 1$ we have, by definition of the sum,

$$\sum_{k=1}^{1} k = 1,$$

and we calculate

$$\frac{1(1+1)}{2} = 1,$$

thus the identity

$$\sum_{k=1}^{1} k = 1, = \frac{1(1+1)}{2}$$

is true.

Now let $n \geq 1$ and assume that

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Then the definition of the sum gives

$$\sum_{k=1}^{n+1} k = \sum_{k=1}^{n} k + (n+1),$$

so the induction hypothesis and arithmetic imply

$$\sum_{k=1}^{n+1} k = \frac{n(n+1)}{2} + (n+1) = \frac{(n+1)(n+2)}{2} = \frac{(n+1)[(n+1)+1]}{2}.$$

More generally, we can define sums over arbitrary "ranges."

7.2.5 Definition.

Let $N \in \mathbb{Z}$ and let $f\colon [N, \infty) \cap \mathbb{Z} \to \mathbb{R}$. For m, $n \in \mathbb{Z}$ with $N \leq m \leq n$, we define

$$\sum_{k=m}^{n} f(k) := \begin{cases} f(n), & n = m \\ \left(\sum_{k=m}^{n-1} f(k)\right) + f(n), & n > m. \end{cases}$$

If $n < m$, it will be convenient to set

$$\sum_{k=n}^{m} f(k) := 0.$$

The following two sums appear in many diverse applications in mathematics.

7.2.6 Example (Geometric partial sum).

Let $r \in \mathbb{R} - \{1\}$ and $n \in \mathbb{N} \cup \{0\}$. Then

$$\sum_{k=0}^{n} r^k = \frac{1 - r^{n+1}}{1 - r}. \tag{7.2.1}$$

Proof. We induct with $n = 0$ as the base case. If $n = 0$, then

$$\sum_{k=0}^{0} r^k = r^0 = 1 \quad \text{and} \quad \frac{1 - r^{0+1}}{1 - r} = \frac{1 - r}{1 - r} = 1,$$

so $\sum_{k=0}^{0} r^k = (1 - r^{0+1})/(1 - r)$. Here, incidentally, we have used $r \neq 1$ so that the second set of equalities made sense.

Assume the result to be true for some $n \geq 0$. Then

$$\sum_{k=0}^{n+1} r^k = \left(\sum_{k=0}^{n} r^k \right) + r^{n+1} \text{ by the definition of the sum}$$

$$= \frac{1 - r^{n+1}}{1 - r} + r^{n+1} \text{ by the induction hypothesis}$$

$$= \frac{1 - r^{n+1} + (1 - r)r^{n+1}}{1 - r} \text{ after adding}$$

$$= \frac{1 - r^{n+1} + r^{n+1} - r^{n+2}}{1 - r}$$

$$= \frac{1 - r^{n+2}}{1 - r}$$

$$= \frac{1 - r^{(n+1)+1}}{1 - r} \text{ since } n + 2 = (n + 1) + 1.$$

■

7.2.7 Remark.

Certainly the formula (7.2.1) for the geometric partial sum does not make sense when $r = 1$, as it would then involve division by 0. Instead, when $r = 1$, we have

$$\sum_{k=0}^{n} r^k = \sum_{k=0}^{n} 1^k = \sum_{k=0}^{n} 1 = (n - 0 + 1) \cdot 1 = n + 1.$$

7.2.8 Example (Telescoping).

Let (a_k) be a sequence of real numbers and let m, $n \in \mathbb{N}$ with $m \leq n$. Then

$$\sum_{k=m}^{n}(a_{k+1} - a_k) = a_{n+1} - a_m.$$

Proof Sketch.

- Here is what is going on. For simplicity, suppose $m = 1$ and $n = 3$. Then

$$\sum_{k=1}^{3}(a_{k+1}-a_k) = (a_{1+1}-a_1)+(a_{2+1}-a_2)+(a_{3+1}-a_3) = (a_2-a_1)+(a_3-a_2)+(a_4-a_3)$$
$$= a_4 - a_1.$$

More generally, after each addition certain terms cancel in a special way, and the sum collapses to the "outermost" numbers (like a telescope of yore expanding and contracting):

$$\sum_{k=m}^{n}(a_{k+1} - a_k) = (a_{m+1} - a_m) + (a_{m+2} - a_{m+1}) + (a_{m+3} - a_{m+2}) + \cdots + (a_{m+n} - a_{m+n-1}).$$

- Where should the induction fall? There are two natural numbers in play here: m and n. It might help to rephrase the statement with quantifiers in a precise order:

$$\forall n \in \mathbb{N} \forall m \in \mathbb{N} : \left(m \leq n \Longrightarrow \sum_{k=m}^{n}(a_{k+1} - a_k) = a_{n+1} - a_m \right).$$

Thus we are trying to prove $\forall n \in \mathbb{N} : P(n)$, where $P(n) :\equiv \forall m \in \mathbb{N} : \big(m \leq n \Longrightarrow \sum_{k=m}^{n}(a_{k+1} - a_k) = a_{n+1} - a_m\big)$. We would prove this by induction on n with the base case of $n = m$.

- But we could also view this statement as choosing m first and then showing that if $n \geq m$, then $\sum_{k=m}^{n}(a_{k+1} - a_k) = a_{n+1} - a_m$. That is, we would be trying to prove $\forall n \in [m, \infty) \cap \mathbb{Z} : Q(n)$, where $Q(n) :\equiv \sum_{k=m}^{n}(a_{k+1} - a_k) = a_{n+1} - a_m$.

Formal Proof. ***Method 1.*** We will prove

$$\forall n \in \mathbb{N} : P(n), \qquad P(n) \equiv \forall m \in \mathbb{N} : \left(m \leq n \Longrightarrow \sum_{k=m}^{n}(a_{k+1} - a_k) = a_{n+1} - a_m \right).$$

We induct with $n = 1$ as the base case. If $n = 1$ and $m \in \mathbb{N}$ with $m \leq n$, then $m \in \mathbb{N}$

and $m \leq 1$, so really $m = 1$. Then we have

$$\sum_{k=m}^{n}(a_{k+1} - a_k) = \sum_{k=1}^{1}(a_{k+1} - a_k) = a_{1+1} - a_1 = a_{n+1} - a_m.$$

Now suppose that for some $n \geq 1$, it is the case that if $m \in \mathbb{N}$ and $m \leq n$, then $\sum_{k=m}^{n}(a_{k+1} - a_k) = a_{n+1} - a_m$. We will show that if $m \in \mathbb{N}$ and $m \leq n+1$, then $\sum_{k=m}^{n+1}(a_{k+1} - a_k) = a_{(n+1)+1} - a_m$. We distinguish two cases: $1 \leq m \leq n$ and $m = n+1$. In the latter, we have

$$\sum_{k=n+1}^{n+1}(a_{k+1} - a_k) = a_{(n+1)+1} - a_{n+1} = a_{(n+1)+1} - a_m,$$

as desired. Otherwise, if $1 \leq m \leq n$, we have

$$\sum_{k=m}^{n+1}(a_{k+1} - a_k) = \sum_{k=m}^{n}(a_{k+1} - a_k) + (a_{(n+1)+1} - a_{n+1}) \text{ by the definition of the sum}$$

$$= (a_{n+1} - a_m) + (a_{(n+1)+1} - a_{n+1}) \text{ by the induction hypothesis}$$

$$= a_{(n+1)+1} - a_m.$$

Method 2. We will prove

$$\forall n \in [m, \infty) \cap \mathbb{Z} : Q(n), \qquad Q(n) \equiv \sum_{k=m}^{n}(a_{k+1} - a_k) = a_{n+1} - a_m.$$

We induct with $n = m$ as the base case. First,

$$\sum_{k=m}^{m}(a_{k+1} - a_k) = \sum_{k=n}^{n}(a_{k+1} - a_k) = a_{n+1} - a_n = a_{n+1} - a_m.$$

Now suppose that the result is true for some $n \geq m$. Then

$$\sum_{k=m}^{n+1}(a_{k+1} - a_k) = \left(\sum_{k=m}^{n}(a_{k+1} - a_k)\right) + (a_{(n+1)+1} - a_{n+1}) \text{ by the definition of the sum}$$

$$= (a_{n+1} - a_m) + (a_{(n+1)+1} - a_{n+1}) \text{ by the induction hypothesis}$$

$$= a_{(n+1)+1} - a_m.$$

7.2.2. Recursion and finite products.

Much like a finite sum, we recursively define the product of finitely many numbers.

7.2.9 Definition.

Let (a_k) be a sequence and m, $n \in \mathbb{N}$ with $m \leq n$. We define

$$\prod_{k=m}^{n} a_k := \begin{cases} a_m, & n = m \\ \left(\prod_{k=m}^{n-1} a_k\right) a_n, & n > m. \end{cases}$$

We may also write $\prod_{k=m}^{n} a_k$. If $n < m$, it will be convenient to define

$$\prod_{k=n}^{m} a_k := 1.$$

7.2.10 Example.

$$\prod_{k=1}^{4} k = \left(\prod_{k=1}^{3} k\right) 4 = \left(\prod_{k=1}^{2} k\right) 3 \cdot 4 = \left(\prod_{k=1}^{1} k\right) 2 \cdot 3 \cdot 4 = 1 \cdot 2 \cdot 3 \cdot 4 = 24.$$

This is where we finished on Monday, November 15, 2021 (Section 53).

The product of all the integers from 1 to a fixed positive integer appears with (un)surprising frequency in mathematics.

7.2.11 Definition.

Let $n \in \mathbb{N} \cup \{0\}$. The number $n!$, pronounced "n **FACTORIAL**,*" is*

$$n! := \begin{cases} 1, & n = 0 \\ \prod_{k=1}^{n} k, & n \geq 1. \end{cases}$$

It turns out that 0! appears often enough in formulas and definitions that we want to give meaning to 0!, and the "natural" meaning from those appearances is $0! = 1$. Also, we quickly calculate the useful identity

$$(n+1)! = \prod_{k=1}^{n+1} k = \left(\prod_{k=1}^{n} k\right)(n+1) = (n+1)n!.$$

7.2.12 Example.

Show that $2^{k-1} \leq k!$ *for all* $k \in \mathbb{N}$.

Proof Sketch. We can check the validity of this inequality for some small k and see along the way why the power $k-1$ is important:

$$2^{1-1} = 2^0 = 1 = 1!,$$

$$2^{2-1} = 2^1 = 2 = 2 \cdot 1 = 2!,$$

and

$$2^{3-1} = 2^2 = 4 \quad \text{while} \quad 3! = 3 \cdot 2 = 6 > 4.$$

More generally, we can look at a factorial for a "large" value of k and see how replacing the factors in that product by 2 give us the inequality:

$$5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \geq 2 \cdot 2 \cdot 2 \cdot 2 \cdot 1 = 2^4 = 2^{5-1}.$$

Formal Proof. We induct on k starting with the base case of $k = 1$, at which we have

$$2^{1-1} = 2^0 = 1 = 1!.$$

Next, assume that for some $k \geq 1$ we have $2^{k-1} \leq k!$. We must show $2^{(k+1)-1} \leq (k+1)!$, equivalently, $2^k \leq (k+1)k!$. The induction hypothesis tells us $2^k = 2(2^{k-1}) \leq 2k!$, and since $k \in \mathbb{N}$ we have $k+1 \geq 2$. Thus $2k! \leq (k+1)k!$, so we conclude $2^k \leq (k+1)k!$, as desired.

This is where we finished on Monday, November 15, 2021 (Section 54).

