12A: Introduction to Logic

Nir Elber

Spring 2022

Contents

Part I

Propositional Logic

Theme 1
Syntax and Semantics

1.1 January 19

Let's go ahead and get going. Today we are hyping the course.

1.1.1 Symbolic Logic

In this course, we are more interested in symbolic logic. More broadly, we are interested in what kind of reasoning is "logical," and we will do this by abstracting what a good argument is.

In symbolic logic, we have lots of symbols for our reasoning words. Here is a table of such symbols.

word	symbol
not	$\neg$
and	$\wedge$
or	$\vee$
if, then	$\to$
for all	$\forall$
there exists	$\exists$

In this course we will be able to give a rigorous definition for what a valid formula is in the language of these symbols. The truth value of a statement will have no dispute.

Example 1.1. The formula

$$(\forall x \mathsf{Red}(x) \vee \exists y \mathsf{Square}(y)) \to \exists z(\mathsf{Red}(z) \wedge \mathsf{Square}(z))$$

asserts that "if everything is red and something is square, then there is something which is both red and square." This is a good, true assertion: that something which is square must also be red, finishing.

1.1.2 Advertisements

This sort of reasoning has applications in lots of fields.

- Logic is a main branch of philosophy. For example, we will study the syllogistic reasoning of Aristotle.[1]

[1] Here is an example of a syllogism: Suppose that all men are mortal, and that Socrates is a man. Then it follows Socrates is mortal.

- Logic is at the base of mathematics, and careful logical reasoning informs foundational mathematics (e.g., Gödel's incompleteness theorems or the independence of the Continuum hypothesis). For example, we will have to understand (basic) mathematical proofs in this course. We will talk about foundational mathematics a bit at the end of the course.
- Logic and its methods (e.g., λ-calculus) impacts how one does computer programming. For a concrete example, logic is used in SQL to give statements for database queries. As another example, formal hardware and software verification comes down to very careful logical analysis.
- One approach to artificial intelligence is by trying to create a machine which spits out true facts from old ones, for which the formal language of first-order logic is quite important.
- The kind of epistemic logic of trying to reason about what people know and do not know is important in game theory and hence has applications to economics. This can quickly get complicated: for example, we might want to keep track of the fact that (e.g., in poker) Player 1 knows that Player 2 knows that Player 3 has an ace card, for this fact might affect Player 2's behavior.
- Linguistics is interested in what sentences mean, for which one had to keep track of formal semantics to determine truth values.
- In cognitive science, how hard it is to understand/learn something turns out to be directly proportional to the length of the shortest logically equivalent propositional formula. In other words, longer formulae are harder to get in one's head.

1.1.3 Logistics

Let's talk about logistics.

- There is a class Piazza, which hopefully will get some use.
- The course outline in the syllabus is more of a guess than a promise; we may get ahead or behind it, but the syllabus will be updated frequently to match.
- There is a textbook. It is more like a math textbook: one is expected to read things multiple times instead of in an English class where one tries to skim as much as possible. While the material is dense, the course has been designed to try to make the course accessible to everyone.
- All reading (including the textbook) will be freely available online.
- It is better to skim the reading before lecture to not completely be lost during lecture. It is also good to do another pass on the reading after lecture.
- There are weekly problem sets, released on Monday (starting next Monday) and due on Sunday midnight. They will be graded via GradeScope, and there are regrade requests (as usual).
- In theory, the problem sets will not depend on a great deal on the lecture Friday before the deadline.
- The class will be curved upwards depending on its difficulty, at the very end.
- Please come to office hours instead of struggling needlessly on one's own.

Next class we will talk about propositional logic.

1.2 January 21

Today we are talking about propositional logic.

1.2.1 Proportions and Connectives

Roughly, propositional logic is about reasoning with propositions with propositional connectives.

Remark 1.2. Propositional logic might also be called sentential logic or boolean logic.

Here are propositions.

Definition 1.3 (Proposition)**.** In this class, a *proposition* will simply be any declarative sentence.

Example 1.4. The sentence "Paris the capital of France" is a proposition.

Non-Example 1.5. The question "Is Paris the capital of France?" is not a proposition.

Remark 1.6. Logic can handle questions, but we will not discuss it. It is called inquisitive logic.

Here are propositional connectives.

Definition 1.7 (Connective)**.** A *propositional connective* is some word or phrase that we can use to build new propositions from old ones.

There are lots of propositional connectives. Have some examples.

Example 1.8. Suppose p and q are propositions. For example, p can be "the thief entered through the back door," and q can be "the thief left through the side door." Then we have the following.

- "p or q" is a proposition.
- "p and q" is a proposition.
- "If p then q" is a proposition.
- "Not p" is a proposition. (Grammatically, "it is not the case that p.")

These are read by replacing p and q with the propositions they represent.

Example 1.9. Propositional connectives do not have to be absolute. For example, "it is more likely that p than it is that q" is a proposition.

1.2.2 Reasoning

So thus far we have propositions and their connectives. How might we reason with them? Here's an example.

Example 1.10. If we happen to know that "p or q," and we know that "it is not the case that p," then q must be true. This is essentially process of elimination; e.g., we might imagine running this argument with p that "the infection is viral" and q that "the infection is bacterial."

The reasoning in the above argument feels quite true, even without making p or q concrete propositions. The reasoning makes us feel good.

Warning 1.11. Suppose we have the following premises.

- p or q.
- p.

It does not follow that q is false. In short, p or q permits both p and q to be true.

Here is another example of reasoning.

Example 1.12. We have the following premises.

- If p, then q.
- Not q.

Then it follows that not p. Concretely, we can set p to be "the reactor is cooling down" and q to be "the blue light is on." Then the fact that the blue light is not on would imply that the reactor is not cooling down.

We do have to be careful with conditionals, however.

Warning 1.13. The premise "if p then q" does not imply that "if q then p."

And here is some reasoning with less concrete connectives.

Example 1.14. Suppose we are given that p is more likely than q. Then it follows that p is more likely than q and r, for any other event r. In essence, trying to make more events happen is harder. For concreteness, think through this argument with the following concrete premises:

- Set p to be "the US will sign the treaty."
- Set q to be "Russia will sign the treaty."
- Set r to be "China will sign the treaty."

Having more people sign the treaty is harder.

Reasoning can be hard, and sometimes our intuition might be wrong. Here is some bad reasoning.

Non-Example 1.15. Suppose we have the following premises.

- If p, then q.
- Not p.

Then it does not follow that q. Concretely, set p to be "the patient is taking her medicine" and q to be "the patient will get better." Then the fact that patient is not taking her medicine does not imply that the patient will not get better: perhaps the patient will get better for some other reason.

The above reasoning is bad because we went from true premises to false conclusions. This is what we try to avoid.

Here is more bad reasoning.

Non-Example 1.16. Suppose we have the following premises.

- It is more likely that p than q.
- It is more likely that p than r.

Then it does not follow that p is more likely than q or r. For example, take p to be the event that a die roll is odd, q to be the event that a die rolls is $\{1, 2\}$, and r to be the vent that a die rolls is $\{3, 4\}$. The probability that q or r exceeds the probability that p in this case.

As an aside, our bad reasoning might still give good premises at the end. The reason that we like good reasoning better is that every single time we will get good premises from good ones; with bad reasoning, we run the risk of getting bad results at the end.

Example 1.17. The argument that the proposition "grass is green" directly implies "the sky is blue" is not valid reasoning because these two propositions have effectively nothing to do with each other. Nevertheless, "the sky is blue" is a true conclusion.

1.2.3 Truth-Functional Connectives

Earlier we gave examples of lots of different propositional connectives. It turns out that we only care about very few of these: the truth-functional propositional connectives.

We need to have a reasonable notion of truth. We adopt the following conventions.

Convention 1.18. In this course, we take the following.

- All propositions are either true or false.
- No proposition is both true and false.

These probably seem obvious, but we need to be careful.

Example 1.19. The following propositions are bad in that they have unclear truth value.

- The proposition "ice cream is delicious" is a proposition of taste (this depends on the person), so we will ignore it.
- "Bob is bald" is a bit vague because "bald" is not well-defined, so we will ignore it.
- "This proposition is false" is self-referential and more or less breaks down truth (if the proposition is true, then the proposition declares its own falsehood), so we will ignore it. Similar is "this proposition is true." (This is known as the liar paradox.)

One way to escape these problems is to simply declare that they are not propositions. We choose to ignore the altogether.

Nevertheless, we go forwards with our notion of truth value.

Definition 1.20 (Truth value)**.** The *truth value* of a proposition is "true" if the proposition is true and "false" if it is false.

Very quickly, we note that the connectives we've talked about come in two classes.

Definition 1.21 (Unary, binary)**.** A propositional connective is *unary* (respectively, *binary*) if and only if it acts on one (respectively, two) proposition.

Example 1.22. The connective "not" is a unary connective. The connective "We know that" is a unary connective.

Definition 1.23. More generally, the *arity* of a connective is the number of propositions the connective acts on.

Example 1.24. The arity of "not" is 1.

Natural language tends to focus on unary and binary connectives.

We are now ready to define what a truth-functional connective is. Here is the definition for unary connectives.

Definition 1.25 (Truth-functional)**.** A unary connective $\#$ is *truth-functional* means that $\#p$ has truth value which is a function of (i.e., is completely determined by) the truth value of p.

Example 1.26. The connective "not" is a truth-functional connective: the truth value of p tells us what the truth value of "not p" is.

Non-Example 1.27. The connective "the police know that" is not a truth-functional connective: a statement being true or false does not immediately tell us whether the police know it. Hopefully if p is false, then the police do not know p; but if p is true, perhaps the police simply do not it yet. The point is that the truth value of "the police know that p" is simply not a function of p.

Here is truth-functionality for binary connectives.

Definition 1.28 (Truth-functional)**.** A binary connective $\#$ is *truth-functional* means that $p\#q$ has truth value which is a function of (i.e., is completely determined by) the truth values of p and q.

Example 1.29. The connective taking the propositions p, q to "p and q" is truth-functional.

1.3 January 24

Okay, welcome back everybody.

1.3.1 Truth-Functionality

Recall the definition.

Definition 1.30 (Truth-functional)**.** A binary connective $\#$ is *truth-functional* means that $p\#q$ has truth value which is a function of (i.e., is completely determined by) the truth values of p and q.

Example 1.31. The connectives "…and …" and "…or …" are both truth-functional.

Non-Example 1.32. The *counterfactual* connective "if it had been the case that p, then it would have been the case that q" is not truth-functional. To see this, note that p being false permits q to be whatever it wants without assigning a truth value depending on p and q. Here are some examples.

- Consider "if I had overslept, then the speaker gave her lecture." In fact, I did not oversleep, and the speaker would give her lecture anyways, so this is p being false and q being true. Here, the counterfactual is in total true.
- Consider "if I had overslept, then I would have arrived late." In fact, I did not oversleep, and in fact I would have arrived on time in this case, so this is p being false and q being true. Here the counterfactual is in total false.