7.3. Recursion and set-theoretic algebra.

7.3.1. Generalized unions and intersections.

So far, we have managed to survive by taking the union or intersection of only two or three sets at a time, e.g., we have (probably) calculated $A \cup B$ and then $(A \cup B) \cup C$, and found that equal to $A \cup (B \cup C)$; thus we have (probably) felt justified in referring to the resulting set as $A \cup B \cup C$ without any parentheses at all. But we have not really explored what it should mean to take the union or intersection of multiple sets in an efficient way. The following definition provides such a mechanism.

7.3.1 Definition.

Let A *be a set and let* (B_k) *be a sequence of subsets of* A*, i.e.,* (B_k) *is a sequence in* $\mathscr{P}(A)$.

The **UNION OF** $B_1, \ldots, B_n$ *is the set*

$$\bigcup_{k=1}^{n} B_k := \begin{cases} B_1, \ n = 1 \\ \left(\bigcup_{k=1}^{n-1} B_k \right) \cup B_n, \ n \geq 2 \end{cases}$$

and the **INTERSECTION OF** $B_1, \ldots, B_n$ *is the set*

$$\bigcap_{k=1}^{n} B_k := \begin{cases} B_1, \ n = 1 \\ \left(\bigcap_{k=1}^{n-1} B_k \right) \cap B_n, \ n \geq 2. \end{cases}$$

We also write these sets as $\cup_{k=1}^{n} B_k$ *and* $\cap_{k=1}^{n} B_k$.

In particular, we have the expected identities

$$\bigcup_{k=1}^{2} B_k = \left(\bigcup_{k=1}^{1} B_k \right) \cup B_2 = B_1 \cup B_2 \quad \text{and} \quad \bigcap_{k=1}^{2} B_k = \left(\bigcap_{k=1}^{1} B_k \right) \cup B_2 = B_1 \cap B_2.$$

7.3.2 Example.

Simplify each of the unions or intersections below as much as possible.

(i) $\cup_{k=1}^{3} \{k\}$

(ii) $\cap_{k=1}^{4} \{1, k\}$

Solution. We emphasize the following calculations in horrible, pedantic detail to distinguish the recursive calculations of the union/intersection from the "ordinary" union/intersection of two sets. And after this exercise in nit-picking, we will never bother to distinguish them again.

(i) We have

$$\bigcup_{k=1}^{3} \{k\} = \left(\bigcup_{k=1}^{2} \{k\} \right) \cup \{3\} \text{ by the recursive definition of the union}$$

$$= \left[\left(\bigcup_{k=1}^{1} \{k\} \right] \cup \{2\} \right) \cup \{3\} \text{ again by the recursive definition of the union}$$

$$= \big(\{1\} \cup \{2\} \big) \cup \{3\} \text{ once more by the recursive definition of the union}$$

$= \{1,2\} \cup \{3\}$ by definition of the union of two sets

$= \{1,2,3\}$ again by definition of the union of two sets.

(ii) We have

$$\bigcap_{k=1}^{4}\{1,k\} = \left(\bigcap_{k=1}^{3}\{1,k\}\right) \cap \{1,4\}$$

$$= \left[\left(\bigcap_{k=1}^{2}\{1,k\}\right) \cap \{1,3\}\right] \cap \{1,4\} \text{ by the recursive definition of the intersection}$$

$$= \left(\bigcap_{k=1}^{2}\{1,k\}\right) \cap \big(\{1,3\} \cap \{1,4\}\big) \text{ by associativity of intersection}^{34}$$

$$= \left(\bigcap_{k=1}^{2}\{1,k\}\right) \cap \{1\} \text{ by definition of the intersection of two sets}$$

$$= \left[\left(\bigcap_{k=1}^{1}\{1,k\}\right) \cap \{1,2\}\right] \cap \{1\} \text{ by the recursive definition of the intersection}$$

$$= \big(\{1,1\} \cap \{1,2\}\big) \cap \{1\} \text{ yet again by the recursive definition of the intersection}$$

$$= \{1\} \cap \{1\} \text{ by the ordinary definition of intersection}$$

$$= \{1\}.$$ ▲

While the recursive definition of a finite union or intersection melds well stylistically with our current studies, it is far from the most convenient definition. Here is, frankly, a better one. Recall that, as usual, $\mathcal{F}_n = [1,n] \cap \mathbb{N}$.

7.3.3 Lemma.

Let (B_k) be a sequence of subsets of A. Then

$$\bigcup_{k=1}^{n} B_k = \{x \in A \mid \exists k \in \mathcal{F}_n : x \in B_k\}$$

[34] For any sets A, B, and C we have $(A \cap B) \cap C = A \cap (B \cap C)$.

and

$$\bigcap_{k=1}^{n} B_k = \{x \in A \mid \forall k \in \mathcal{F}_n : x \in B_k\}.$$

Proof. We give the proof just for the union and leave the proof for the intersection, which is mostly a matter of adjusting quantifiers, as an exercise.

We induct on n. If $n = 1$, then

$$\bigcup_{k=1}^{1} B_k = B_1 = \{x \in A \mid x \in B_1\} = \{x \in A \mid \exists k \in \{1\} : x \in B_k\}.$$

Assume that the identity is true for some $n \geq 1$. Then the recursive definition of the union gives

$$\bigcup_{k=1}^{n+1} B_k = \left(\bigcup_{k=1}^{n} B_k\right) \cup B_{n+1} = \{x \in A \mid \exists k \in \mathcal{F}_n : x \in B_k\} \cup B_{n+1} =: U_{n+1}.$$

We need to show that $U_{n+1} = V_{n+1}$, where

$$V_{n+1} := \{x \in A \mid \exists k \in \mathcal{F}_{n+1} : x \in B_k\}.$$

We are abbreviating with U and V just for convenience.

If $y \in U_{n+1}$, then either $y \in B_{n+1}$ or there is $k \in \mathcal{F}_n$ such that $y \in B_k$. In either case, we find that $y \in B_\ell$ for some $\ell \in \mathcal{F}_{n+1}$. Thus $y \in V_{n+1}$. Conversely, if $y \in V_{n+1}$, then $y \in B_k$ for some $k \in \mathcal{F}_{n+1}$. Then either $k \in \mathcal{F}_n$, in which case $y \in B_k$ for $k \in \mathcal{F}_n$, or $k = n + 1$, in which case $y \in B_{n+1}$. In either case, we find that $y \in U_{n+1}$. ■

7.3.4 Example.

Simplify each of the unions or intersections below as much as possible. Consider each set as a subset of $\mathbb{R}$.

(i) $\cup_{k=1}^{3}[0, k]$

(ii) $\cap_{k=1}^{10}(0, 10]$

Solution. (i) By definition,

$$\bigcup_{k=1}^{3}[0, k] = \{x \in \mathbb{R} \mid \exists k \in \{1, 2, 3\} : x \in [0, k]\}.$$

By intuitively expanding

$$\bigcup_{k=1}^{3}[0, k] = [0, 1] \cup [0, 2] \cup [0, 3],$$

we expect $\cup_{k=1}^{3}[0, k] = [0, 3]$.

Let us prove this equality. First, if $x \in \cup_{k=1}^{3}[0,k]$, then $x \in [0,k]$ for some $k \in \{1,2,3\}$. Thus $0 \le x \le k$ for some $k \in \{1,2,3\}$, and so $0 \le x \le 3$. Hence $x \in [0,3]$.

Conversely, if $x \in [0,3]$, then there are three mutually exclusive possibilities for x:

$$\big(0 \le x \le 1\big) \vee \big(1 < x \le 2\big) \vee \big(2 < x \le 3\big).$$

In the first case, we have $x \in [0,1]$; in the second we have $x \in (1,2] \subseteq [1,2]$, and in the third we have $x \in (2,3] \subseteq [2,3]$.

(ii) By definition,

$$\bigcap_{k=1}^{10}(0,k] = \{x \in \mathbb{R} \mid \forall k \in [1,10] \cap \mathbb{N} : x \in (0,k]\}\,.$$

Intuitively,

$$\bigcap_{k=1}^{10}(0,k] = (0,1] \cap (0,2] \cap (0,3] \cap \cdots \cap (0,9] \cap (0,10].$$

It looks like $(0,1]$ is contained in each of these sets, and so we expect $\cap_{k=1}^{10}(0,k] = (0,1]$.

We prove this. First let $x \in \cap_{k=1}^{10}(0,k]$. Then in particular $x \in (0,1]$.

Conversely, let $x \in (0,1]$. Hence $0 < x \le 1$. We need to show that $x \in (0,k]$ for each $k \in [1,10] \cap \mathbb{N}$. But if $k \in [1,10] \cap \mathbb{N}$, then $k \ge 1$, and so $0 < x \le 1 \le k$. Thus $x \in (0,k]$. ▲

7.3.2. Generalized Cartesian products.

Let A_1, A_2, and A_3 be sets. Then

$$(A_1 \times A_2) \times A_3 = \{(x,y) \mid x \in A_1, y \in A_2\} \times A_3 = \big\{\big((x,y),z\big) \;\big|\; x \in A_1, y \in A_2, z \in A_3\big\}\,.$$

And

$$A_1 \times (A_2 \times A_3) = A_1 \times \{(y,z) \mid y \in A_2, z \in A_3\} = \big\{\big(x,(y,z)\big) \;\big|\; x \in A_1, y \in A_2, z \in A_3\big\}\,.$$

Consequently the elements of $(A_1 \times A_2) \times A_3$ and $A_1 \times (A_2 \times A_3)$ are not the same, since the ordered pairs $\big((x,y),z\big)$ and $\big(x,(y,z)\big)$ are technically different objects[35]. But both ordered pairs capture the idea that $x \in A_1$ comes first, then $y \in A_2$, and then $z \in A_3$. If we want to "order" the elements x, y, and z so that x comes first, followed by y, and last by z, why not just consider the set of pairs $\{(1,x),(2,y),(3,z)\}$? This set of pairs is really a function from $f\colon \{1,2,3\} \to A_1 \cup A_2 \cup A_3$ satisfying $f(1) \in A_1$, $f(2) \in A_2$, and $f(3) \in A_3$.

We have seen an idea like this before. In the following we recall the notations of Definitions 6.1.4 and 6.1.23 and the work of Example 6.1.25.

[35] A truly horrible exercise is to express these ordered pairs without any parentheses. Here is a start:

$$\big((x,y),z\big) = \{(x,y)\},\{(x,y),z\}\} = \Big\{\{\{x\},\{x,y\}\},\{\{(x,y)\},\{(x,y),z\}\Big\} = \cdots\,.$$

7.3.5 Example.

Let A and B be sets and let

$$\mathcal{X} := \left\{ f \in (A \cup B)^{\mathcal{F}_2} \;\middle|\; f(1) \in A, f(2) \in B \right\}.$$

Define

$$G\colon A \times B \to \mathcal{X}\colon (x, y) \mapsto \{(1, x), (2, y)\}.$$

Then G is a bijection.

Inspired by this, we make the following definition.

7.3.6 Definition.

Let (A_k) be a sequence of sets. For $n \in \mathbb{N}$, the **Cartesian product of the sets** *$A_1, \ldots, A_n$ is the set*

$$\prod_{k=1}^{n} A_k := \left\{ f \in \left(\bigcup_{k=1}^{n} A_k \right)^{\mathcal{F}_n} \;\middle|\; \forall k \in \mathcal{F}_n : f(k) \in A_k \right\}.$$

We may also write this set as $\prod_{k=1}^{n} A_k$.

This is where we finished on Wednesday, November 17, 2021.