1.3.2 The Material Conditional

Let's talk about conditionals; they are potentially confusing. The motivation here is that "if ... then ..." is used mathematically quite often, so we are going to formalize this. The following is our truth table.

Definition 1.33 (Material conditional ($\to$))**.** Given two propositions p and q, we define the *material conditional* read as "if p then q" is defined by the following truth table.

p	q	if p, then q
T	T	T
T	F	F
F	T	T
F	F	T

Mnemonically, the only way to make an if-then statement false is for the premise to be true and the conclusion to be false.

Warning 1.34. The statement "if I were 90, then I would be king" is a true statement because the premise is false. However, natural language would dictate this would probably not be accepted as true.

Example 1.35. The only way for Goldbach's conjecture (all even numbers greater than 2 are the sum of two prime numbers) to be false is for there to exist an even number which is not the sum of two prime numbers.

Notably, we do not falsify by showing the existence of odd numbers (such as 11) which are not the sum of two prime numbers. Explicitly, the statement

$$\text{If } 11 \text{ is an even number greater than } 2, \text{ then } 11 \text{ is the sum of two primes.}$$

is true because the premise is false.

Remark 1.36. We can think of the material conditional $p \to q$ as $(\neg p) \vee q$. Namely, as long as p is false or q is true, then $p \to q$ will be true (and conversely).

In this course, we will focus on truth-functional connectives.

1.3.3 Validity and Soundness

For the previous classes, we have been talking about what "good" arguments feel like. The technical version of this is "valid."

Definition 1.37 (Valid)**.** A form of argument is *valid* if, no matter the truth values of the propositions, whenever the premises are true, the conclusion will also be true.

Example 1.38. The following argument form is valid.

1. p or q.
2. Not p.
3. Therefore, q.

Explicitly, all cases where the first two premises hold require q to be true. (In fact, the only way for this to be true is for q to be true and p to be false.)

We can explicitly plug in to the above "argument form" to generate a real, physical argument.

Example 1.39. The following argument is of the above form.

1. Paris is the capital of France or Berlin is the capital of Belgium.
2. It is not the case that Paris is the capital of France.
3. Therefore, Berlin is the capital of Belgium.

Note that this argument is valid, even though the second premise is simply false. Giving false premises provides no guarantees that our conclusion is true, and indeed, the conclusion is false.

Take note that many arguments take the given form.

Warning 1.40. We apply the word "valid" to forms of arguments, not actual arguments. This prevents confusion about the truth-values of the actual premises.

To deal with the above confusion, we have the following definition.

Definition 1.41 (Sound)**.** If an argument is of valid form, and its premises are true (!), then the argument is *sound*.

Importantly, note that soundness is applies to arguments while validity applies to argument forms.

Example 1.42. The following argument is sound.

1. The number 527 is prime, or the number 527 is composite.
2. It is not the case that the number 527 is prime.
3. Therefore, the number 527 is composite.

Non-Example 1.43. The following argument from earlier is not sound.

1. Paris is the capital of France or Berlin is the capital of Belgium.
2. It is not the case that Paris is the capital of France.
3. Therefore, Berlin is the capital of Belgium.

Notably, the second premise here is false, so even though the argument has valid form, the entire argument is no longer sound.

In this case, we will mostly not care too much about soundness because we don't want to consider the truth-values of premises. Our job is to describe how to reason, which means we care more about valid argument forms than actually sound arguments.

Remark 1.44. We do not say that arguments are "true" or "false" because those words belong to propositions in this class. We will only say "sound" or "unsound."

Let's see some more examples.

Non-Example 1.45. The following is an invalid form of argument.

1. It is not the case that (p and q).
2. Therefore, it is not the case that p.

To see that this is invalid, we note that it's possible for p to be true while q is false. Explicitly, take p to be "5 is prime" and q to be "4 is prime." Here, the premise is true (not both 4 and 5 are prime), but the conclusion is false (5 is actually prime.)

However, we could get lucky. For example, in the above argument form, if we swap the roles of p and q, then the conclusion (4 is not prime) will be true. This is an invalid argument still producing true conclusions; the issue is that we are not guaranteed true conclusions from true premises and invalid arguments.

1.3.4 Truth Tables

We would like to have a more mechanical procedure to check the validity of arguments. For this class, we are restricting our attention to the truth-functional case, which means that we directly compute everything depending on the truth values of the propositions. This computation is typically organized into a truth table.

Here is an example.

Exercise 1.46. The following argument form is valid.

1. p or q.
2. It is not the case that p.
3. Therefore, q.

Proof. The only possibilities for p and q are to be true or false in various combinations. We run through all possible cases in the following table.

p	q	p or q	It is not the case that p	q
T	T	T	F	T
T	F	T	F	F
F	T	T	T	T
F	F	F	T	F

From this table we see that the only possibility where all the premises (i.e., both "p and q" and "It is not the case that p") are true is when p is false and q is the highlighted row, and in this case we do see that q is true. Thus, the argument is valid: all cases where the premises are true also happen to have that the conclusion is true. ■

Remark 1.47. An alternate way to do this computation would be to note that the truth table computation comes down to showing the single formula

$$((p \vee q) \wedge \neg p) \to q$$

is always true, independent of the truth values of p and q.

The previous example is a bit misleading because the simple example had only one row in which all premises are true: in general we must check all rows which have the conclusion true. Here is another example.

Exercise 1.48. The following argument form is valid.

1. p.
2. q or r.
3. Therefore, (p and q) or (q and r).

Proof. Here is our truth table.

p	q	r	p	q or r	p and q	p and r	(p and q) or (p and r)
T	T	T	T	T	T	T	T
T	T	F	T	T	T	F	T
T	F	T	T	T	F	T	T
T	F	F	T	F	F	F	F
F	T	T	F	T	F	F	F
F	T	F	F	T	F	F	F
F	F	T	F	T	F	F	F
F	F	F	F	F	F	F	F

The first three rows are the only ones where all the premises are true, and from there we can see that in all these rows the conclusion is true. (In fact, those are exactly the rows where the conclusion is true.) ■

1.4 January 26

Welcome back everyone.

1.4.1 Validity and Truth Tables

Today we're doing some formal propositional logic. It should be fun. Last time we were discussing validity of an argument form, which means that whenever the premises are true, the conclusion will also be true.

Example 1.49. Validity can be counterintuitive. For example, the following argument form is valid.

1. p.
2. $\neg p$.
3. Therefore, q.

This argument form is in fact valid because every time that the premises are true (which happens to be never) also has the conclusion true.

Let's also give an example of an invalid argument.

Exercise 1.50. The following argument is invalid.

1. It is not the case that (p and q).
2. Therefore, it is not the case that p.

Proof. It suffices to give one row of the truth table with true premises but false conclusion. Here it is.

p	q	p and q	not (p and q)	not p
T	F	F	T	F

This finishes because we have found a way to make the premise "not (p and q) true but "not p" false. ■

Importantly, for invalidity it suffices to give the single inconsistent row, though potentially one might have to compute all rows before finding the inconsistent one.

Warning 1.51. However, for the validity check, one does need to write out the full truth table in order to be sure that one has found all cases where the premises are true.

Note that we don't actually care about the rows of the truth table where at least one of the premises is false, but we must include them in order to be sure that we don't care about them. Anyways, in the future we will have other ways to check validity, but for now this is all that we have.

1.4.2 Symbology

Our goal is to define a language of propositional logic. Here are the ideas.

Idea 1.52. To create a formula in propositional logic, we have three ingredients:

1. We will work abstract propositions as "letters" $\{p, q, \ldots\}$.
2. We will change natural language connectors to symbolic ones, essentially to shorten things.
3. We will allow extra connections between formulae by using parentheses in cases of ambiguity.

Note that we have already done the first step above when we moved from concrete arguments to the more abstract argument forms. For the second step, we introduce the following symbols.

English	Symbology	Name
not p	$\neg p$	*negation* of p
p and q	$p \wedge q$	*conjugation* of p and q
p or q	$p \vee q$	*disjunction* of p or q
if p, then q	$p \to q$	*material conditional* with *antecedent* p and *conditional* q
p if and only if q	$p \leftrightarrow q$	*biconditional*

And our third step is more or less automatic after allowing parentheses into our language. For example, $p \to (q \wedge r)$ is a valid formula.

We can translate from English to our formal language with a little of effort.

Example 1.53. Goldbach's conjecture, that

> If n is an integer and n is even and n is greater than 2, then n is the sum of two prime numbers

can be translated into a formula as follows: let p be the proposition that n is an integer, q that n is even, r that n is greater than 2, and s is the sum of two prime numbers. Then the above becomes

$$(p \wedge q \wedge r) \to s.$$

Remark 1.54. Later in life we will even be able to talk about a proposition via predicate logic, in which case we can let $E(n)$ be true if and only if n is even, and $P(n)$ be true if and only if n is prime. Then Goldbach's conjecture becomes

$$\forall n((E(n) \wedge (n > 2)) \to \exists p_1 \exists p_2 (P(p_1) \wedge P(p_2) \wedge (n = p_1 + p_2))).$$

Anyways, with the above discussion, we can now fix the symbols of our language.

Definition 1.55 (Symbols)**.** For propositional logic, we have the following propositional symbols.

- A set of propositional symbols $\{p_1, p_2, \ldots\}$.
- A set of connectives $\{\neg, \wedge, \vee, \to, \leftrightarrow\}$.
- A set of parentheses $\{(,)\}$.

1.4.3 Towards Formulae

We will be defining our formulae by building them up inductively. This inductive process is formalized by "expressions."

Definition 1.56 (Expression)**.** An *expression* is any finite sequence of symbols $\langle s_1, s_2, \ldots, s_n \rangle$ of our language.

Not all of these expressions are good.

Example 1.57. The sequence $\langle \to, \neg, (p \rangle$ is an expression.

We can also "concatenate" expressions together by just stringing one after another. If we have three expressions e_1 and e_2 and e_3, then it is not too hard to see that

$$e_1(e_2e_3) = (e_1e_2)e_3,$$

where the implicit operation is concatenation.

By convention, we will avoid writing the $\langle$ and $\rangle$ as well as commas from our expressions as much as possible.

Example 1.58. We will write "$\to \neg p($" for the expression $\langle \to, \neg, (p \rangle$.

Now, again not all expressions make grammatical sense, so we would like to define which expressions actually have some meaning, and they will be our formulae.

Definition 1.59 (Formula, I)**.** A *formula* is an expression with some meaning. We will define $\mathcal{L}(P)$ to be the set of these well-formed expressions.

Notably we so far still have not described how we are going to build these formulae. Here is a first attempt.

Definition 1.60 (Formula, II)**.** The set of *formulae* $\mathcal{L}(P)$ is defined as a subset of expressions as follows.