In words, each element of $\prod_{k=1}^{n} A_k$ is a function $f\colon \mathcal{F}_n \to \cup_{k=1}^{n} A_k$ such that $f(k) \in A_k$ for each k. We typically and euphemistically write such a function as an **ordered n-tuple**:

$$f = \big(f(1), \ldots, f(n)\big) = \big(x_1, \ldots, x_n\big),$$

if we define $x_k := f(k)$. Thus, for example, the true definition of the symbol $(4, 5, 6, 7)$, which is an **ordered triple**, is the function $f\colon \{1, 2, , 43\} \to \mathbb{N}$ such that $f(1) = 4$, $f(2) = 5$, $f(3) = 6$, and $f(4) = 7$. (Here the codomain of f is irrelevant; we could replace $\mathbb{N}$ with $\mathbb{Z}$ or $\mathbb{R}$ or $\{4, 5, 6, 7\}$ or really any set of which $\{4, 5, 6, 7\}$ is a subset.) In other words,

$$(4, 5, 6, 7) = \{(1, 4), (2, 5), (3, 6), (4, 7)\}.$$

Of course, thinking of an ordered triple as a set of three ordered pairs is unnecessary and baroque, and no one ever actually does this. Far more important is that an ordered n-tuple is completely determined by its components:

$$(x_1, \ldots, x_n) = (y_1, \ldots, y_n) \iff x_1 = y_1, \ldots, x_n = y_n.$$

The $\prod$ notation is not always used when dealing with the product of a "small" number of sets, and we often write

$$\prod_{k=1}^{3} A_k = A_1 \times A_2 \times A_3.$$

But doing so illustrates another drawback of this definition. We have $\prod_{k=1}^{1} A_k \neq A_1$ and $\prod_{k=1}^{2} A_k \neq A_1 \times A_2$. Instead,

$$\prod_{k=1}^{1} A_k = \{(1,x) \mid x \in A_1\} = \{1\} \times A_1 \quad \text{and} \quad \prod_{k=1}^{2} A_k = \big\{\big\{(1,x),(2,y)\big\} \bigm| (x,y) \in A_1 \times A_2\big\}. \tag{7.3.1}$$

However, if we agree from now on to identify the symbol $(x_1, \ldots, x_n)$ with the function from $\mathcal{F}_n$ to $\cup_{k=1}^{n} A_k$ whose value at k is x_k, then the differences disappear.

Although we have $\cup_{k=1}^{n+1} A_k = \big(\cup_{k=1}^{n} A_k\big) \cup A_{n+1}$ and $\cap_{k=1}^{n+1} A_k = \big(\cap_{k=1}^{n} A_k\big) \cap A_{n+1}$, the sets $\prod_{k=1}^{n+1} A_k$ and $\big(\prod_{k=1}^{n} A_k\big) \times A_{n+1}$ are different. The former is a set of functions and the latter is a set of ordered pairs whose first component is a function and whose second component is an element of A_{n+1}. But, again, under the lens of bijections, the differences disappear. Specifically, any function $f \in \prod_{k=1}^{n+1} A_k$ is uniquely determined by its restriction $f\big|_{\mathcal{F}_n}$ to $\mathcal{F}_n$ and its value $f(n+1)$ at $n+1$, and so we have the following result.

7.3.7 Lemma.

Let (A_k) be a sequence of sets. The map

$$G\colon \prod_{k=1}^{n+1} A_k \to \left(\prod_{k=1}^{n} A_k\right) \times A_{n+1}\colon f \mapsto \big(f\big|_{\mathcal{F}_n}, f(n+1)\big)$$

is a bijection.

8. Counting

> "Fiver?" said the other rabbit. "Why's he called that?"
>
> "Five in the litter, you know: he was the last — and the smallest."
>
> Rabbits can count up to four. Any number above four is *hrair* — "a lot," or "a thousand." Thus they say *U Hrair* — "The Thousand" — to mean, collectively, all the enemies (or *elil*, as they call them) of rabbits — fox, stoat, weasel, cat, owl, man, etc. There were probably more than five rabbits in the litter when Fiver was born, but his name, *Hrairoo*, means "Little Thousand" — i.e., the little one of a lot or, as they say of pigs, "the runt."
>
> —Richard Adams, *Watership Down*

The ability to count certainly separates us from the rabbits. But how do we count? And why do we know there are four dots below?

We probably counted the dots by assigning to each dot one and only one of the numbers 1, 2, 3 and 4, perhaps (though not necessarily), proceeding from left to right.

In other words, we constructed a bijection from the set $\{1, 2, 3, 4\}$ to the set of dots.

It will be the concept of bijection that captures precisely the notion of how many elements a set has. And if we are only interested in how many elements a set has, not what those elements are or how they otherwise interact, at the level of *finite* sets, the only sets that will matter will be $\mathcal{F}_n = \{1, \ldots, n\}$, the prototypical set of n elements. But if we broaden our horizons to *infinite* sets, our world will become much more complicated.

8.1. Finite sets.

8.1.1. Basic definitions.

Recall that for $n \in \mathbb{N}$ we write

$$\mathcal{F}_n := \mathbb{N} \cap [1, n] = \{k \in \mathbb{N} \mid 1 \leq k \leq n\}.$$

We formalize the notion of "counting a set A by pairing elements with numbers 1 through n via a bijection from $\mathcal{F}_n$ to A."

8.1.1 Definition.

A set A is **FINITE** *if $A = \varnothing$ or if there exist $n \in \mathbb{N}$ and a bijection $f : \mathcal{F}_n \to A$. A set A is* **INFINITE** *if it is not finite.*

Note that A is infinite if

$$\forall n \in \mathbb{N}\ \forall f\colon \mathcal{F}_n \to A : f \text{ is not bijective.}$$

It is possible to replace "f is not bijective" above with "f is not surjective," although that will not be relevant for some time. Also, if there exist $n \in \mathbb{N}$ and a bijection $f\colon \mathcal{F}_n \to A$, then $A \neq \varnothing$, since $f(1) \in A$. Thus the "or" in Definition 8.1.1 really is exclusive.

Although we will not discuss the details, the following result (which we prove as Corollary C.0.5 in Appendix C) assures us that the n in Definition 8.1.1 really is unique. Informally, if we have counted a set (correctly) and we count it again, the number of elements in the set will not change.

8.1.2 Lemma.

Let A be a set and suppose there exist $m, n \in \mathbb{N}$ and bijections $f\colon \mathcal{F}_m \to A$ and $g\colon \mathcal{F}_n \to A$. Then $m = n$.

The following numbers are therefore unambiguously defined.

8.1.3 Definition.

Let A be a finite set.

(i) *The* **CARDINALITY** *of A is $|A| = 0$ if $A = \varnothing$.*

(ii) *If $A \neq \varnothing$, the* **CARDINALITY** *of A is $|A| = n$, where $n \in \mathbb{N}$ is the (necessarily unique) natural number for which there exists a bijection from $\mathcal{F}_n$ to A.*

(iii) *If B is another finite set, then A and B* **HAVE THE SAME CARDINALITY** *if $|A| = |B|$.*

In symbols,

$$|A| = \begin{cases} 0, & A = \varnothing \\ n, & A \neq \varnothing \wedge \exists n \in \mathbb{N}\ \exists f\colon \mathcal{F}_n \to A \text{ bijective.} \end{cases}$$

In particular, if A is finite, then $|A| \in \mathbb{N}$ if and only if $A \neq \varnothing$.

The concept of cardinality is precisely the notion of "how many elements" a finite set has. If $|A| = n$ and $f\colon \mathcal{F}_n \to A$ is bijective, then we can index A as

$$A = \{f(k) \mid k \in \mathcal{F}_n\} =: \{f(k)\}_{k=1}^{n}, \quad \text{where} \quad 1 \leq j < \ell \leq n \Longrightarrow f(j) \neq f(\ell).$$

We will often say something like "Let A be a finite set and write $A = \{a_k\}_{k=1}^{n}$." In doing so we implicitly assume that the a_k are all distinct, i.e., $a_j \neq a_\ell$ for $j \neq \ell$.

The proofs of the next two theorems follow directly from the definition of cardinality and the results in Appendix C. The first set of results should feel obvious — for example, any subset of a finite set is finite and is no larger than the whole set — but the proofs of the theorems in Appendix C on which these results depend are deliciously challenging.

8.1.4 Theorem.

Let A be a finite set and $B \subseteq A$.

(i) *Then B is also finite and $|B| \leq |A|$.*

(ii) *$|B| = |A|$ if and only if $B = A$.*

The second theorem tells us about how nicely we can pair elements of sets of different sizes. One should be convinced of the truth of this theorem just by drawing a couple of arrow diagrams. Again, the proofs are totally nontrivial.

8.1.5 Theorem.

Let A and B be finite sets.

(i) *$|A| = |B|$ if and only if there is a bijection from A to B.*

(ii) *If $|A| < |B|$, then there is no injection from B to A and no surjection from A to B.*

(iii) *If $|A| = |B|$, then a function $f\colon A \to B$ is injective if and only if f is surjective.*

We will prove several results that detail the interaction of cardinality and set-theoretic algebra and then explore some applications of these results to counting both routine and esoteric quantities.

8.1.2. Unions of finite sets.

If we form a set by combining two (or more) sets that have no element in common, then the number of elements in our new set should be the sum of the elements in each "component" set. Here is a precise statement of that idea.

8.1.6 Theorem (Addition rule).

Let A and B be finite sets with $A \cap B = \varnothing$. Then

$$|A \cup B| = |A| + |B|.$$

Proof. If $A = \varnothing$, then

$$A \cup B = B \qquad |A \cup B| = |B|, \quad \text{and} \quad |A| = 0,$$

in which case the desired formula follows. The same holds if $B = \varnothing$.

Now suppose both A and B are nonempty. To simplify notation, suppose $|A| = n$ and $|B| = m$ for some $m, n \in \mathbb{N}$. Let $f\colon \mathcal{F}_n \to A$ and $g\colon \mathcal{F}_m \to B$ be bijections. We have the vision that "the first n elements of $A \cup B$ are the elements of A, and the last m elements of $A \cup B$ are the elements of B." Thus we define

$$h\colon \mathcal{F}_{n+m} \to A \cup B\colon k \mapsto \begin{cases} f(k), & 1 \leq k \leq n \\ g(k-n), & n+1 \leq k \leq n+m. \end{cases}$$

We leave it as an exercise to check that h is bijective. Thus $|A \cup B| = m + n = |A| + |B|$. ■

This result generalizes to the union of any finite number of sets.

8.1.7 Corollary (Generalized addition rule).

Let (A_k) be a sequence of **PAIRWISE DISJOINT** *sets: $A_j \cap A_\ell = \varnothing$ for every j, $\ell \in \mathbb{N}$ with $j \neq \ell$. Then for each $n \in \mathbb{N}$, the union $\cup_{k=1}^{n} A_k$ is finite, and*

$$\left| \bigcup_{k=1}^{n} A_k \right| = \sum_{k=1}^{n} |A_k|.$$

We leave the proof as an exercise in induction and the recursive definition of the union.

This is where we finished on Friday, November 19, 2021.

8.1.8 Corollary (Difference rule).

Let A be a finite set and $C \subseteq A$. Then $A - C$ is finite and $|A - C| = |A| - |C|$. Does this equality[36] make sense if C is a finite set that is not a subset of A?

Proof. Since A is finite and $A - C \subseteq A$, $A - C$ is also finite, by Theorem 8.1.4. Next, we have the decomposition

$$A = (A - C) \cup C, \tag{8.1.1}$$

and this union is disjoint; we leave the proof of both the equality (8.1.1) and the disjointness as an exercise. Thus $|A| = |A - C| + |C|$ by Theorem 8.1.6, and so $|A - C| = |A| - |C|$. ■

8.1.9 Example.

Does the conclusion of the difference rule make sense if C is not a subset of A?

Solution. If C is not a subset of A, then $A - C$ is still a subset of A, and so $A - C$ is still finite. But the equality $|A - C| = |A| - |C|$ may not hold, or may not make sense. First, if C is infinite, then we have not defined $|C|$, and so this equality is meaningless. Next, if C is finite but $|C| > |A|$, then $|A| - |C| < 0$. We cannot have $|A - C| = |A| - |C|$ here, since $|A - C| \geq 0$.