- Any propositional p in P makes an "atomic" formula "p."
- Unary connective: if $\varphi \in \mathcal{L}(P)$ is a formula, then $\neg\varphi$ is a formula.
- Binary connectives: if $\varphi_1, \varphi_2 \in \mathcal{L}(P)$ is a formula, then both $(\varphi_1 \wedge \varphi_2)$ and $(\varphi_2 \vee \varphi_2)$ and $(\varphi_1 \to \varphi_2)$ and $(\varphi_1 \leftrightarrow \varphi_2)$ are formulae.
- There are no other formulae than these.

Example 1.61. Here are some examples.

- All propositions are atomic formulae.
- Given a proposition p, "$\neg p$" is a formula. In fact, $\neg\neg p$ will also be a formula.
- Given propositions p and q, $(p \wedge q)$ is a formula. In fact, from here it follows that $(p \to (p \wedge q))$ is a formula. We can continue this process.

The issue with Definition 1.60 is its rigor in the last point: it is not completely clear how to reduce the set of formulae to exclude other formulae. For example, it is a bit annoying to prove that some expression is not a formula.

Example 1.62. All the rules in our definition preserve that there must be a proposition in a formula, so we can immediately say that the expression "$\neg()$" is in fact not a formula.

Next time we will make precise our definition of a formula. Later on we will give them some meaning ("semantics"), but for now we are just arguing about syntax.

1.5 January 28

Here we go.

1.5.1 Formulae

Our story today continues trying to define what a grammatical formula is in our language. Last time we gave an almost-precise definition of formula; today we will formalize it.

The idea is that complex formulae can be built from simpler ones. Namely, we bring in the idea of "construction sequences."

Example 1.63. We can build the formula $(\neg p_1 \to (p_2 \vee p_3))$ by the construction sequence

$$\langle p_1, \neg p_1, p_2, p_3, (p_2 \vee p_3), (\neg p_1 \to (p_2 \vee p_3))\rangle.$$

The point is that each formula in the list is made via some concatenation rule applied to previous formulae in the list.

Here is our definition.

Definition 1.64. A *construction sequence* from a set of letters P is a finite sequence of expressions $\langle \varphi_1, \dots, \varphi_n \rangle$ satisfying the following conditions.

(a) φ_k may be an atomic formula in P.

(b) If $k < \ell$, then φ_ℓ may be $\neg\varphi_k$.

(c) If $k, \ell < m$, then φ_m may be $(\varphi_k \wedge \varphi_\ell)$ or $(\varphi_k \vee \varphi_\ell)$ or $(\varphi_k \to \varphi_\ell)$ or $(\varphi_k \leftrightarrow \varphi_\ell)$.

We repeat the same example.

Example 1.65. Let's build our previous construction sequence.

- We start with $\langle p_1 \rangle$.
- From here we can get $\langle p_1, \neg p_1 \rangle$.
- From here we may add many atomic formulae, giving $\langle p_1, \neg p_1, p_2, p_3 \rangle$.
- With p_2 and p_3, we can build to $\langle p_1, \neg p_1, p_2, p_3, (p_2 \vee p_3) \rangle$.
- With $\neg p_1$ and $(p_2 \vee p_3)$, we can build the full $\langle p_1, \neg p_1, p_2, p_3, (p_2 \vee p_3), (\neg p_1 \to (p_2 \vee p_3)) \rangle$.

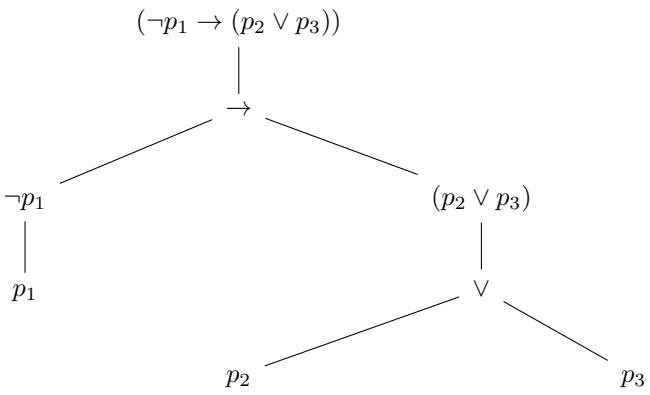
This is a construction sequence for $(\neg p_1 \to (p_2 \vee p_3))$.

The point is that construction sequences let us define formulae.

Definition 1.66 (Formula)**.** An expression φ is a *formula* of $\mathcal{L}(P)$ if and only if it is it an element of some construction sequence from P.

Note that there are infinitely many expressions.

Here is an alternate way to think about this construction sequence: we can build it as a tree.



Observe that this really does convey the same information.

1.5.2 Formulae, Again

Last time we defined what it means to "construct" a formula by manually describing how to construct formulae. We now provide a separate way to do this, which will be helpful for proofs later.

The key here is to view construction steps as "operations" on $\mathcal{L}(P)$. For example, $o_\neg$ is a function which takes the formula φ and returns $\neg\varphi$. In general, for some connector $\#$ such as in $\{\neg, \wedge, \vee, \to, \leftrightarrow\}$, we can write

$$o_\#(\varphi, \varphi') \coloneqq (\varphi \# \varphi').$$

Now we can provide an "inductive" definition of our formulae.

Definition 1.67 (Formula, again)**.** The set of formulae $\mathcal{L}(P)$ to be a set of expressions satisfying the following.

(a) $\mathcal{L}(P)$ contains all atomic formulae from P.

(b) $\mathcal{L}(P)$ is closed under the operations $o_\neg, o_\wedge, o_\vee, o_\to, o_\leftrightarrow$.

(c) $\mathcal{L}(P)$ is minimal with respect to the properties (a) and (b).

Non-Example 1.68. The set

$$S \coloneqq \{p, \neg p, \neg\neg p, \dots\}$$

is closed under $o_\neg$ because appending a $\neg$ to any element in S stays in S. However, this set is not closed under $o_\wedge$ because $p \in S$ but $(p \wedge p) \notin S$.

Remark 1.69. Closure is a nice property. For example, $\mathbb{N}$ is closed under $+$ and $\times$. If we want it to be closed under $-$ (say), we should introduce $\mathbb{Z}$ instead of $\mathbb{N}$.

Note that the conditions (a) and (b) in the definition is too much.

Non-Example 1.70. The set of all possible expressions certainly satisfies (a) and (b), but it has many expressions which we don't want to call formulae, such as $\neg\neg) \, p$.

One possible issue with (c) in Definition 1.67 is that it is not obvious what it means, and once we agree what it means, it's not obvious that $\mathcal{L}(P)$ actually exists as a set. Well, we choose "minimal" to mean that any set S satisfying the properties (a) and (b) immediately implies $\mathcal{L}(P) \subseteq S$.

Thus, to verify that Definition 1.67 does have some $\mathcal{L}(P)$ which exists, we can define

$$\mathcal{L}(P) = \bigcap_{S \text{ satisfies (a), (b)}} S.$$

Then it is not hard to check that $\mathcal{L}(P)$ satisfies (a)—P is a subset of each set we intersect, so $P \subseteq \mathcal{L}(P)$—as well as satisfies (b)—if φ and φ' are two formulae in all sets satisfying (b), then after applying the operation they will still be in all sets satisfying (b).

Anyways, Definition 1.67 is called an "inductive" definition because it will turn out that it gives us an induction: if we want to show that some property holds for all formulae $\mathcal{L}(S)$, we show that the property holds for all propositions and that the property is closed under the operation.

Example 1.71. The natural numbers $\mathbb{N}$ are minimal with respect to $0 \in \mathbb{N}$ and closure under $+1$; its existence can be guaranteed via the same intersection idea. This means that any set containing 0 and closed under $+1$ will contain $\mathbb{N}$.

Remark 1.72. Observe that Definition 1.67 is a bit weird: it's saying an expression φ is a formula if and only if it is a member of each set satisfying (a) and (b). This makes it a little harder to prove that something is a formula because this approach is more "top-down."

Example 1.73. We claim that $(p \wedge q)$ is a formula. Well, let S be some set of expressions satisfying (a) and (b), and we will show $(p \wedge q) \in S$. Then $p, q \in S$ by (a), from which (b) implies $(p \wedge q) \in S$.

Notice how similar the proof in the above example is to actually giving the construction sequence $\langle p, q, (p \wedge q)\rangle$.

1.6 January 31

So class is in-person today. The lecture hall is very large and quite empty.

1.6.1 Formula Induction

Big-picture, we are focusing on symbolic logic. The main point of introducing our formal language right now is that it will help us formally define what rigorous reasoning is and how it works.

Today we're going to discuss induction on formulae, which will be the main technique to show that some property holds for all formulae. Namely, we will be using Definition 1.67 in order to do out proofs.

Remark 1.74. We never actually showed that Definition 1.66 and Definition 1.67 are equivalent, but they are. The annoying thing to do is to check that any set satisfying Definition 1.66 will automatically satisfy Definition 1.67.

Let's do an example proof.

Proposition 1.75. All expressions in $\mathcal{L}(P)$ have the same number of left and right parentheses.

Proof. We proceed by induction. Let S be the set of expressions that have the same number of left and right parentheses, and we show that $\mathcal{L}(P) \subseteq S$. We have the following two checks.

(a) Each atomic expressions $p \in P$ have no parentheses at all, so $p \in S$.

(b) Suppose that $\varphi, \psi \in S$; suppose φ has x left parentheses (and therefore x right parentheses) and ψ has y left parentheses (and therefore y right parentheses). Then we check what happens with our connectives.

- We see that $\neg\varphi$ has the same number left/right parentheses as φ, so these quantities are equal, so $\neg\varphi \in S$.
- We see that $(\varphi \vee \psi)$ and $(\varphi \wedge \psi)$ and $(\varphi \to \psi)$ and $(\varphi \leftrightarrow \psi)$ have $x+y+1$ left and right parentheses, so we are done.

Because $\mathcal{L}(P)$ is the minimal set satisfying (a) and (b), it follows $\mathcal{L}(P) \subseteq S$, so we are done. ∎

An alternate way to see that satisfying (a) and (b) is enough is because we can build any formula in $\mathcal{L}(P)$ by using the steps (a) and (b), by definition of $\mathcal{L}(P)$. The minimality of $\mathcal{L}(P)$ is a way to formalize this intuition.

Remark 1.76. The part where we assume $\varphi, \psi \in S$ before checking $\neg\varphi$ and $(\varphi \vee \psi)$ and its friends is called the "inductive hypothesis." It is very important: if $\varphi = (p \notin S$, then in fact $\neg\varphi = \neg(p \notin S$ as well.

Here is another example: we show that $\neg \to p$ is not a formula, via the following lemma.

Lemma 1.77. Every formula neither starts with $\neg \to$ nor with $\to$.

Proof. Let S be the set of expressions that neither start with $\neg \to$ nor with $\to$. We now induct.

(a) No atomic formula starts with $\neg \to$ nor with $\to$.

(b) If we have a formula φ not starting with $\neg \to$ nor with $\to$, then $\neg\varphi$ will not start with $\neg \to$ because φ did not start with $\to$.

On the other hand, if φ and ψ neither start with $\neg \to$ nor with $\to$, then (for any binary connector $\#$) $(\varphi \# \psi)$ will start with a parenthesis and not with $\neg \to$ nor with $\to$.

Thus, because $\mathcal{L}(P)$ is minimal satisfying (a) and (b), it follows $\mathcal{L}(P) \subseteq S$, so we are done. ■

Here is another result, which we can show without much effort by induction.

Lemma 1.78. If we have letters $P \subseteq Q$, then $\mathcal{L}(P) \subseteq \mathcal{L}(Q)$. We will not prove this here, but it is just another induction.

The point of this is to say that a single formula can belong to multiple languages. For example,

$$(p_1 \wedge p_2) \in \mathcal{L}(\{p_1, p_2\}) \cap \mathcal{L}(\{p_1, p_2, p_3\}).$$

The proof of the lemma is by induction and not hard, so we will omit it.

1.6.2 Subformula

Here is a motivating example.

Example 1.79. The formula $\varphi = (\neg p_1 \to \neg(p_2 \vee p_3))$ has the following subformulae.

$$p_1, \quad \neg p_1, \quad p_2, \quad p_3, \quad (p_2 \vee p_3), \quad \neg(p_2 \vee p_3), \quad (\neg p_1 \to (\neg(p_2 \vee p_3))).$$

The set of proper formulae is the same set except for φ.

Non-Example 1.80. The formula $(p_1 \vee p_2)$ is not a subformula of $(p_1 \wedge p_2)$.

We would like to define a function $\mathrm{sub} : \mathcal{L}(P) \to \mathcal{P}(\mathcal{L}(P))$. For this, we will create a recursive definition for sub.

Definition 1.81 (Subformula)**.** Given a formula $\varphi \in \mathcal{L}(P)$, we define the *subformulae* $\mathrm{sub}(\varphi)$ to be defined as follows.

- $\mathrm{sub}(p) \coloneqq \{p\}$ for each atomic formula $p \in P$.
- $\mathrm{sub}(\neg\varphi) = \mathrm{sub}(\varphi) \cup \{\neg\varphi\}$.
- $\mathrm{sub}((\varphi \# \psi)) = \mathrm{sub}(\varphi) \cup \mathrm{sub}(\psi) \cup \{(\varphi \# \psi)\}$ for any two formulae $\varphi, \psi \in \mathcal{L}(P)$ and binary connective $\# \in \{\wedge, \vee, \to, \leftrightarrow\}$.

The above definition is called "recursive" because we did not explicitly define it but only how to do this by breaking down connectives. We can show that sub is defined on $\mathcal{L}(P)$ by yet another induction which we will omit.[2]

Let's see some examples.

Example 1.82. We have the following computation.

$$\begin{aligned}
\mathrm{sub}(\neg(p \wedge q)) &= \mathrm{sub}((p \wedge q)) \cup \{\neg(p \wedge q)\} \\
&= \mathrm{sub}(p) \cup \mathrm{sub}(q) \cup \{(p \wedge q)\} \cup \{\neg(p \wedge q)\} \\
&= \{p\} \cup \{q\} \cup \{(p \wedge q)\} \cup \{\neg(p \wedge q)\} \\
&= \{p, q, (p \wedge q), \neg(p \wedge q)\}.
\end{aligned}$$

[2] By definition, sub is defined on atomic formulae, and the domain is closed under connectives, so sub is defined on $\mathcal{L}(P)$.

Example 1.83. We have the following computation.

$$\begin{aligned}\operatorname{sub}((p \wedge p)) &= \operatorname{sub}(p) \cup \operatorname{sub}(p) \cup \{(p \wedge p)\} \\ &= \{p\} \cup \{p\} \cup \{(p \wedge p)\} \\ &= \{p, p \wedge p)\}.\end{aligned}$$

Note that the union of the two sets has killed the duplicate p.

Example 1.84. We have the following computation.

$$\begin{aligned}\operatorname{sub}((\neg p_1 \to p_2)) &= \operatorname{sub}(\neg p_1) \cup \operatorname{sub}(p_2) \cup \{(\neg p_1 \to p_2)\} \\ &= \operatorname{sub}(p_1) \cup \{\neg p_1\} \cup \operatorname{sub}(p_2) \cup \{(\neg p_1 \to p_2)\} \\ &= \{p_1\} \cup \{\neg p_1\} \cup \{p_2\} \cup \{(\neg p_1 \to p_2)\} \\ &= \{p_1, \neg p_1, p_2, (\neg p_1 \to p_2)\}.\end{aligned}$$

It is true that the definition of "subformula" feels intuitively obvious, but we have given the above rigorous definition to introduce the idea of a recursive definition.

Let's discuss the idea of recursion a little more closely because it will come up again.

- Define a set S inductively as the smallest set containing some base objects and closed under some operations. We will also require the following coherence conditions.[3]
  - We will require that the operations never output the same formula. For example, for two formulae φ and ψ, we have
    $$(\varphi \wedge \psi) \neq (\varphi \vee \psi).$$
  - No operation can output a base object. For example, no operation can output an atomic formula because operations will always start with $\neg$ or a parenthesis.
- Then to define a function f inductively, we need to define $f(b)$ for each base object $b \in S$ and provide some function g such that
  $$f(o(a_1, \ldots, a_n)) = g_o(f(a_1), \ldots, f(a_n), a_1, \ldots, a_n)$$
  for each n-ary operation o. For example, we had
  $$\operatorname{sub}(o_\neg(\varphi)) = g_\neg(\operatorname{sub}(\varphi), \varphi),$$
  where $g_\neg(S, \varphi) \coloneqq S \cup \varphi$.

Let's do another example recursive definition.

Definition 1.85 (Depth)**.** We define the *depth* of a formula $\varphi \in \mathcal{L}(P)$ to be given be a function $\operatorname{depth} : \mathcal{L}(P) \to \mathbb{N}$ defined as followed.

- $\operatorname{depth}(p) = 0$ for each atomic formula $p \in P$.
- $\operatorname{depth}(\neg\varphi) = \operatorname{depth}(\varphi) + 1$ for each $\varphi \in \mathcal{L}(P)$. In other words, adding $\neg$ adds one level of depth.
- $\operatorname{depth}((\varphi \# \psi)) = \max\{\operatorname{depth}(\varphi), \operatorname{depth}(\psi)\} + 1$ for each $\varphi, \psi \in \mathcal{L}(P)$ and binary connective $\#$. In other words, the depth of $(\varphi \# \psi)$ is one more than the larger of the two depths of φ and ψ.

[3] The coherence conditions ensure that the function we create is well-defined.

Example 1.86. We have the following computation.

$$\begin{aligned}\operatorname{depth}((\neg p_1 \to p_2)) &= \max\{\operatorname{depth}(\neg p_1), \operatorname{depth}(p_2)\} + 1 \\ &= \max\{\operatorname{depth}(p_1) + 1, \operatorname{depth}(p_2)\} + 1 \\ &= \max\{0 + 1, 0\} + 1 \\ &= 2.\end{aligned}$$

We can see that this definition is indeed recursive: we computed it as 0 for the basic objects, and then we had

$$f(o_\neg(\varphi)) = g_\neg(f(\varphi), \varphi),$$

where $g_\neg(n, \varphi) = n + 1$, and

$$f(o_\#(\varphi, \psi)) = g_\#(f(\varphi), f(\psi), \varphi, \psi),$$

where $g_\#(n, m, \varphi, \psi) = \max\{n, m\} + 1$.

1.7 February 2

Here we go again.

1.7.1 Semantics for Connectives

Last time we finished our discussion of creating the language of propositional logic. Today we will actually give formulae meaning as a "truth" function of the propositions.

Example 1.87. We hinted for $\neg p$ to mean "not p."

Our goal is to give precise definitions to all of our symbols, in a mathematically precise way.

The way we do this is by "truth-conditional semantics." Namely, we will give meaning to certain formulae by describing when a formula is true or false.

Example 1.88. We can compute the meaning of $\neg p$ by the following truth table.

p	$\neg p$
T	F
F	T

This truth table fully captures what $\neg$ should mean.

Convention 1.89. For the remainder of the class, we will replace "T" with 1 and "F" with 0.

Example 1.90. We have that the truth value of $\neg p$ is $1-$ the truth value of p.

So using numbers will have some utility in this class.

Example 1.91. We have that

p	q	$p \wedge q$
1	1	1
1	0	0
0	1	0
0	0	0

In short, we can verify that $p \wedge q$ has truth value equal to the minimum of the truth values of p and q.

We could also say that $p \wedge q$ has truth value equal to the product of the truth values of p and q. However, we do not do this to make better analogy with $\vee$, as follows.

Example 1.92. We have that

p	q	$p \vee q$
1	1	1
1	0	1
0	1	1
0	0	0

In short, we can verify that $p \wedge q$ has truth value equal to the maximum of the truth values of p and q.

Let's wrap up with the last two connectives.

Example 1.93. We have that

p	q	$p \to q$
1	1	1
1	0	0
0	1	1
0	0	1

We can say this as the truth value of $p \to q$ is the indicator for when the truth value of p is less than or equal to the truth value of q.

Example 1.94. We have that

p	q	$p \leftrightarrow q$
1	1	1
1	0	0
0	1	0
0	0	1

Now, we can say this as the truth value of $p \to q$ is the indicator for when the truth value of p is equal to the truth value of q.

1.7.2 Semantics for Formulae

The key idea to extending semantics to all formulae is to use a recursive definition, using the connectives to build arbitrary semantics for formulae. We will need to be somewhat technical for this.

Definition 1.95 (Valuation)**.** A *valuation* V for P is a function $V : P \to \{0, 1\}$ which assigns each proposition $p \in P$ a value of 0 or 1.

Example 1.96. There is a valuation $V : P \to \{0, 1\}$ which gives $V(p) = 1$ for each $p \in P$. There is also a valuation $V : P \to \{0, 1\}$ which gives $V(p) = 0$ for each $p \in P$.

Example 1.97. The following describes a valuation for $P = \{p, q\}$.

p	1
q	0

Remark 1.98. The "valuation" $\hat{V}$ is meant to be an abbreviation of "the truth value of."

Intuitively, a valuation is a description of the world of P: maybe $p_1 \in P$ should be true with p_2 false, maybe something else.

Remark 1.99. If P is finite, then there are only finitely many valuations. Precisely, there are $2^{\#P}$ total valuations: each proposition $p \in P$ has two options (namely, 0 or 1), and we have to make $\#P$ total choices (simultaneously) to determine a unique valuation V.