Even stranger, consider a situation like $A = \{1, 2, 3\}$ and $C = \{4, 5, 6, 7\}$. Then $A - C = A$, so $|A - C| = |A|$. ▲

More generally, we can find a formula for the cardinality of $A \cup B$ even if we do not assume $A \cap B = \varnothing$. The intuition behind the following formula is that we count the elements of A, then count the elements of B (possibly double-counting some elements of A along the

[36] This equality is a nice example of a symbol doing double duty. On the left, $-$ denotes the set-theoretic difference. On the right, $-$ denotes the algebraic difference of two integers.

way), and then subtract the number of elements of $A \cap B$ to compensate for any double-counting. While the statement of the corollary subsumes that of Theorem 8.1.6, the proof of the corollary very much hinges on establishing Theorem 8.1.6 first[37].

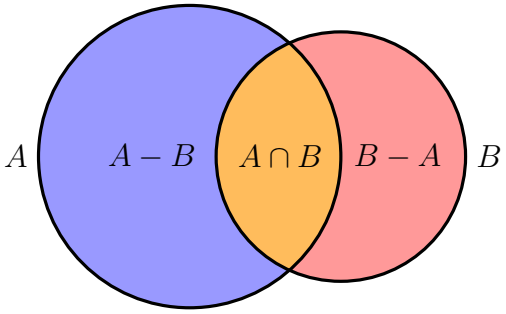
8.1.10 Corollary (Inclusion-exclusion principle).

Let A and B be finite sets. Then $A \cup B$ is finite, and

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Proof. The following Venn diagram suggests the disjoint decomposition

$$A \cup B = (A - B) \cup (A \cap B) \cup (B - A). \tag{8.1.2}$$



$A \cup B$

We leave it as an exercise to prove (8.1.2) formally. Then Corollary 8.1.7 implies

$$|A \cup B| = |A - B| + |A \cap B| + |B - A|. \tag{8.1.3}$$

On the other hand, we also have

$$A = (A - B) \cup (A \cap B) \quad \text{and} \quad B = (B - A) \cup (B \cap A), \tag{8.1.4}$$

and these are also disjoint unions. Thus

$$|A| = |A - B| + |A \cap B| \quad \text{and} \quad |B| = |B - A| + |B \cap A| = |B - A| + |A \cap B|. \tag{8.1.5}$$

(The statements in (8.1.4) and (8.1.5) are really generalizations of Corollary 8.1.8.) Hence

$$|A| + |B| = |A - B| + |B - A| + 2|A \cap B|,$$

and so

$$|A| + |B| - |A \cap B| = |A - B| + |B - A| + |A \cap B| = |A \cup B|$$

by (8.1.3). ■

There is a much more intricate inclusion-exclusion principle that gives a formula for $\left| \cup_{k=1}^{n} A_k \right|$ without the assumption of pairwise disjointness; we will survive without it.

[37] Translation: do not think that old results can be discarded just because they are restated in new results.

8.1.3. Cartesian products of finite sets.

From our experience with calculating Cartesian products, we might expect that the cardinality of the Cartesian product of two finite sets is the product of the cardinalities. Indeed, if $A = \{a_k\}_{k=1}^n$ and $B = \{b_k\}_{k=1}^m$, we might list the ordered pairs in $A \times B$ as the following "matrix":

$$\begin{bmatrix} (a_1, b_1) & (a_2, b_1) & \cdots & (a_n, b_1) \\ (a_1, b_2) & (a_2, b_2) & \cdots & (a_n, b_2) \\ \vdots & \vdots & \ddots & \vdots \\ (a_1, b_m) & (a_2, b_m) & \cdots & (a_n, b_m) \end{bmatrix}.$$

There are m rows, and each row has n ordered pairs, thus there are $mn = |A| \cdot |B|$ ordered pairs altogether. Of course, this argument needs some technical refinements and niceties, which we provide below.

This is where we finished on Monday, November 29, 2021.

8.1.11 Theorem (Product rule).

Let A and B be finite sets. Then $A \times B$ is finite, and

$$|A \times B| = |A| \cdot |B|. \tag{8.1.6}$$

Proof. If either A or B is empty, then the product $A \times B$ is empty, and consequently both sides of (8.1.6) are 0.

Suppose now that A and B are both nonempty. Write $A = \{a_k\}_{k=1}^n$ for some $n \in \mathbb{N}$. We claim that

$$A \times B = \bigcup_{k=1}^{n} \{a_k\} \times B \tag{8.1.7}$$

and that the sets in this union are disjoint, i.e., if $1 \leq j < \ell \leq n$, then

$$\big(\{a_j\} \times B\big) \cap \big(\{a_\ell\} \times B\big) = \varnothing.$$

Assuming these claims to be true, (8.1.7) and Corollary 8.1.7 imply

$$|A \times B| = \left| \bigcup_{k=1}^{n} \{a_k\} \times B \right| = \sum_{k=1}^{n} |\{a_k\} \times B|. \tag{8.1.8}$$

We leave it as an exercise to verify that if $|B| = m$ and $B = \{b_k\}_{k=1}^m$, then the map $f\colon \mathcal{F}_m \to \{a\} \times B\colon k \mapsto (a, b_k)$ is bijective for any $a \in A$. That is, $|\{a_k\} \times B| = m$. Then (8.1.8) implies

$$|A \times B| = \sum_{k=1}^{n} |\{a_k\} \times B| = \sum_{k=1}^{n} m = mn = |A| \cdot |B|. \qquad \blacksquare$$

8.1.12 Corollary (Generalized product rule).

Let (A_k) be a sequence of finite sets. Then the Cartesian product $\prod_{k=1}^{n} A_k$ is finite for any $k \in \mathbb{N}$ and

$$\left| \prod_{k=1}^{n} A_k \right| = \prod_{k=1}^{n} |A_k|.$$

Proof. If $n = 1$, then

$$\prod_{k=1}^{1} A_k = \{1\} \times A_1$$

by (7.3.1). By Theorem 8.1.11, this gives $|\prod_{k=1}^{1} A_k| = |A_1|$.

To prove the result for $n \geq 2$ we use induction, starting with the base case at $n = 2$. This base case is just Theorem 8.1.11 (again). We leave the details of the inductive step as an exercise but mention that it uses the bijection between $\prod_{k=1}^{n+1} A_k$ and $(\prod_{k=1}^{n} A_k) \times A_{n+1}$ from Example 7.3.7. The proof of this inductive step relies not only on the induction hypothesis but also on the $n = 2$ result. ■

8.1.4. Applications to choice counting.

> "The game is pretty straightforward. You can choose to spin or you can choose to choose. If you choose to spin you can land on Spin, or Choice, or Lose a Spin, or Lose a Choice, or Free Spin, or Free Choice, or Spin Again."
>
> —Peggy Hill, *King of the Hill*

When we make choices, we select one or more options from one or more sets of possibilities. Perhaps we do this multiple times, either with the same set(s) of possibilities or with different sets in a certain order. Such processes have fairly vast outcomes, of course. We begin our study of choices with two "simple" cases: first where we are just selecting one choice out of a number of non-overlapping possibilities, and next where we are making several successive choices. To count the number of possibilities in the first case, we will need the addition rule (Corollary 8.1.7).

8.1.13 Theorem (Informal addition rule).

Suppose that an individual has to select an element from one, and only one, of n disjoint sets. For $k = 1, \ldots, n$, there are m_k elements in each set. Then the individual has $\sum_{k=1}^{n} m_k$ possible choices.

Proof. Call the sets A_k for $k = 1, \ldots, n$, so $|A_k| = m_k$. The individual is choosing an element of $\cup_{k=1}^{n} A_k$, where the A_k are pairwise disjoint. Hence there are $\left| \cup_{k=1}^{n} A_k \right| = \sum_{k=1}^{n} |A_k| = \sum_{k=1}^{n} m_k$ elements from which to choose. ■

8.1.14 Example.

Dinner at the Pop Shop in Collingswood, NJ, is a sheer delight. A main dish is a burger, a grilled cheese sandwich, or, if one really must, a salad. There are 5 kinds of burgers, 5 kinds of grilled cheeses (of which the superior is clearly the "Arlington": mac n' cheese, cheddar, and crumbled bacon on thick-sliced sourdough), and 6 kinds of salads. If a diner orders one, and and only one, main dish, how many delicious dinners are possible?

Solution. 16: 5 burgers + 5 grilled cheeses + 6 salads. ▲

To count the number of possible options in a *finite sequence* of choices, we need the product rule (Corollary 8.1.12).

8.1.15 Theorem (Informal product rule).

Suppose that an individual has to make a finite sequence of n choices, one after another. For the kth choice, $1 \leq k \leq n$, the individual has m_k options. Then there are $\prod_{k=1}^{n} m_k$ possible sequences of choices.

Proof. For $k = 1, \ldots, n$, let A_k be the set of options for choice k, so $|A_k| = m_k$. Each sequence of choices is an n-tuple $(x_1, \ldots, x_n)$, where $x_k \in A_k$ for each k. That is, each sequence of choices corresponds to an element of $\prod_{k=1}^{n} A_k$, and so there are $\left|\prod_{k=1}^{n} A_k\right| = \prod_{k=1}^{n} |A_k| = \prod_{k=1}^{n} m_k$ choices. ■

8.1.16 Example.

A mathematician who specializes in nonlinear wave dynamics owns 10 distinct shirts (all plaid, naturally, as this is obligatory for analysts of nonlinear waves), 5 distinct pairs of pants, and 3 distinct pairs of shoes. How many distinct outfits can he wear to class, if an outfit consists of a shirt, a pair of paints, and a pair of shoes?

Solution. Each outfit corresponds to making three choices (of shirts, of pants, and of shoes), and there are $10 \cdot 5 \cdot 3 = 150$ such choices. Thus he has 150 dapper outfits. ▲

8.1.17 Example.

Let A and B be finite, nonempty sets. How many functions exist from A to B? That is, what is $|B^A|$?

Solution. ***Method 1.*** Let $|A| = n$ and $|B| = m$. To construct a function $f\colon A \to B$, each element of A needs to be paired with exactly one element of B. For the value of $f(a_1)$ we have m options, for $f(a_2)$ we also have m options and so on. In total we must must

make n choices, and we have m options for each choice. Hence the total number of (finite) sequences of choices is $\prod_{k=1}^{n} m = m^n = |B|^{|A|}$. This motivates the notation B^A for the set of all functions from A to B.

Method 2. Let $|A| = n$ and $|B| = m$. We induct on n and consider $m \in \mathbb{N}$ fixed throughout the proof. If $n = 1$, then $A = \{a_1\}$ for some element a_1; thus any function from A to B has the form $\{(a_1, y)\}$ for some $y \in B$, and therefore $B^A = \{a_1\} \times B$. Then $|B^A| = |B| = m = m^1$.

We make the induction hypothesis that if A and B are any sets with $|A| = n$ for some $n \geq 1$ and $|B| = m$, then $|B^A| = m^n$. Now suppose $|A| = n + 1$. Write $A = \{a_k\}_{k=1}^{n+1}$. We claim that the map

$$B^A \to B^{A-\{a_{n+1}\}} \times B \colon f \mapsto \left(f\big|_{\{a_k\}_{k=1}^{n}}, f(a_{n+1})\right)$$

is a bijection. Thus

$$|B^A| = |B^{A-\{a_{n+1}\}} \times B| = |B^{A-\{a_{n+1}\}}| \cdot |B| = m^n \cdot m = m^{n+1}.$$

To get the penultimate equality we used the induction hypothesis, which was permissible since $|A - \{a_{n+1}\}| = n$. ▲

8.1.18 Example.

A discrete mathematics student needs to write a truth table for a statement form with n "component" statements. How many rows are in the truth table?

Solution. ***Method 1.*** Each row in the truth table corresponds to a particular selection of truth values for the n component statements. Each of the n statements can be true or false. So, the student must make n choices with 2 options per choice. Thus there are $\prod_{k=1}^{n} 2 = 2^n$ choices.

Method 2. We could also view this question as asking how many functions there are from $\mathcal{F}_n$ to $\mathcal{F}_2$; the value of this function at $k \in \mathcal{F}_n$ is the truth value (say, 0 is false, 1 is true) of the kth component statement in the statement form. By Example 8.1.17 we know there are 2^n such functions. ▲

This is where we finished on Wednesday, December 1, 2021.

8.1.19 Example.

Let A be a finite set. Show that $|\mathscr{P}(A)| = 2^{|A|}$. This motivates the occasional notation of 2^A instead of $\mathscr{P}(A)$ for the power set of A.

Solution. ***Method 1.*** We can construct subsets of A by deciding whether or not to include elements of A. To obtain a subset C of A, we parse the elements of A and either we choose to include a given element in C or we choose to exclude it. If $|A| = n$, then we must make n choices, and for each choice we have 2 options (include or exclude). Thus there are $2^n = 2^{|A|}$ total ways to choose and so 2^n subsets.

Method 2. Each subset of A has a "natural" correspondence to a function $f\colon \mathcal{F}_n \to \mathcal{F}_2$, where if $f(k) = 1$ we say that a_k is in the subset, and if $f(k) = 0$ we say that a_k is not in the subset. More precisely, the map

$$\mathscr{P}(A) \to \mathcal{F}_2^{\mathcal{F}_n}\colon A \mapsto \chi_A, \qquad \chi_A(x) := \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

is a bijection. (Recall from Definition 6.1.4 that we denote by $\mathcal{F}_2^{\mathcal{F}_n}$ the set of all functions from $\mathcal{F}_n$ to $\mathcal{F}_2$.) Thus $|\mathscr{P}(A)| = \left|\mathcal{F}_2^{\mathcal{F}_n}\right| = 2^n$.

Method 3. Let $|A| = n$. We induct on n, starting with the base case $n = 0$. Here $A = \varnothing$, so $\mathscr{P}(\varnothing) = \{\varnothing\}$, and clearly $|\{\varnothing\}| = 1$.

Now suppose that if $|A| = n$ for some $n \geq 1$, then $|\mathscr{P}(A)| = n$. Let A be a set with $|A| = n + 1$. Fix $x \in A$. The induction hypothesis gives $|\mathscr{P}(A - \{x\})| = 2^n$. We claim that we have the following disjoint union:

$$\mathscr{P}(A) = \mathscr{P}(A - \{x\}) \cup \{C \cup \{x\} \mid C \in \mathscr{P}(A - \{x\})\}.$$

Then Theorem 8.1.6 implies

$$|\mathscr{P}(A)| = |\mathscr{P}(A - \{x\})| + |\{C \cup \{x\} \mid C \in \mathscr{P}(A - \{x\})\}|.$$

We also claim that

$$|\{C \cup \{x\} \mid C \in \mathscr{P}(A - \{x\})\}| = |\mathscr{P}(A - \{x\})|.$$

Thus

$$|\mathscr{P}(A)| = 2|\mathscr{P}(A - \{x\})| = 2(2^n) = 2^{n+1}. \qquad \blacktriangle$$

Each choice in the preceding sequences was independent of its predecessors and successors. Knowledge of what happened in choice $k - 1$ did not affect choice k, nor did choice k affect choice $k + 1$. However, there are many situations in which subsequent choices *do* affect each other. We can still count using the product rule, but our choices will now become much more "k-dependent."

For example, it will be useful to know how many bijections exist between two sets (of the same cardinality, naturally). Any bijection between sets A and B, where necessarily $|A| = |B|$, amounts to pairing elements of A in a one-to-one fashion with elements of B. Say that $A = \{a_k\}_{k=1}^n$. To get a bijection $f\colon A \to B$, we have to select one of n values in B for $f(a_1)$. That leaves us with $n - 1$ values for $f(a_2)$, and $n - 2$ values for $f(a_3)$, all the way down to 2 values for $f(a_{n-1})$ and only one value for $f(a_n)$ at the very end. The product rule suggests that there are $n!$ such ways to construct f, and induction makes this discussion of "selecting values" more rigorous.

8.1.20 Theorem.

Suppose that A and B are sets with $|A| = |B| = n$ for some $n \in \mathbb{N}$. There are $n!$ bijections

from A to B. That is,

$$\left|\left\{f \in B^A \;\middle|\; f \text{ is bijective}\right\}\right| = n!.$$

Proof. We denote the set of bijections from A to B by $\mathbf{BIJ}(A, B)$. We prove this theorem by induction on n with the base case $n = 1$.

At $n = 1$ we can write $A = \{a_1\}$ and $B = \{b_1\}$ for some elements a_1 and b_1. Then the only function from A to B is $\{(a_1, b_1)\}$, and this is a bijection. Hence $|\mathbf{BIJ}(A, B)| = 1 = 1!$.

Now we make the inductive hypothesis that for $n \geq 1$ and *all* sets A and B with $|A| = |B| = n$, we have $|\mathbf{BIJ}(A, B)| = n$. We have

$$\mathbf{BIJ}(A, B) = \bigcup_{k=1}^{n+1} \mathbf{BIJ}_k(A, B), \qquad \mathbf{BIJ}_k(A, B) := \left\{f \in \big(A, B\big) \;\middle|\; f(a_k) = b_k\right\}.$$

We leave it as an exercise to show that the sets in $\{\mathbf{BIJ}_k(A, B)\}_{k=1}^{n+1}$ are pairwise disjoint. Thus

$$|\mathbf{BIJ}(A, B)| = \sum_{k=1}^{n+1} |\mathbf{BIJ}_k(A, B)|.$$

If, for $f \in \mathbf{BIJ}_k(A, B)$, we "eliminate the pair (a_k, b_k) from consideration," then f is really a bijection from $A - \{a_k\}$ to $B - \{b_k\}$. That is, the map

$$g\colon \mathbf{BIJ}_k(A, B) \to \mathbf{BIJ}(A - \{a_k\}, B - \{b_k\})\colon f \mapsto f\big|_{A-\{a_k\}}$$

is itself a bijection. Thus

$$|\mathbf{BIJ}_k(A, B)| = |\mathbf{BIJ}(A - \{a_k\}, B - \{b_k\})| = n!$$

by the induction hypothesis, since $|A - \{a_k\}| = |A| - 1 = n$ by Example 8.1.8, and likewise $|B - \{b_k\}| = n$. Hence

$$|\mathbf{BIJ}(A, B)| = \sum_{k=1}^{n+1} |\mathbf{BIJ}_k(A, B)| = \sum_{k=1}^{n+1} n! = (n + 1)n! = (n + 1)!. \qquad \blacksquare$$

Here is a useful paraphrase of the preceding result.

8.1.21 Theorem (Ordered selection without replacement).

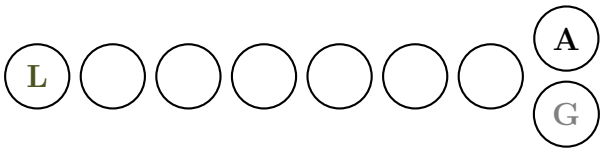
Suppose that an individual needs to arrange or select n items in some particular recognizable order. There are $n!$ such ways to order these items.

Proof. Let A be the set of items under consideration, so $|A| = n$. Any "arrangement" or "ordering" of A is equivalent to pairing the elements of A in a one-to-one fashion with the numbers 1 through n. There are $n!$ bijections from $\mathcal{F}_n$ to A. $\blacksquare$

8.1.22 Example.

The Company of the Ring is setting out on the Quest of Mount Doom. There are nine members of the Company ("set against the Nine Riders that are evil"). Each day they walk in a particular arrangement. Gandalf and Aragorn walk side-by-side in front. The other seven members walk in single file behind these two. Legolas the Elf is always the last member of the line. Suppose that all four hobbits in the Company desire to walk one behind the other. Under these strictures, how many ways can the Company arrange themselves?

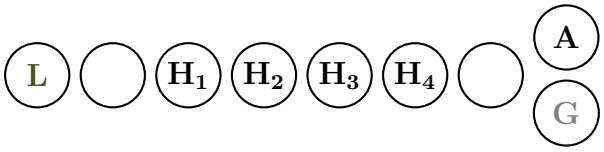
Solution. Here is a typical arrangement of the Company. (We presume that Gandalf walks to the right of Aragorn, chiefly for the sake of Footnote 39.)



The four hobbits and the remaining two members of the Company must fill the six empty circles.

We first pick where the hobbits walk. This amounts to choosing a circle to the right of Legolas (**L**) and then filling in that circle and the next three with hobbits. This requires us to have at least three empty circles between the first hobbit circle and the Aragorn (**A**)/Gandalf (G) leading pair. Only the first three circles after Legolas (**L**) will work. Thus we have 3 choices for where the first hobbit goes.

Next we need to arrange our hobbits. There are four hobbits ("Hobbits! Four hobbits! And what's more, out of the Shire by their talk"), so there are $4! = 24$ arrangements of them in a line. Here is one such possible arrangement.



Finally, we need to place our remaining two members of the Company (Boromir and Gimli). There are only two circles left, and so there are $2! = 2$ ways[38] to arrange Boromir and Gimli.

All together, we have made a sequence of three choices: the starting position of the hobbits, the arrangement of the hobbits, and the arrangement of the last two members. Thus there are

$$3 \cdot 24 \cdot 2 = 144$$

ways[39] for the Company to walk. ▲

[38] Note that if, for cinematographic purposes, the Company walked *all* in a genuine line and Aragorn did not go side-by-side with Gandalf, we would have three members of the Company to fill in, and there would be $3! = 6$ ways to do that. The point is to recall that there are $n!$ ways to fill n slots with n different Walkers.

[39] The eagle-eyed reader will recall that 144 is equal to "twelve dozen (a number also called by the hobbits one Gross, though the word was not considered proper to use of people)."

We might also ask how many injections exist from a set A to a set B. By Theorem 8.1.5, we need to require $|A| \le |B|$. Just as in the discussion preceding Theorem 8.1.20, we can use the product rule to calculate this number. Suppose $|A| = m \le n = |B|$ and write $A = \{a_k\}_{k=1}^m$. We need to choose m distinct values for our injection $f\colon A \to B$. For $f(a_1)$ there are n values in B, and then for $f(a_2)$ there are $n-1$, all the way down to $n-(m-2)$ values for $f(a_{m-1})$ and finally $n-(m-1)$ values for $f(a_m)$. That is, we choose $f(a_k)$ from a set of $n-(k-1)$ values in B, and so the product rule says that there are $\prod_{k=0}^{m-1} n-(k-1)$ ways to do this. A rather more painful way is to proceed by induction, and we sketch that proof below.

8.1.23 Theorem.

Suppose that A and B are sets with $|A| = m \le n = |B|$. Then there are

$$\prod_{j=0}^{m-1}(n-j).$$

injections from A to B.

Proof. Let $\mathbf{INJ}(A,B)$ denote the set of injections from A to B. We prove this theorem for $A = \mathcal{F}_m$ and $B = \mathcal{F}_n$. That is, we show

$$\forall n \in \mathbb{N}\ \forall m \in \mathcal{F}_n : |\mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_n)| = \prod_{j=0}^{m-1}(n-j).$$

We induct on n. When $n = 1$, we must take $m = 1$ as well, and there is exactly one injection (and one function) from $\mathcal{F}_1$ to $\mathcal{F}_1$: it is $\{(1,1)\}$. Observe that $\prod_{j=0}^{1-1}(1-j) = \prod_{j=0}^{0} 1 = 1$.