Now here is our promised idea.

Idea 1.100. Recursion can extend any valuation $V : P \to \{0,1\}$ uniquely to all of $\hat{V} : \mathcal{L}(P) \to \{0,1\}$.

Example 1.101. Once we know $V(p)$ and $V(q)$, we get $V((p \wedge q))$ for free.

Let's give the recursive definition for $\hat{V}$; it works as follows.

- For each atomic formula $p \in P \subseteq \mathcal{L}(P)$, we have $\hat{V}(p) \coloneqq V(p)$. (This is our base case.)
- We have $\hat{V}(\neg\varphi) \coloneqq 1 - \hat{V}(\varphi)$.
- We have $\hat{V}((\varphi \wedge \psi)) \coloneqq \min\{\hat{V}(\varphi), \hat{V}(\psi)\}$.
- We have $\hat{V}((\varphi \vee \psi)) \coloneqq \max\{\hat{V}(\varphi), \hat{V}(\psi)\}$.
- We have $\hat{V}((\varphi \to \psi)) \coloneqq 1_{\hat{V}(\varphi) \le \hat{V}(\psi)}$.
- We have $\hat{V}((\varphi \leftrightarrow \psi)) \coloneqq 1_{\hat{V}(\varphi) = \hat{V}(\psi)}$.

A standard induction can show that the domain of $\hat{V}$ is indeed $\mathcal{L}(P)$. In particular, the domain of $\hat{V}$ contains all the letters $p \in P$ and is closed under our connectives.

Remark 1.102. We could write the above discussion by writing out the full truth tables for each connective, but the above is arguably a cleaner connective.

Now that we've extended V, we can make terminology for when formulae might be true or false.

Definition 1.103 (Satisfies)**.** A valuation $V : P \to \{0,1\}$ *satisfies* $\varphi \in \mathcal{L}(P)$ if and only if $\hat{V}(\varphi) = 1$. We will notate this by $V \vDash \varphi$.

Intuitively, the world that V describes makes the formula φ true.

Let's do an example.

Exercise 1.104. Fix $P = \{p, q, r\}$ and V a valuation by $V(p) = V(q) = 1$ and $V(r) = 0$. We determine if V satisfies $(p \wedge (r \to q))$.

Proof. We do the computation by hand.

$$\begin{aligned}
\hat{V}(((p \wedge (r \to q)))) &= \min\{\hat{V}(p), \hat{V}((r \to q))\} \\
&= \begin{cases} \min\{\hat{V}(p), 1\} & \hat{V}(r) \le \hat{V}(q), \\ \min\{\hat{V}(p), 0\} & \hat{V}(r) \le \hat{V}(q), \end{cases} \\
&= \min\{\hat{V}(p), 1\} \\
&= \min\{1, 1\} \\
&= \boxed{1}.
\end{aligned}$$

■

The above approach felt more top-down as an evaluation, in contrast to the ore bottom-up a typical truth-table computation, as follows.

p	q	r	$r \to q$	$p \wedge (r \to q)$
1	1	0	1	1

1.7.3 Validity, Again

We would like to use our precise definitions for logic to recreate our definitions for "valid."

We start with a small remark.

Remark 1.105. Suppose $\varphi \in \mathcal{L}(P_1) \cap \mathcal{L}(P_2)$ for some proposition letters P_1 and P_2. (Namely, the letters of φ live in both P_1 and P_2.) Then if $V_1 : P_1 \to \{0,1\}$ and $V_2 : P_2 \to \{0,1\}$ are valuations such that each p present in φ has $V_1(p) = V_2(p)$, then

$$\hat{V}_1(\varphi) = \hat{V}_2(\varphi).$$

To formalize this, we would need a notion of which propositions are present in φ, which we could define as a function recursively. We will not bother to do this here.

Here is our definition of an argument form.

Definition 1.106 (Argument form)**.** An *argument form* in $\mathcal{L}(P)$ is a pair of *premises* $\{\varphi_1, \dots, \varphi_n\} \subseteq \mathcal{L}(P)$ and a *conclusion* $\psi \in \mathcal{L}(P)$.

And here is our definition of validity.

Definition 1.107 (Valid)**.** An argument form $(\{\varphi_1, \dots, \varphi_n\}, \psi)$ is *valid* if and only if each valuation $V : P \to \{0,1\}$ has

$$\hat{V}(\varphi_1) = \cdots = \hat{V}(\varphi_n) = 1 \implies \hat{V}(\psi) = 1.$$

We notate this by $\varphi_1, \dots, \varphi_n \vDash \psi$ and say "ψ is a *consequence* of $\{\varphi_1, \dots, \varphi_n\}$."

Example 1.108. We have that $(p \vee q), \neg q \vDash p$.

Intuitively, an argument form is valid if, when the premises are true, the conclusion is also true. This is essentially the definition that we wanted.

Remark 1.109. One can extend this definition to work with infinitely many premises, but we will not do this here.

Observe that checking validity is a finite matter because, to check $\varphi_1, \dots, \varphi_n \vDash \psi$, we only need to consider the finitely many propositions that take place in any of $\varphi_1, \dots, \varphi_n, \psi$. And then each proposition has only two options, so in total we are safely in a finite universe.

1.8 February 4

Here we go.

1.8.1 Validity for Argument Forms

Last time we left off talking about the validity of an argument form. We had the following definition.

Definition 1.110 (Valid)**.** An argument form $(\{\varphi_1, \dots, \varphi_n\}, \psi)$ is *valid* if and only if each valuation $V : P \to \{0,1\}$ has

$$\hat{V}(\varphi_1) = \cdots = \hat{V}(\varphi_n) = 1 \implies \hat{V}(\psi) = 1.$$

We notate this by $\varphi_1, \dots, \varphi_n \vDash \psi$ and say "ψ is a *consequence* of $\{\varphi_1, \dots, \varphi_n\}$."

Remark 1.111. Importantly, a computer could always decide validity in finite time, essentially by making some huge truth table. In other words, determining validity is "decidable."

Let's see an example.

Exercise 1.112. We show that $\{\neg(p \wedge q), p\} \vDash \neg q$.

Proof. We reason using the equations for the connectives. Namely, suppose $V : \{p, q\} \to \{0,1\}$ is a valuation such that $\hat{V}(\neg(p \wedge q)) = \hat{V}(p) = 1$. Now, we see

$$1 = \hat{V}(\neg(p \wedge q)) = 1 - \hat{V}((p \wedge q)) = 1 - \min\{\hat{V}(p), \hat{V}(q)\}.$$

Thus, $\min\{\hat{V}(p), \hat{V}(q)\} = 0$, so one of $\hat{V}(p) = 0$ or $\hat{V}(q) = 0$. But $\hat{V}(p) = 1$, so $\hat{V}(q) = 0$ is forced instead. Thus,

$$\hat{V}(\neg q) = 1 - \hat{V}(q) = 1 - 0 = 1,$$

which is what we wanted. ∎

In fact, the above exercise shows the following.

Proposition 1.113. Fix $\alpha, \beta \in \mathcal{L}(P)$. Then $\{\neg(\alpha \wedge \beta), \beta\} \vDash \beta$.

Proof. Reuse the proof above. ∎

More generally, we have the following notion of substitution.

Proposition 1.114 (Substitution)**.** Suppose that $\{\varphi_1, \dots, \varphi_n\} \vDash \psi$ in $\mathcal{L}(P)$, then any substitution of the atomic formulae P by formulae in $\mathcal{L}(Q)$ (which induces a map $f : \mathcal{L}(P) \to \mathcal{L}(Q)$), then $\{f\varphi_1, \dots, f\varphi_n\} \vDash \psi$.

Proof. Induction to make f and then show the statement. ∎

Example 1.115. We know that

$$\{\neg((r \vee s) \wedge (s \to p)), (r \vee s)\} \vDash \neg(s \to p)$$

is valid by Proposition 1.113.

Here are some more examples.

Exercise 1.116. We show that $\{(p \to q), \neg p\} \nvDash \neg q$.

Proof. We set a valuation V by $V(p) = 0$ and $V(q) = 1$. Then

$$\hat{V}((p \to q)) = 1_{\hat{V}(p) \leq \hat{V}(q)} = 1_{0 \leq 1} = 1,$$

and $\hat{V}(\neg p) = 1 - \hat{V}(p) = 1 - 0 = 1$. Thus, our premises are true. But $\hat{V}(\neg q) = 1 - \hat{V}(q) = 1 - 1 = 0$ means the conclusion is true, so we are done. ∎

Remark 1.117. Propositional logic cannot see the validity of all arguments. For example, the following argument cannot be broken down into propositions in the ways that $\mathcal{L}(P)$ provides.

1. Every integer greater than 1 is a product of prime numbers.
2. 999 is an integer greater than 1.
3. Therefore, 999 is a product of prime numbers.

The issue is that we need quantifiers: we need a way to plug in 999 into "every integer," which propositional logic does not provide. We will fix this when we talk about predicate logic later in the course.

Let's run down some valid forms of argument. They can be proven from a similar logic to the above.

- Modus ponens: $\{\varphi \to \psi, \varphi\} \vDash \psi$.
- Modus tollens: $\{\varphi \to \psi, \neg\psi\} \vDash \neg\varphi$.
- Contraposition: $\{\varphi \to \psi\} \vDash \neg\psi \to \neg\varphi$.
- Disjunctive syllogism: $\{\varphi \vee \psi, \neg\varphi\} \vDash \psi$ and $\{\varphi \vee \psi, \neg\psi\} \vDash \varphi$.
- Hypothetical syllogism: $\{\varphi \to \psi, \psi \to \chi\} \vDash \varphi \to \chi$.
- Proof by cases: $\{\varphi \vee \psi, \varphi \to \chi, \psi \to \chi\} \vDash \chi$. The point is that either antecedent φ or ψ lead to the same conclusion χ.

1.8.2 Validity for Formulae

Here is our definition.

Definition 1.118 (Valid formula, tautology)**.** A formula $\varphi \in \mathcal{L}(P)$ is *valid* or a *tautology* if and only if each valuation $V : P \to \{0, 1\}$ has $\hat{V}(\varphi) = 1$.

Example 1.119. Given any formula $\varphi \in \mathcal{L}(P)$, then any valuation V has $\hat{V}(\varphi) = 1$ or $\hat{V}(\varphi) = 0$, so $\hat{V}(\varphi \vee \neg\varphi) = 1$ in all cases.

Here are some examples. The main point is that valid forms of arguments can turn $\vDash$ into $\to$ to make a tautology.

- $\varphi \to \varphi$.
- $\varphi \to (\varphi \vee \psi)$ and $\psi \to (\varphi \wedge \psi)$.
- $\varphi \to (\varphi \vee \psi)$ and $\psi \to (\varphi \vee \psi)$.
- "Modus ponens": $((\varphi \to \psi) \wedge \varphi) \to \psi$.
- $((\varphi \to \psi) \wedge \neg\psi) \to \neg\varphi$.
- $((\varphi \to \psi) \wedge (\psi \to \chi)) \to (\varphi \to \chi)$.
- $\varphi \to (\psi \to \varphi)$.
- Peirce's Law: $((\varphi \to \psi) \to \varphi) \to \varphi$.

The last one is not actually a tautology, but it is. The only care we have to worry about is when φ is true, but when φ is true, $(\varphi \to \psi) \to \varphi$ can only be false when $\varphi \to \psi$ can only be false when φ is false.

Remark 1.120. Another way to say that φ is a valid formula is that $\varnothing \vDash \varphi$. So we will notationally write $\vDash \varphi$.

Validity of arguments can be turned into validity of formulae, as follows.

Theorem 1.121 (Deduction)**.** An argument form $\{\varphi_1, \ldots, \varphi_n\}$ with conclusion ψ is valid if and only if $(\varphi_1 \wedge \ldots \wedge \varphi_n) \to \psi$ is valid. In other words, $\{\varphi_1, \ldots, \varphi_n\} \vDash \psi$ if and only if $\psi(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \psi$.

Proof. By definition, $(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \psi$ will be valid if and only if every valuation has V

$$1 = \hat{V}((\varphi_1 \wedge \cdots \wedge \varphi_n) \to \psi),$$

which is equivalent to $\hat{V}(\varphi_1 \wedge \cdots \wedge \varphi_n) \leq \hat{V}(\psi)$ which is equivalent to

$$\min\{\hat{V}(\varphi_1), \ldots, \hat{V}(\varphi_n)\} \leq \hat{V}(\psi).$$

The only case above that we actually care about is that $\hat{V}(\varphi_1) = \cdots = \hat{V}(\varphi_n) = 1$ implies $\hat{V}(\psi) = 1$, which is exactly what $\{\varphi_1, \ldots, \varphi_n\} \vDash \psi$. ■

1.8.3 Equivalence

We have the following definition.

Definition 1.122 (Equivalence)**.** Two formulae $\varphi, \psi \in \mathcal{L}(P)$ are *equivalent* if and only if

$$\hat{V}(\varphi) = \hat{V}(\psi)$$

for each valuation $V : P \to \{0, 1\}$.

Example 1.123. We have that $\neg\neg p$ is equivalent to p.

Remark 1.124. Saying that φ and ψ are equivalent is the same as asserting $\vDash \varphi \leftrightarrow \psi$. In fact, we can define φ being equivalent if and only if φ is equivalent to $p \vee \neg p$.

Here are some more examples.

- Idempotence: $\varphi \leftrightarrow (\varphi \wedge \varphi)$.
- Commutativity: $(\varphi \wedge \psi) \leftrightarrow (\psi \wedge \varphi)$ and $(\varphi \vee \psi) \leftrightarrow (\psi \vee \varphi)$.
- Associativity: $(\varphi \wedge \psi) \wedge \chi \leftrightarrow \varphi \wedge (\psi \wedge \chi)$ and $(\varphi \vee \psi) \vee \chi \leftrightarrow \varphi \vee (\psi \vee \chi)$.
- Absorption: $\varphi \leftrightarrow ((\varphi \wedge (\varphi \vee \psi)))$, which we can see by a direct computation.
- Distributivity: $(\varphi \wedge (\psi \vee \chi)) \leftrightarrow ((\varphi \wedge \psi) \vee (\varphi \wedge \chi))$.
- Distributivity: $(\varphi \vee (\psi \wedge \chi)) \leftrightarrow ((\varphi \vee \psi) \wedge (\varphi \vee \chi))$.
- De Morgan Laws: $\neg(\psi \wedge \varphi) \leftrightarrow (\neg\psi \vee \neg\varphi)$.
- De Morgan Laws: $\neg(\psi \vee \varphi) \leftrightarrow (\neg\psi \wedge \neg\varphi)$.

1.9 February 7

We are still talking about semantics. Today we are focusing on satisfiability.

1.9.1 Satisfiability

Satisfiability is dual to validity: if there is any way for the formula to be true, then the formula is satisfiable.

Non-Example 1.125. The formula $p \wedge \neg p$ is unsatisfiable: it is always false.

To help build some intuition, we note that we have some feeling for when multiple propositions may be true or false.

Exercise 1.126. Suppose we have the following statements of agencies responsible for an action.

1. Agency A or Agency B is responsible.
2. It's not the case that both Agencies A and B are responsible.
3. If Agency A is responsible, then the money is coming from fun X.
4. If Agency A is not responsible, then the money is coming from fun Y.
5. If Agency B is responsible, then the money is not coming from fun Y.
6. If Agency B is responsible, then the money is not coming from fun X.

We show not all of these can be true simultaneously; namely, this set of premises is "inconsistent."

Proof. We can translate as follows into our formal language. Here a (resp., b) is "Agency A (resp., B) is responsible" and x (resp., y) is "the money is coming from X (resp., Y)."

1. $a \vee b$.
2. $\neg(a \wedge b)$.
3. $a \to x$.
4. $\neg a \to y$.
5. $b \to \neg y$.
6. $\neg b \to \neg x$.

There is no valuation will make all of these true. If a were true, then x; but also a means $\neg b$, which means $\neg x$. Conversely, if a were not true, then y; but also $\neg a$ means b, which means $\neg y$. So all cases don't make sense. ■

Here are our formal definitions.

Definition 1.127 (Inconsistent)**.** A set of formulae is *inconsistent* if and only if one can prove contradiction from the formulae, using some proof system.

Definition 1.128 (Satisfiable)**.** A set of formulae $\{\varphi_1, \ldots, \varphi_n\} \subseteq \mathcal{L}(P)$ is *satisfiable* if and only if there is a valuation $V : P \to \{0, 1\}$ which satisfies all the formulae. A set of formulae is *unsatisfiable* if and only if it is not satisfiable.

We say that a single formula φ is *(un)satisfiable* if and only if $\{\varphi\}$ is (un)satisfiable. In other words, there is a valuation V such that $V \vDash \varphi$.

Example 1.129. The formula $p \wedge \neg q$ is satisfiable: $p = 1$ and $q = 0$.

Example 1.130. The formula $p \wedge \neg p$ is unsatisfiable: $p = 0$ and $p = 1$ both give $p \wedge \neg p$ false.

We could also say that $\{\varphi_1, \dots, \varphi_n\}$ is satisfiable if and only if $\varphi_1 \wedge \cdots \wedge \varphi_n$ is satisfiable.

Remark 1.131. In fact, it will be true that any unsatisfiable set of formulae will be able to derive contradiction from our proof system. So unsatisfiability will be equivalent to inconsistent, though this is not immediately obvious.

Satisfiability in fact is more directly dual to validity.

Proposition 1.132. Fix $\varphi \in \mathcal{L}(P)$. Then φ is satisfiable if and only if $\neg\varphi$ is not valid.

Proof. We have φ is satisfiable if and only if there exists some $V : P \to \{0, 1\}$ such that $V \vDash \varphi$.

Conversely, $\neg\varphi$ is not valid if and only if there exists a valuation $V : P \to \{0, 1\}$ such that $\hat{V}(\neg\varphi) = 0$ if and only if $\hat{V}(\varphi) = 1$. ■

Example 1.133. The expression $p \wedge \neg p$ is unsatisfiable, so $\neg(p \wedge \neg p)$ is valid.

Corollary 1.134. Fix $\varphi \in \mathcal{L}(P)$. Then φ is valid if and only if $\neg\varphi$ is not satisfiable.

Proof. This is the contraposition of Proposition 1.132. ■

To review, we have defined the following semantic notions.

- Valid argument forms.
- Valid formulae.
- Satisfiable sets of formulae.
- Satisfiable formulae.

In fact, we know that studying any one of these can study any other, assuming finiteness.

1.9.2 Infinite Arguments

We quickly remark that we can generalize some of our notions to the infinite case.

Definition 1.135 (Satisfiable)**.** Any set $S \subseteq \mathcal{L}(P)$ is *satisfiable* if and only if there is a valuation $V : P \to \{0, 1\}$ satisfying all of them.

However, this is no longer the same as some formula

$$\bigwedge_{\varphi \in S} \varphi$$

where we and everything together: all of our formulae have finite length. One could generalize formulae to allow infinite formulae, but we won't do so here.

We remark while we are here that we have the following notion of compactness.

Theorem 1.136 (Compactness)**.** A set S of formulae is satisfiable if and only if every finite subset of S is satisfiable.

Proof. The forwards direction is clear: if a valuation satisfies S, then the valuation will satisfy any subset.

The backwards direction is significantly harder. In essence, one shows that a set of formula is unsatisfiable if and only if one is able to derive contradiction. However, deriving contradiction is a finite process which only takes finitely many propositions, so this finite subset would also be unsatisfiable. ■

Example 1.137. Any finite graph can be 4-colored, so by compactness, any graph (possibly infinite) can be 4-colored.

Next class we will start talking about economy of language: exactly what connectives do we need to construct all possible truth functions? It will turn out that we do not need many.

Theme 2
Basic Theory

2.1 February 9

Welcome back everybody.

2.1.1 Defining Translation

As a fun exercise, for the next few lectures we will be reducing the number of connectives we really have to talk about. Recall that two formulae $\varphi, \psi \in \mathcal{L}(P)$ are equivalent if and only if each valuation $V : P \to \{0, 1\}$ has $\hat{V}(\varphi) = \hat{V}(\psi)$.

Our main goal for today is to show that any formula in our language is equivalent to a formula whose only connectives are $\neg$ and $\wedge$.

> **Remark 2.1.** One reason we might care is that this reduces the number of logic gates that we would need. Another reason we might care is that, for proofs on truth values, it reduces our headaches in checking if something is true for any connective if we simply get rid of all connectives.

The main idea in the proof is to define a translation function $T : \mathcal{L}(P) \to \mathcal{L}(P)$, where the output should only use the connectives $\neg, \wedge, \vee$ and be equivalent to the original input formula. As usual, we define T recursively, as follows; fix $\varphi, \psi \in \mathcal{L}(P)$.

- We define $T(p) \coloneqq p$ for any atomic formula $p \in P$.
- We define $T(\neg\varphi) \coloneqq \neg T(\varphi)$. Note that this does not mean we output $\neg\varphi$: we still need to translate φ, but the negation in front of φ is safe.
- We define $T((\varphi \wedge \psi)) \coloneqq (T(\varphi) \wedge T(\psi))$. Again, the point is that we don't care about the connective, but we still need to translate φ and ψ.

The next three clauses are more difficult.