Now assume that $|\mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_n)| = \prod_{j=0}^{m-1}(n-j)$ for some $n \ge 1$ and every $m \in \mathcal{F}_n$. We must show $|\mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_{n+1})| = \prod_{j=0}^{m-1}(n+1-j)$ for all $m \in \mathcal{F}_{n+1}$. We consider two cases on m.

Case 1. $m = 1$. Then every function from $\mathcal{F}_1$ to $\mathcal{F}_{n+1}$ is injective, and by Example 8.1.17 there are $(n+1)^1 = n+1$ functions from $\mathcal{F}_1$ to $\mathcal{F}_{n+1}$. And

$$\prod_{j=0}^{1-1}(n+1-j) = \prod_{j=0}^{0}(n+1-j) = n+1.$$

Case 2. $m = n+1$. By Theorem 8.1.5, every injection from $\mathcal{F}_{n+1}$ to $\mathcal{F}_{n+1}$ is a surjection and thus a bijection. That is, $\mathbf{INJ}(\mathcal{F}_{n+1}, \mathcal{F}_{n+1}) = \mathbf{BIJ}(\mathcal{F}_{n+1}, \mathcal{F}_{n+1})$, to use the notation of the proof of Theorem 8.1.20. We know from that theorem that $|\mathbf{BIJ}(\mathcal{F}_{n+1}, \mathcal{F}_{n+1})| = (n+1)!$, and we have

$$\prod_{j=0}^{(n+1)-1}(n+1-j) = \prod_{j=0}^{n}(n+1-j) = (n+1)!.$$

Case 3. $2 \leq m \leq n$. We distinguish the injections from $\mathcal{F}_m$ to $\mathcal{F}_{n+1}$ from those whose ranges contain $n+1$ and those whose ranges do not. In the latter case, such an injection really maps $\mathcal{F}_m$ to $\mathcal{F}_n$, and we can therefore apply the induction hypothesis to them. In the former case, if we remove the preimage of $n+1$ from $\mathcal{F}_m$, a preimage that (by injectivity) consists of precisely one element of $\mathcal{F}_m$, we are left with an injection from $\mathcal{F}_{m-1}$ to $\mathcal{F}_n$, and again the induction hypothesis will apply.

Specifically, we write

$$\mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_{n+1}) = \mathbf{INJ}^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1}) \cup \mathbf{INJ}^{\notin}(\mathcal{F}_m, \mathcal{F}_{n+1}), \tag{8.1.9}$$

where

$$\mathbf{INJ}^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1}) := \{f \in \mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_{n+1}) \mid n+1 \in f(\mathcal{F}_m)\}$$

and

$$\mathbf{INJ}^{\notin}(\mathcal{F}_m, \mathcal{F}_{n+1}) := \{f \in \mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_{n+1}) \mid n+1 \notin f(\mathcal{F}_m)\}.$$

The two sets in the union in (8.1.9) are disjoint, so

$$|\mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_{n+1})| = |\mathbf{INJ}^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1})| + |\mathbf{INJ}^{\notin}(\mathcal{F}_m, \mathcal{F}_{n+1})|. \tag{8.1.10}$$

We can further write

$$\mathbf{INJ}^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1}) = \bigcup_{k=1}^{m} \mathbf{INJ}_k^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1}),$$

where

$$\mathbf{INJ}_k^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1}) := \{f \in \mathbf{INJ}^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1}) \mid f(k) = n+1\}.$$

The m sets here are disjoint, so

$$|\mathbf{INJ}^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1})| = \sum_{k=1}^{m} |\mathbf{INJ}_k^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1})|.$$

We claim that one can construct a bijection from $\mathbf{INJ}_k^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1})$ to $\mathbf{INJ}(\mathcal{F}_{m-1}, \mathcal{F}_n)$ by identifying $f \in \mathbf{INJ}_k^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1})$ with $f\big|_{\mathcal{F}_m - \{k\}}$; we leave the details as an exercise. Then $|\mathbf{INJ}_k^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1})| = |\mathbf{INJ}(\mathcal{F}_{m-1}, \mathcal{F}_n)|$, so the induction hypothesis implies

$$|\mathbf{INJ}^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1})| = \sum_{k=1}^{m} |\mathbf{INJ}(\mathcal{F}_{m-1}, \mathcal{F}_n)| = m|\mathbf{INJ}(\mathcal{F}_{m-1}, \mathcal{F}_n)| = m \prod_{j=0}^{m-2} (n-j). \tag{8.1.11}$$

Note that this product is defined since $m \geq 2$ here.

More simply, if $f \in \mathbf{INJ}^{\notin}(\mathcal{F}_m, \mathcal{F}_{n+1})$, then $f(\mathcal{F}_m) \subseteq \mathcal{F}_n$, and so $f \in \mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_n)$. Conversely, if $f \in \mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_n)$, then because $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, we also have $f \in \mathbf{INJ}^{\notin}(\mathcal{F}_m, \mathcal{F}_{n+1})$. Thus $\mathbf{INJ}^{\notin}(\mathcal{F}_m, \mathcal{F}_{n+1}) = \mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_n)$, and so

$$|\mathbf{INJ}^{\notin}(\mathcal{F}_m, \mathcal{F}_{n+1})| = |\mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_n)| = \prod_{j=0}^{m-1} (n-j) \tag{8.1.12}$$

by the induction hypothesis.

We conclude from (8.1.10), (8.1.11), and (8.1.12) that

$$|\mathbf{INJ}(\mathcal{F}_m, \mathcal{F}_{n+1})| = |\mathbf{INJ}^{\in}(\mathcal{F}_m, \mathcal{F}_{n+1})| + |\mathbf{INJ}^{\notin}(\mathcal{F}_m, \mathcal{F}_{n+1})| = m \prod_{j=0}^{m-2}(n-j) + \prod_{j=0}^{m-1}(n-j).$$

We leave the calculation

$$m \prod_{j=0}^{m-2}(n-j) + \prod_{j=0}^{m-1}(n-j) = \prod_{j=0}^{m-1}(n+1-j)$$

as an exercise. (Hint: each side equals $(n+1)\prod_{j=0}^{m-2}(n-j)$.) ■

8.2. Binomial coefficients.

Suppose that we wish to select k distinct elements from a set of n elements. In other words, how many subsets of cardinality k does a set of cardinality n have? In symbols, if $|A| = n$, what is

$$\big|\{C \in \mathscr{P}(A) \mid |C| = k\}\big|?$$

By the way, since $\{C \in \mathscr{P}(A) \mid |C| = k\} \subseteq \mathscr{P}(A)$, and since $\mathscr{P}(A)$ is finite, the cardinality above is actually defined. And we leave it as an exercise to verify that if B is another set with $|B| = n$, then

$$\big|\{C \in \mathscr{P}(A) \mid |C| = k\}\big| = \big|\{D \in \mathscr{P}(B) \mid |D| = k\}\big|. \tag{8.2.1}$$

Thus we are really speaking of the same number when we ask how many subsets of cardinality k does a set of cardinality n have, regardless of what that set of cardinality n is.

So, given n, $k \in \mathbb{N} \cup \{0\}$ and a set A with $|A| = n$, how many ways can we choose $C \subseteq A$ with $|C| = k$? If $k = 0$, there is only one way to do this by definition of cardinality: $C = \varnothing$. If $k = n$ and $C \subseteq A$ with $|C| = n = |\mathcal{F}_n|$, then there is also only one way to do this: $C = A$, per Theorem 8.1.4. And if $k > n$, then there certainly are no subsets of A with cardinality k, again per Theorem 8.1.4.

What about the intermediate cases? For simplicity, we may as well choose $A = \mathcal{F}_n$, per (8.2.1). Let us write

$$\binom{n}{k} := \big|\{C \in \mathscr{P}(\mathcal{F}_n) \mid |C| = k\}\big|.$$

This symbol is pronounced "n choose k," and, in words, it is the number of subsets of k elements of a set of n elements. We also write it as $\binom{n}{k}$.

Let us settle down to pick k elements (numbers) out of $\mathcal{F}_n$ without repetition. Of course we do not want to pick the same element repeatedly, because then we would not end up with a set of k distinct elements. Then for the first element we have n choices, for the second $n-1$, for the third $n-2$, and so on down to $n-(k-1)$ for the kth element. The product rule tells us that we can make $\prod_{j=0}^{k-1}(n-j)$ such sequences of choices. This, by the way, is exactly the same as the number of injections from $\mathcal{F}_m$ to $\mathcal{F}_n$, per Theorem 8.1.23.

In fact, this is exactly the same way we rationalized the result of Theorem 8.1.23. However, we have introduced some artificial structure here, and we do not really have $\binom{n}{k} = \prod_{j=0}^{k-1}(n-j)$. By saying "for the first element we have n choices, for the second $n-1\ldots$" above, we are not only selecting elements of $\mathcal{F}_n$ but *ordering* them as well. But we do not care about order when we are just picking k numbers from $\mathcal{F}_n$. For example, if we are picking 2 elements out of $\mathcal{F}_3$, then picking 1 first and 3 second gives us the same result as picking 3 first and then 1 second: in each case we have chosen the subset $\{1,3\} = \{3,1\}$.

This is where we finished on Friday, December 3, 2021.

We can see the (enlightening) error of our ways if we return to the method of constructing injections that preceded Theorem 8.1.23. We can define an injection $f\colon \mathcal{F}_k \to \mathcal{F}_n$ via a sequence of two choices. First we pick the image $f(\mathcal{F}_k)$; that is, we pick k elements from $\mathcal{F}_n$, and there are $\binom{n}{k}$ ways to do this. Then we arrange those k elements in some order to pair them up with the numbers in $\mathcal{F}_k$. That is, we establish a bijection from $\mathcal{F}_k$ to $f(\mathcal{F}_k)$. And there are $k!$ such bijections. Thus the number of injections from $\mathcal{F}_k$ to $\mathcal{F}_n$ should be $k!\binom{n}{k}$, and so

$$\binom{n}{k} = \frac{1}{k!}\prod_{j=0}^{k-1}(n-j).$$

This number, by the way, should be no larger than the number of injections from $\mathcal{F}_k$ to $\mathcal{F}_n$.

There is a slightly more transparent formula for $\binom{n}{k}$, which we can also develop via induction. Like the inductive proof of Theorem 8.1.23, the following inductive proof has less motivation than our product rule argument above.

8.2.1 Theorem.

Let $n \in \mathbb{N}$ and $k \in \mathcal{F}_n \cup \{0\}$. Then

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Proof. First we treat the case $k = 0$. If $A \subseteq \mathcal{F}_n$ for some $n \in \mathbb{N}$ and $|A| = 0$, then $A = \varnothing$. That is,

$$\forall n \in \mathbb{N} : |\{C \in \mathscr{P}(\mathcal{F}_n) \mid |C| = 0\}| = |\{\varnothing\}| = 1.$$

And

$$\frac{n!}{0!(n-0)!} = 1.$$

Here we used the convention $0! = 1$.

For the remainder of the proof we may therefore assume that $1 \le k \le n$. To be explicit, we will prove the statement

$$\forall n \in \mathbb{N}\ \forall k \in \mathcal{F}_n : \big|\{C \in \mathscr{P}(\mathcal{F}_n) \mid |C| = k\}\big| = \frac{n!}{k!(n-k)!}.$$

We induct on n. If $n = 1$ and $1 \leq k \leq n$, then $k = 1$. That is, if $C \subseteq \mathcal{F}_1$ with $|C| = 1$, then $C = \mathcal{F}_1 = \{1\}$ by Theorem 8.1.4. Thus

$$|\{C \in \mathscr{P}(\mathcal{F}_1) \mid |C| = 1\}| = |\{\mathcal{F}_1\}| = 1.$$

And

$$\frac{1!}{1!(1-1)!} = 1,$$

again using $0! = 1$.

Now suppose that for some $n \geq 1$ and all $k \in \mathcal{F}_n$ we have

$$\left|\{C \in \mathscr{P}(\mathcal{F}_n) \mid |C| = k\}\right| = \frac{n!}{k!(n-k)!}.$$

Consider a subset $C \subseteq \mathcal{F}_{n+1}$ such that $|C| = k$. Either $n+1 \in C$ or $n+1 \notin C$. If $n+1 \in C$, then $C - \{n+1\} \subseteq \mathcal{F}_n$ and $|C - \{n+1\}| = k-1$. If $n+1 \notin C$, then $C \subseteq \mathcal{F}_n$. This suggests that we have the disjoint decomposition

$$\{C \in \mathscr{P}(\mathcal{F}_{n+1}) \mid |C| = k\} = \{C \in \mathscr{P}(\mathcal{F}_n) \mid |C| = k\} \cup \{D \cup \{n+1\} \mid D \in \mathscr{P}(\mathcal{F}_n), |D| = k-1\}.$$

We leave the formal proof as an exercise. Then

$$\begin{aligned} &|\{C \in \mathscr{P}(\mathcal{F}_{n+1}) \mid |C| = k\}| \\ &\qquad = |\{C \in \mathscr{P}(\mathcal{F}_n) \mid |C| = k\}| + |\{D \cup \{n+1\} \mid D \in \mathscr{P}(\mathcal{F}_n), |D| = k-1\}|. \end{aligned} \tag{8.2.2}$$

The induction hypothesis tells us

$$|\{C \in \mathscr{P}(\mathcal{F}_n) \mid |C| = k\}| = \binom{n}{k}. \tag{8.2.3}$$

If $k \geq 2$, the induction hypothesis also tells us

$$|\{D \cup \{n+1\} \mid D \in \mathscr{P}(\mathcal{F}_n), |D| = k-1\}| = \binom{n}{k-1} \tag{8.2.4}$$

due to the bijection

$$\{D \cup \{n+1\} \mid D \in \mathscr{P}(\mathcal{F}_n), |D| = k-1\} \to \{E \in \mathscr{P}(\mathcal{F}_n) \mid |E| = k-1\} : D \cup \{n+1\} \mapsto D.$$

If $k = 1$, then we still have (8.2.4) from the work with $k = 0$ above. We combine (8.2.2), (8.2.3), and (8.2.4) to conclude

$$|\{C \in \mathscr{P}(\mathcal{F}_{n+1}) \mid |C| = k\}| = \binom{n}{k} + \binom{n}{k-1}.$$

We leave it as an exercise to verify that this sum equals $\binom{n+1}{k}$. ■

8.2.2 Example.

Recalling the situation of Example 1.0.1, suppose that every person in a group of n people shakes the hand of every other person in the group. How many handshakes take place?

Solution. As in Example 1.0.1, a handshake is completely determined by the two people shaking hands, and if person A shakes hands with person B, that is the same as person B shaking hands with person A. Thus each handshake can be uniquely identified with a pair of people from our set, and so we are choosing subsets of cardinality 2 from a set of cardinality n. There are $\binom{n}{2}$ ways to do this.

In particular, if $n = 4$, there are

$$\binom{4}{2} = \frac{4!}{2!(4-2)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{(2 \cdot 1) \cdot (2 \cdot 1)} = \frac{24}{4} = 6$$

handshakes, which is exactly what we argued back in Example 1.0.1. ▲

This is where we finished on Monday, December 6, 2021.

A. Thesaurus

We discourage the use of symbols like $\forall$, $\exists$, $\implies$, and $\iff$ "in text" and provide some alternatives in English below. For example, within a paragraph, within a proof, we prefer to write "For all $x \in \mathbb{R}$ there is $y \in \mathbb{R}$ such that $x + y = 0$" instead of something totally symbolic like $\forall x \in \mathbb{R}\ \exists y \in \mathbb{R} : x + y = 0$ or a combination like "For all $x \in \mathbb{R}\ \exists y \in \mathbb{R}$ such that $x + y = 0$." These are not totally absolute rules, and we will probably encounter valid exceptions. The following table is far from exhaustive.

Symbol	**Synonym(s)**
$\forall$	for all, for each, for every, all, each, every
$\exists$	there exists, there is, there are some
$\implies$	if-then, since, because, consequently, thus, hence, so

B. Russell's Paradox

If the empty set contains *nothing*, then its mortal enemy should be the set that contains *everything*. Can such a "universal" set exist? Does there exist a set $\mathcal{U}$ that contains every possible element? If so, we would expect, for example, both $\mathbb{R} \subseteq \mathcal{U}$ and $\mathbb{R} \in \mathcal{U}$, and, for that matter, $\{\mathbb{R}\} \in \mathcal{U}$.

Suppose that a universal set $\mathcal{U}$ containing every possible element does exist. In particular, $\mathcal{U}$ contains every set. Let $P(X)$ be the predicate "X is a set and $X \notin X$" with domain $\mathcal{U}$. The axiom of separation allows us to form the set

$$\mathcal{R} := \{X \in \mathcal{U} \mid P(X)\} = \{X \in \mathcal{U} \mid X \text{ is a set and } X \notin X\}.$$

We show that the contradiction $\mathcal{R} \in \mathcal{R} \iff \mathcal{R} \notin \mathcal{R}$ results.

First, if $\mathcal{R} \in \mathcal{R}$, then it is the case that $P(\mathcal{R})$. In particular, $\mathcal{R} \notin \mathcal{R}$. Thus the statement $\mathcal{R} \in \mathcal{R} \implies \mathcal{R} \notin \mathcal{R}$ is true. (Note that if P is false, then the statement $P \implies \sim P$ is true, and so, as strange as $\mathcal{R} \in \mathcal{R} \implies \mathcal{R} \notin \mathcal{R}$ sounds, it is not yet a contradiction, i.e., a statement that is always false.)

But if $\mathcal{R} \notin \mathcal{R}$, then by part (iii) of Example 5.1.2, it must be the case that either $\mathcal{R} \notin \mathcal{U}$ or $\sim P(\mathcal{R})$. Since $\mathcal{U}$ contains everything, including all sets, and since $\mathcal{R}$ is a set, we must have $\mathcal{R} \in \mathcal{U}$. So, the only possibility left is that it is the case that $\sim P(\mathcal{R})$. But then either $\mathcal{R}$ is not a set or $\mathcal{R} \notin \mathcal{R}$. The axiom of separation has ensured that $\mathcal{R}$ is a set, and so the last possibility, which must be the case, is that $\mathcal{R} \notin \mathcal{R}$.

The problem is either that a universal set containing in particular all sets cannot exist, or the axiom of separation cannot be true. We can live without a universal set; we cannot live without the axiom of separation.

C. Counting the sets $\mathcal{F}_n = \{1, \ldots, n\}$

THERE. ARE. FOUR. LIGHTS.

—Jean-Luc Picard, Captain, USS *Enterprise* NCC-1701-D

Recall that for $n \in \mathbb{N}$ we write

$$\mathcal{F}_n := \mathbb{N} \cap [1, n] = \{k \in \mathbb{N} \mid 1 \leq k \leq n\}.$$

We prove a number of counting results for the sets $\mathcal{F}_n$. Most of these results are universal statements of the form

$$\forall n \in \mathbb{N} : P(n) \quad \text{where} \quad P(n) \equiv \forall m \in \mathcal{F}_n : Q(m, n),$$

and where $Q(m, n)$ is a(nother) predicate, now with domain $\mathbb{N} \times \mathbb{N}$. Usually $Q(m, n)$ will tell us something about how $\mathcal{F}_m$ and $\mathcal{F}_n$ interact. Typically, once we have specified n, we will use m to denote an integer less than or equal to n.

The natural way to prove these statements is by induction on n. The $n = 1$ base case usually involves very facile results about the set $\{1\}$ or, at worst, $\{1, 2\}$. The proof of the inductive step will usually be rather demanding and often proceed by contradiction. We state a number of technical results as lemmas and corollaries to highlight the logical dependencies among different ideas and to emphasize how we can use the same idea over and over.

Our first lemma captures the idea that if we remove one element from a set of n elements, then we are left with $n - 1$ elements.

C.0.1 Lemma.

Let $n \in \mathbb{N}$ with $n \geq 2$ and $k \in \mathcal{F}_n$. There exists a bijection $f : \mathcal{F}_{n-1} \to \mathcal{F}_n - \{k\}$.

Proof. We consider three possibilities on k. In each case, we "slide" the elements of $\mathcal{F}_{n-1}$ around to "cover" $\mathcal{F}_n - \{k\}$, allowing for a "gap" at k.

First suppose $k = 1$ and define

$$f : \mathcal{F}_{n-1} \to \mathcal{F}_n - \{1\} : j \mapsto j + 1.$$

Next, suppose $1 < k < n$ and define

$$g : \mathcal{F}_{n-1} \to \mathcal{F}_n - \{k\} : j \mapsto \begin{cases} j, & 1 \leq j < k \\ j + 1, & k \leq j \leq n - 1. \end{cases}$$

Last, suppose $k = n$ and let h be the "inclusion" map

$$h : \mathcal{F}_{n-1} \to \mathcal{F}_n - \{n\} : j \mapsto j.$$

We leave it as an exercise to check that each of these maps is bijective. ■

The next lemma captures the idea that we cannot pair $n + 1$ elements with n elements in a one-to-one fashion. (If there are more people than chairs and everyone wants to sit, at least two people will have to share a chair.)

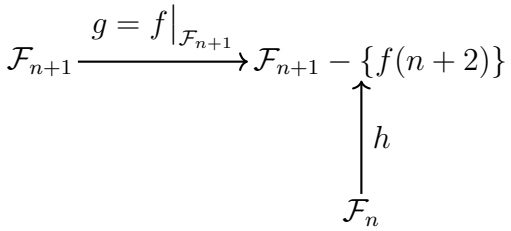
C.0.2 Lemma.

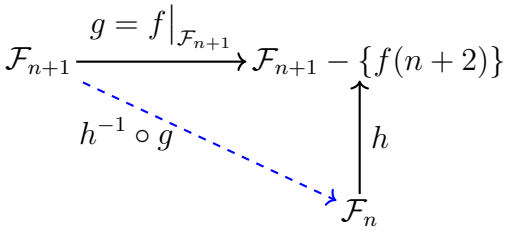
Let $n \in \mathbb{N}$. No function $f\colon \mathcal{F}_{n+1} \to \mathcal{F}_n$ is injective.

Proof. We induct on n with the base case $n = 1$. Here $\mathcal{F}_1 = \{1\}$ and $\mathcal{F}_2 = \{1, 2\}$. The only function from $\mathcal{F}_2$ to $\mathcal{F}_1$ is the constant function $\{(1, 1), (2, 1)\}$, which is not injective.

Suppose that given $n \geq 1$, we know there is no injection from $\mathcal{F}_{n+1}$ to $\mathcal{F}_n$. Now we must show that there is no injection from $\mathcal{F}_{(n+1)+1} = \mathcal{F}_{n+2}$ to $\mathcal{F}_{n+1}$. We prove this induction step by contradiction: suppose there *does* exist an injection $f\colon \mathcal{F}_{n+2} \to \mathcal{F}_{n+1}$.