- We define $T((\varphi \vee \psi)) \coloneqq \neg(\neg T(\varphi) \wedge \neg T(\psi))$. The equivalence is De Morgan's laws: saying that "we are going to the beach or hiking" is the same as "it is not the case that we are neither going to the beach nor going hiking."
- We define $T((\varphi \to \psi)) \coloneqq \neg(T(\varphi) \wedge \neg T(\psi))$. The equivalence is by noting $(\varphi \to \psi)$ simply means $(\neg\varphi \vee \psi)$, so we get this by applying De Morgan's laws.
- We define $T((\varphi \leftrightarrow \psi)) = \big(\neg(T(\varphi) \wedge \neg T(\psi)) \wedge \neg(T(\psi) \wedge \neg T(\varphi))\big)$. The idea here is that $(\varphi \leftrightarrow \psi)$ means $(\varphi \to \psi) \wedge (\psi \to \varphi)$.

Exercise 2.2. We compute $T(\neg(p \to (q \vee r)))$.

Proof. We mindlessly apply our translation rules. This gives

$$\begin{aligned} T(\neg(p \to (q \vee r))) &= \neg T((p \to (q \vee r))) \\ &= \neg\neg(T(p) \wedge \neg T(q \vee r)) \\ &= \neg\neg(T(p) \wedge \neg\neg(\neg T(q) \wedge \neg T(r))) \\ &= \boxed{\neg\neg(p \wedge \neg\neg(\neg q \wedge \neg r))}, \end{aligned}$$

which finishes. This formula might look worse (it's longer), but it only involves $\neg$ and $\wedge$. We remark that we can get rid of $\neg\neg$ to get an equivalent formula, which gives $p \wedge \neg q \wedge \neg r$. ■

2.1.2 Rigorizing Translation

Now let's actually prove that we translated correctly.

Theorem 2.3. Any formula $\varphi \in \mathcal{L}(P)$ is equivalent to a formula whose only connectives are $\neg$ and $\wedge$.

Proof. The main idea is to use our translation T defined above. We have two different main claims.

Lemma 2.4. Any formula $\varphi \in \mathcal{L}(P)$ only uses $\neg$ and $\wedge$ as its connectives.

Proof. This is an induction which we can see directly. Let $S \subseteq \mathcal{L}(P)$ be the set of formula such that $\varphi \in S$ if and only if $T(\varphi)$ only uses the connectives $\neg$ and $\wedge$. We have the following checks.

- We see that each $p \in P$ has $T(p) = p$ has no connectives at all.
- For the inductive step, suppose $\varphi, \psi \in S$. Then we have the following checks.
  - We see $T(\neg\varphi) = \neg T(\varphi)$ will only use $\neg$ and $\wedge$ because $T(\varphi)$ will only use $\neg$ and φ.
  - We see $T((\varphi \wedge \psi)) \coloneqq (T(\varphi) \wedge T(\psi))$ will only use $\neg$ and $\wedge$ because $T(\varphi)$ and $T(\psi)$ will only use $\neg$ and φ.
  - We see $T((\varphi \vee \psi)) \coloneqq \neg(\neg T(\varphi) \wedge \neg T(\psi))$ will only use $\neg$ and $\wedge$ because $T(\varphi)$ and $T(\psi)$ will only use $\neg$ and φ.
  - We see $T((\varphi \to \psi)) \coloneqq \neg(T(\varphi) \wedge \neg T(\psi))$ will only use $\neg$ and $\wedge$ because $T(\varphi)$ and $T(\psi)$ will only use $\neg$ and φ.
  - We see $T((\varphi \leftrightarrow \psi)) = \big(\neg(T(\varphi) \wedge \neg T(\psi)) \wedge \neg(T(\psi) \wedge \neg T(\varphi))\big)$ will only use $\neg$ and $\wedge$ because $T(\varphi)$ and $T(\psi)$ will only use $\neg$ and φ.

From all these checks, we see that S must contain $\mathcal{L}(P)$, so $S = \mathcal{L}(P)$, so we are done. ■

Lemma 2.5. Any formula $\varphi \in \mathcal{L}(P)$ is equivalent to $T(\varphi)$.

Proof. We proceed by induction. For our base case, we are saying that $p \in P$ is equivalent to $T(p) = p$, which is clear.

For the inductive step, fix φ and ψ which are equivalent to their translation. We have the following checks.

- We show $(\varphi \vee \psi)$ is equivalent to $T((\varphi \vee \psi))$. Well, fix any valuation V, and we see that

$$\begin{aligned}\hat{V}((\varphi \vee \psi)) &= \max\{\hat{V}(\varphi), \hat{V}(\psi)\} \\ &\stackrel{*}{=} 1 - \min\{1 - \hat{V}(\varphi), 1 - \hat{V}(\psi)\}.\end{aligned}$$

  The point is that we can check $\stackrel{*}{=}$ is true by a computation. Here is the table.

x	y	$\max\{x, y\}$	$1-x$	$1-y$	$1-\min\{1-x, 1-y\}$
1	1	1	0	0	1
1	0	1	0	1	1
0	1	1	1	0	1
0	0	0	1	1	0

  We continue. We see, because φ is equivalent to $\hat{V}(\varphi)$ and similar for ψ, we see

$$\begin{aligned}\hat{V}((\varphi \vee \psi)) &= 1 - \min\{1 - \hat{V}(\varphi), 1 - \hat{V}(\psi)\} \\ &= 1 - \min\{1 - \hat{V}(T(\varphi)), 1 - \hat{V}(T(\psi))\} \\ &= 1 - \min\{\hat{V}(\neg T(\varphi)), \hat{V}(\neg T(\psi))\} \\ &= 1 - \hat{V}((\neg T(\varphi) \wedge \neg T(\psi))) \\ &= \hat{V}(\neg(\neg T(\varphi) \wedge \neg T(\psi))).\end{aligned}$$

  Thus, $(\varphi \vee \psi)$ is equivalent to $\neg(\neg T(\varphi) \wedge \neg T(\psi)) = T((\varphi \vee \psi))$, so we are done.

- The cases of $\to$ and $\leftrightarrow$ are similar, so we will omit them.

- We show $(\varphi \wedge \psi)$ is equivalent to $T((\varphi \wedge \psi)) = (T(\varphi) \wedge T(\psi))$. But we can compute, for any valuation V,

$$\begin{aligned}\hat{V}((\varphi \wedge \psi)) &= \min\{\hat{V}(\varphi), \hat{V}(\psi)\} \\ &= \min\{\hat{V}(T(\varphi)), \hat{V}(T(\psi))\} \\ &= \hat{V}((T\varphi) \wedge T(\psi)),\end{aligned}$$

  which finishes.

- We show $\neg\varphi$ is equivalent to $T(\neg\varphi) = \neg T(\varphi)$. But we can compute, for any valuation V,

$$\begin{aligned}\hat{V}(\neg\varphi) &= 1 - \hat{V}(\varphi) \\ &= 1 - \hat{V}(T(\varphi)) \\ &= \hat{V}(\neg T(\varphi)),\end{aligned}$$

  which finishes.

The above checks complete our induction. ■

The above two claims show that a formula $\varphi \in \mathcal{L}(P)$ is equivalent to some formula $T(\varphi)$, where $T(\varphi)$ only uses the connectives $\neg$ and $\wedge$. ■

> **Remark 2.6.** We might want to optimize the check for, say, $(\varphi \vee \psi)$ being equivalent to $T((\varphi \vee \psi))$ by avoiding the table. However, this cannot really be done because we must deal with what $\neg(\neg T(\varphi) \wedge \neg T(\psi))$ means sometime in the proof, which is where the $\stackrel{*}{=}$ equality comes from.

Remark 2.7. The above checks were pretty annoying. But for any proof about truth in the future will now only have to check $\neg$ and $\wedge$ in an inductive step.

Explicitly, if we are trying to prove some property about formulae which is preserved by equivalence (namely, one that really only cares about formulae as truth functions), then our inductive step only has to deal with formulae which use $\wedge$ and $\neg$.

Non-Example 2.8. The property that a formula only contains one $\neg$ connector is not preserved by equivalence. For example, p is equivalent to $\neg\neg p$.

2.1.3 More Economy

There are other ways we can be economical about our connectives. For example, we could only use $\neg$ and $\vee$ using a translation function $S : \mathcal{L}(P) \to \mathcal{L}(P)$ as follows.

- We define $S(p) \coloneqq p$ for any $p \in P$.
- We define $S(\neg\varphi) \coloneqq \neg S(\varphi)$.
- We define $S((\varphi \wedge \psi)) \coloneqq \neg(\neg S(\varphi) \vee \neg S(\psi))$.
- We define $S((\varphi \wedge \psi)) \coloneqq \neg(\neg S(\varphi) \vee \neg S(\psi))$.
- We define $S((\varphi \to \psi)) \coloneqq (\neg S(\varphi) \vee S(\psi))$.
- We define $S((\varphi \leftrightarrow \psi)) \coloneqq (\neg(\neg S(\varphi) \vee S(\psi)) \vee \neg(\neg S(\psi) \vee S(\varphi)))$. Again, this comes from noting $(\varphi \leftrightarrow \psi)$ is the same as $((\varphi \to \psi) \wedge (\psi \to \varphi))$.

We will omit the checks that this is a valid translation.

2.2 February 11

Here we go.

2.2.1 A Little More Economy

Let's continue with our translation. We will again note that every formula in a language is equivalent to one in which the only connectives are $\{\neg, \to\}$. Here is our translation.

- We define $U(p) \coloneqq p$ for each atomic formula $p \in P$.
- We define $U(\neg\varphi) \coloneqq \neg U(\varphi)$.
- We define $U(\varphi \wedge \psi) \coloneqq \neg(U(\varphi) \to \neg U(\psi))$.
- We define $U(\varphi \vee \psi) \coloneqq$
- We define $U(\varphi \to \psi) \coloneqq (U(\varphi) \to U(\psi))$.
- We define $U(\varphi \leftrightarrow \psi) \coloneqq (U(\varphi) \to U(\psi)) \wedge (U(\psi) \to U(\varphi))$.

2.2.2 Too Much Economy

We quickly remark that we cannot be too callous. For example, we claim that the connectives $\{\wedge, \vee, \to, \leftrightarrow\}$ are not enough. Here is the main claim.

Lemma 2.9. Any formula $\varphi \in L(\{p_1, \dots, p_n\})$ where φ has only the connectives $\{\wedge, \vee, \to, \leftrightarrow\}$ will have $p_1, \dots, p_n \vDash \varphi$.

This will be enough because it means any such formula φ cannot be equivalent to $\neg p_1$ because $\{p_1, \dots, p_n\} \nvDash \varphi$.