Let $g = f\big|_{\mathcal{F}_{n+1}}$. Since $f(\mathcal{F}_{n+2}) \subseteq \mathcal{F}_{n+1}$, we have $g(\mathcal{F}_{n+1}) \subseteq \mathcal{F}_{n+1} - \{f(n+2)\}$. Also, $g\colon \mathcal{F}_{n+1} \to \mathcal{F}_{n+1} - \{f(n+2)\}$ is injective, since g is the restriction of an injection. Now let $h\colon \mathcal{F}_n \to \mathcal{F}_{n+1} - \{f(n+2)\}$ be a bijection from Lemma C.0.1.





Then $h^{-1}\colon \mathcal{F}_{n+1} - \{f(n+2)\} \to \mathcal{F}_n$ is bijective and in particular injective. Thus $h^{-1} \circ g\colon \mathcal{F}_{n+1} \to \mathcal{F}_n$ is injective, which contradicts the induction hypothesis. ■

We put these two lemmas together to show that if $m < n$, then a set of m elements does not have the same number of elements as a set of n elements.

C.0.3 Theorem.

Let $m, n \in \mathbb{N}$ with $m < n$. No function $f\colon \mathcal{F}_n \to \mathcal{F}_m$ is bijective.

Proof. It suffices to show that no function from $\mathcal{F}_n$ to $\mathcal{F}_m$ is injective. (It turns out that a surjection from $\mathcal{F}_n$ to $\mathcal{F}_m$ always exists: let $f(k) = k$ for $1 \leq k \leq m$ and $f(k) = 1$ for $m + 1 \leq k \leq n$. Also, the strict inequality is necessary: if $n = m$, let the bijection be $f(k) = k$.)

We induct on n starting with the base case of $n = 2$. (If $n = 1$ then there is no $m \in \mathbb{N}$ such that $m < 1$.) If $n = 2$, then the only $m \in \mathbb{N}$ such that $m < 2$ is $m = 1$. Lemma C.0.2 then applies to prevent the existence of an injection.

Suppose now that for a given $n \geq 2$ there is no injection from $\mathcal{F}_n$ to $\mathcal{F}_m$ for each $m < n$. We must show that there is no injection from $\mathcal{F}_{n+1}$ to $\mathcal{F}_m$ for each $m < n + 1$, i.e., for each $m \leq n$.

First suppose $m < n$. If there is an injection $f\colon \mathcal{F}_{n+1} \to \mathcal{F}_m$, then $f\big|_{\mathcal{F}_n}\colon \mathcal{F}_n \to \mathcal{F}_m$ is also injective. But since $m < n$ this contradicts the induction hypothesis.

The other possibility is that $m = n$. But Lemma C.0.2 forbids the existence of an injection from $\mathcal{F}_{n+1}$ to $\mathcal{F}_n$. ■

Now we develop two useful corollaries of this result.

C.0.4 Corollary.

Let m, $n \in \mathbb{N}$. There exists a bijection $f\colon \mathcal{F}_n \to \mathcal{F}_m$ if and only if $n = m$.

Proof. ($\Longrightarrow$) Suppose there exists $f\colon \mathcal{F}_n \to \mathcal{F}_m$ bijective. If $m < n$, this would contradict Theorem C.0.3, and so it must be the case that $n \leq m$.

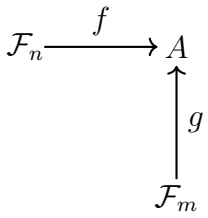
The map $f^{-1}\colon \mathcal{F}_m \to \mathcal{F}_n$ is also bijective. If $n < m$, this would contradict Theorem C.0.3 again (interchange the roles of m and n in the statement of that theorem). The only possibility let is that $n = m$.

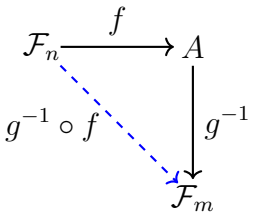
($\Longleftarrow$) If $m = n$, then $\mathcal{F}_n = \mathcal{F}_m$. Let $f\colon \mathcal{F}_n \to \mathcal{F}_n$ be the identity map, i.e., $f(k) = k$. This is a bijection (although it is not necessarily the only bijection from $\mathcal{F}_n$ to $\mathcal{F}_n$, at least if $n \geq 2$.) ■

C.0.5 Corollary.

Let A be a set and suppose there exist m, $n \in \mathbb{N}$ and bijections $f\colon \mathcal{F}_m \to A$ and $g\colon \mathcal{F}_n \to A$. Then $m = n$.

Proof. Since $g\colon \mathcal{F}_n \to A$ is bijective, the map $g^{-1}\colon A \to \mathcal{F}_n$ is also bijective.





Then $g^{-1} \circ f\colon \mathcal{F}_n \to \mathcal{F}_m$ is bijective, so by Corollary C.0.4 we must have $n = m$. ■

Finally, we show that any subset of a set of n elements contains at most n elements, and the subset contains exactly n elements if and only if it is the whole set.

C.0.6 Theorem.

Let $n \in \mathbb{N}$ and $A \subseteq \mathcal{F}_n$ be nonempty.

(i) *There exist $m \in \mathbb{N}$ with $m \leq n$ and a bijection $f\colon \mathcal{F}_m \to A$. This m is unique.*

(ii) *We have $m = n$ if and only if $A = \mathcal{F}_n$.*

Proof. (i) The uniqueness claim follows from Corollary C.0.5. We prove the existence claim by induction on n.

If $n = 1$, then $\mathcal{F}_1 = \{1\}$, and the only nonempty subset of $\mathcal{F}_1$ is $\mathcal{F}_1$ itself. Then the bijection f is given by $f(1) = 1$.

Now suppose the result is true for a given $n \geq 1$ and let $A \subseteq \mathcal{F}_{n+1}$ be nonempty. We consider two cases on A.

Case 1. $n + 1 \notin A$. Then $A \subseteq \mathcal{F}_n$. By the induction hypothesis there exist $m \in \mathbb{N}$ with $m \leq n$ and a bijection $f\colon \mathcal{F}_m \to A$. Since $m \leq n$, we have $m < n + 1$.

Case 2. $n+1 \in A$. Then $A - \{n+1\} \subseteq \mathcal{F}_n$, and so, again, there exist $m \in \mathbb{N}$ with $m \leq n$ and a bijection $f\colon \mathcal{F}_m \to A - \{n+1\}$. Now define

$$g\colon \mathcal{F}_{m+1} \to A\colon k \mapsto \begin{cases} f(k), & k \in \mathcal{F}_m \\ n+1, & k = m+1. \end{cases}$$

We claim that $g\colon \mathcal{F}_{m+1} \to A$ is a bijection and leave the proof as an exercise. Since $m \leq n$, we have $m+1 \leq n+1$.

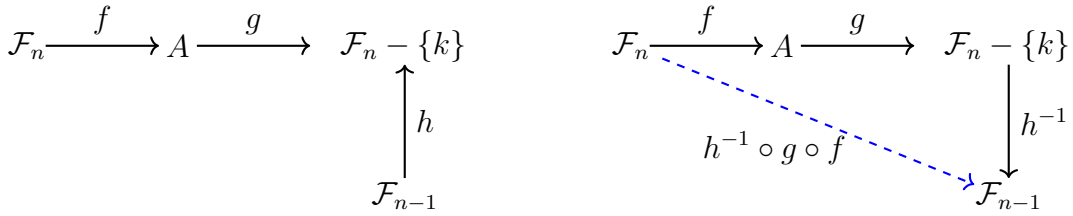
(ii) ($\Longleftarrow$) If $A = \mathcal{F}_n$, then the identity map $\mathcal{F}_n \to A\colon k \mapsto k$ is a bijection.

($\Longrightarrow$) If $n = 1$, then since $m \in \mathbb{N}$ and $m \leq 1$ we must have $m = 1$, too.

For the case of $n \geq 2$, we proceed by contradiction. That is, we suppose there is a bijection $f\colon \mathcal{F}_n \to A$, but $A \neq \mathcal{F}_n$. Then since $A \subseteq \mathcal{F}_n$, there is $k \in \mathcal{F}_n$ such that $k \notin A$, and so $A \subseteq \mathcal{F}_n - \{k\}$. The "inclusion" map

$$g\colon A \to \mathcal{F}_n - \{k\}\colon j \mapsto j$$

is then injective. Last, let $h\colon \mathcal{F}_{n-1} \to \mathcal{F}_n - \{k\}$ be a bijection from Lemma C.0.1. Then $h^{-1}\colon \mathcal{F}_n - \{k\} \to \mathcal{F}_{n-1}$ is bijective and therefore injective.



We conclude that $h^{-1} \circ g \circ f\colon \mathcal{F}_n \to \mathcal{F}_{n-1}$ is injective, which contradicts Lemma C.0.2. ■

Parallel to Lemma C.0.2 is the next result, which tells us that if $m < n$, then we cannot "cover" n elements with m elements.

C.0.7 Lemma.

Let $m, n \in \mathbb{N}$ with $m < n$. There does not exist a surjection $f\colon \mathcal{F}_m \to \mathcal{F}_n$.

Proof. We outline the proof and leave the details as an exercise. Induct on n starting with $n \geq 2$. For the base case $n = 2$ we can explicitly list the functions from $\mathcal{F}_1$ to $\mathcal{F}_2$ and see that neither is a surjection.

Assume that for some $n \geq 2$ there are no surjections from $\mathcal{F}_m$ to $\mathcal{F}_n$ with $m < n$. Suppose, by way of contradiction, that for some $m < n+1$ there is a surjection $f\colon \mathcal{F}_m \to \mathcal{F}_{n+1}$. Abbreviate $\mathcal{G}_n := \mathcal{F}_m - f^{-1}(\{n+1\})$. Explain why $f\big|_{\mathcal{G}_n}\colon \mathcal{G}_n \to \mathcal{F}_n$ is surjective. Check that $\mathcal{G}_n \subseteq \mathcal{F}_m$ and $\mathcal{G}_n \neq \mathcal{F}_m$; then use Theorem C.0.6 to show that there exist $\ell \in \mathbb{N}$ with $\ell < m$ and a bijection $g\colon \mathcal{F}_\ell \to \mathcal{G}_n$. (Be careful not to assume $\ell = m - 1$.) Conclude that $f\big|_{\mathcal{G}_n} \circ g\colon \mathcal{F}_\ell \to \mathcal{F}_n$ is surjective and deduce a contradiction to the induction hypothesis. ■

We conclude with the remarkable fact that injectivity and surjectivity are equivalent if the domain and codomain have the same number of elements.

C.0.8 Theorem.

A function $f\colon \mathcal{F}_n \to \mathcal{F}_n$ is injective if and only if it is surjective.

Proof. Again, we just outline the proof and leave the details as an exercise. We use contradiction for both directions.

($\Longrightarrow$) Suppose $f\colon \mathcal{F}_n \to \mathcal{F}_n$ is injective but not surjective. Explain why $f\colon \mathcal{F}_n \to f(\mathcal{F}_n)$ is bijective. Use Theorem C.0.6 to find $m \in \mathbb{N}$ with $m < n$ and a bijection $g\colon \mathcal{F}_m \to f(\mathcal{F}_n)$. Deduce that there exists a bijection $\mathcal{F}_n \to \mathcal{F}_m$ where $n \neq m$. This contradicts Corollary C.0.4.

($\Longleftarrow$) Suppose $f\colon \mathcal{F}_n \to \mathcal{F}_n$ is surjective but not injective. Explain why there exist k_1, $k_2 \in \mathcal{F}_n$ such that $f(k_1) = f(k_2)$ and $k_1 \neq k_2$. Abbreviate $\mathcal{G}_n := \mathcal{F}_n - \{k_2\}$. Explain why $f\big|_{\mathcal{G}_n}\colon \mathcal{G}_n \to \mathcal{F}_n$ is still surjective. Use Lemma C.0.1 to produce a surjection from $\mathcal{F}_{n-1}$ to $\mathcal{F}_n$, which contradicts Lemma C.0.7. ■