Proof. We induct. For our base case, we note that any of the propositions p_i will have $\{p_1, \dots, p_n\} \vDash p_i$ for free. Now for our inductive hypothesis, suppose $\{p_1, \dots, p_n\} \vDash \varphi, \psi$. Then for any $\# \in \{\wedge, \vee, \to, \leftrightarrow\}$ will have

$$\hat{V}(\varphi) = \hat{V}(\psi) = 1 \implies \hat{V}((\varphi \# \psi)) = 1$$

by hand. So $\hat{V}(p_i) = 1$ for each p_i implies $\hat{V}(\varphi) = \hat{V}(\psi) = 1$ (by hypothesis), so $\hat{V}((\varphi \# \psi)) = 1$, finishing. ∎

Remark 2.10. Intuitively, when we set everything to be true, all the given connectives will send true to true.

We close with a question.

Question 2.11. Is any formula equivalent to one involving the symbols $\{\neg, \leftrightarrow\}$?

2.2.3 The Most Economy

We start by defining a new connective.

Definition 2.12 (Sheffer stroke, NAND)**.** Given two formulae φ, ψ, we define $(\varphi \uparrow \psi)$ by the following truth table.

$\hat{V}(\varphi)$	$\hat{V}(\psi)$	$\hat{V}((\varphi \uparrow \psi))$
1	1	0
1	0	1
0	1	1
0	0	1

Intuitively, $\hat{V}((\varphi \uparrow \psi))$ is true if and only if at least one of φ or ψ is false. In symbols, $\hat{V}((\varphi \uparrow \psi)) = 1 - \hat{V}(\varphi)\hat{V}(\psi)$.

We note that $\neg\varphi$ is equivalent to $(\varphi \uparrow \varphi)$ because

$$\hat{V}((\varphi \uparrow \varphi)) = 1 - \hat{V}(\varphi)^2 = 1 - \hat{V}(\varphi) = \hat{V}(\neg\varphi),$$

where $\hat{V}(\varphi)^2 = \hat{V}(\varphi)$.

Further, $(\varphi \vee \psi)$ is equivalent to $\neg(\neg\varphi \wedge \neg\psi)$ is equivalent to $(\neg\varphi \uparrow \neg\psi)$ is equivalent to

$$((\varphi \uparrow \varphi) \uparrow (\psi \uparrow \psi)).$$

Thus, we note that any formula is equivalent to one involving only the connectives $\vee$ or $\neg$, which we can then translate to a formula only involving the connectives.

Of course, there are other translations.

Example 2.13. We note $(\varphi \wedge \psi)$ is equivalent to $\neg\neg(\varphi \wedge \psi)$ is equivalent to $\neg(\varphi \uparrow \psi)$ is equivalent to $(\varphi \uparrow \psi) \uparrow (\varphi \uparrow \psi)$.

Remark 2.14. Any formula is also equivalent to one which only uses the connective "$\downarrow$," which has the following truth table.

$\hat{V}(\varphi)$	$\hat{V}(\psi)$	$\hat{V}((\varphi \downarrow \psi))$
1	1	0
1	0	0
0	1	0
0	0	1

To see the above remark, we note that the top row must be false (or else we preserve truth), and the bottom row must be true (or else we preserve false). But the truth table

$\hat{V}(\varphi)$	$\hat{V}(\psi)$	$\hat{V}((\varphi \# \psi))$
1	1	0
1	0	1
0	1	0
0	0	1

does not work because it does not depend on $\hat{V}(\varphi)$. Similarly,

$\hat{V}(\varphi)$	$\hat{V}(\psi)$	$\hat{V}((\varphi \# \psi))$
1	1	0
1	0	0
0	1	1
0	0	1

does not work because it does not depend on $\hat{V}(\psi)$.

2.2.4 Truth Functions

We should probably go back and show that any function on truth can be achieved by our language.

Definition 2.15. Fix a positive integer n. Then an *n-ary truth function* is a function $f : \{0,1\}^n \to \{0,1\}$.

Example 2.16. The function $f(x) = 1 - x$ corresponds to $\neg$. The function $f(x,y) = \min\{x,y\}$ corresponds to $\wedge$.

Example 2.17. The function

$$f(x,y,z,w) = \min\{x,y,z,1-w,xy,\max\{x,y\}\}$$

is a truth function.

We note that, because there are only finitely many inputs in $\{0,1\}^n$, we can simply tabulate to give the function. For example, here is a truth function.

p	q	r	$f(p,q,r)$
1	1	1	1
1	1	0	1
1	0	1	0
1	0	0	0
0	1	1	1
0	1	0	0
0	0	1	1
0	0	0	0

This might look like a mess, but it is perfectly reasonable.

Remark 2.18. Note that, among the 2^n inputs for an n-ary truth function, we have two choices for inputs, so there are 2^{2^n} total possible truth functions. That's a lot.

So far we have been noting that formulae in our language correspond to truth functions.

Example 2.19. The formula $\neg(p \leftrightarrow q)$ corresponds to the truth function described by the following table.

p	q	$f(p,q)$
1	1	0
1	0	1
0	1	1
0	0	0

This is not unique: we could also do $(p \leftrightarrow \neg q)$.

Let's make this rigorous.

Definition 2.20 (Truth functions from formulae)**.** Fix $\varphi \in \mathcal{L}(\{p_1, \ldots, p_n\})$. We then say φ *defines the n-ary truth function* f^n_φ defined as follows: any $(x_1, \ldots, x_n) \in \{0,1\}^n$ defines a valuation $V(p_k) \coloneqq x_k$, from which we define

$$f^n_\varphi(x_1, \ldots, x_n) \coloneqq \hat{V}(\varphi).$$

Intuitively, the truth tables for f^n_φ and $\hat{V}(\varphi)$ "match up" in the way that the row for some $(x_1, \ldots, x_n) \in \{0,1\}^n$ corresponds to the row $V(p_1) = x_1, \ldots, V(p_n) = 1$.

Let's see some examples.

Example 2.21. The following truth tables match up.

p	$\neg p$
1	0
0	1

x	$f(x)$
1	0
0	1

So $\neg p$ defines the function f.

Example 2.22. We find a formula to define the following truth function.

p	q	r	$f(p,q,r)$
1	1	1	1
1	1	0	1
1	0	1	0
1	0	0	0
0	1	1	1
0	1	0	0
0	0	1	1
0	0	0	0

We claim that $((p \to q) \wedge (\neg p \to r))$ will do the trick. Checking this is a matter of writing out the truth table, but we won't do the work here.

2.2.5 Defining Truth Functions

There is actually an algorithm which will always be able to produce a truth table. Let's work with the following example.

p	q	r	$f(p,q,r)$
1	1	1	1
1	1	0	1
1	0	1	0
1	0	0	0
0	1	1	1
0	1	0	0
0	0	1	1
0	0	0	0

(In the following discussion, we will drop parentheses when the meaning is clear or the parentheses do not matter.)

Above we have highlighted all the cases which are true. To create a formula for this, we simply hard-code in all possible cases where are true. To be explicit, the top highlighted row is $(p \wedge q \wedge r)$, the next row is $(p \wedge q \wedge \neg r)$, and so on. Then we just need to ensure that we live in exactly one of these cases, which looks like

$$(p \wedge q \wedge r) \vee (p \wedge q \wedge \neg r) \vee (\neg p \wedge q \wedge r) \vee (\neg p \wedge \neg q \wedge r).$$

This is long and horrendous, but it is correct.

Remark 2.23. To feel how bad this algorithm is, consider how annoying it would be to try to create a truth table for "$f(x_1, \dots, x_n)$ returns true if and only if at least $\lfloor n/2 \rfloor$ of the x_k are true."

Example 2.24. The truth table

p	q	$f(p,q)$
1	1	0
1	0	0
0	1	0
0	0	1

can be defined by the formula $(\neg p \wedge \neg q)$.

We quickly remark that there will be lots of different formulae which can give the same truth function.

Lemma 2.25. Two formula $\varphi, \psi \in \mathcal{L}(\{p_1, \dots, p_n\})$ have $f_\varphi^n = f_\psi^n$ if and only if φ and ψ are logically equivalent.

Proof. To say that φ and ψ means that any valuation $V : \{0,1\}^n \to \{0,1\}$ gives $\hat{V}(\varphi) = \hat{V}(\psi)$. Translating this into discussion about truth functions, a valuation corresponds to an input $(x_1, \dots, x_n) \in \{0,1\}^n$, so being logically equivalent is saying that f_φ^n and f_ψ^n are equal on all inputs $(x_1, \dots, x_n) \in \{0,1\}^n$. ∎

2.3 February 14

We're back in action everybody.

2.3.1 Equivalence Classes

Recall that it is possible for two formulae to define the same truth function. For example, consider the truth table as follows.

x_1	x_2	$\max\{x_1, x_2\}$
1	1	1
1	0	1
0	1	1
0	0	0

This function is equal to either $f^2_{p\vee q}$ or $f^2_{\neg(\neg p\wedge\neg q)}$. Indeed, we can compute as follows.

p	q	$p\vee q$	$\neg p$	$\neg q$	$\neg(\neg p\neg q)$
1	1	1	0	0	1
1	0	1	0	1	1
0	1	1	1	0	1
0	0	0	1	1	0

We also recall that we had the following lemma from last class.

Lemma 2.25. Two formula $\varphi, \psi \in \mathcal{L}(\{p_1, \dots, p_n\})$ have $f^n_\varphi = f^n_\psi$ if and only if φ and ψ are logically equivalent.

The above idea gives us a notion of an equivalence class of formulae.

Definition 2.26 (Eqiuvalence class)**.** Fix a formula $\varphi \in \mathcal{L}(P)$. Then we define the *equivalence class* of φ in $\mathcal{L}(P)$ to be the set of formulae which are logically equivalent to φ. We denote this set by $[\varphi]$, for which φ is a representative.

Example 2.27. We have $\neg(\neg p \wedge \neg q) \in [p \vee q]$ by de Morgan's laws. Similarly, $\neg\neg(p \vee q) \in [p \vee q]$.

The point of the equivalence class is that it ignores the underlying syntax of a formula and only cares about its semantics. If the only thing that we care about is what a formula "means" instead of what it looks like, then it makes sense to lump together formulae which are equivalent.

Our notion of equivalence class makes sense, as follows.

Proposition 2.28. Logical equivalence is reflexive, symmetric, and transitive. That is, for any $\alpha, \beta, \gamma \in \mathcal{L}(P)$, we have the following.

- Reflexive: $\alpha \equiv \alpha$.
- Symmetric: $\alpha \equiv \beta$ implies $\beta \equiv \alpha$.
- Transitive: $\alpha \equiv \beta$ and $\beta \equiv \gamma$ implies $\alpha \equiv \gamma$.

Proof. We show the claims as follows. Fix $\alpha, \beta, \gamma \in \mathcal{L}(P)$. Denote equivalence by $\equiv$.

- Reflexive: for any valuation $V : P \to \{0, 1\}$, we have $\hat{V}(\alpha) = \hat{V}(\alpha)$, so $\alpha \equiv \alpha$.
- Symmetric: if $\alpha \equiv \beta$, then any valuation $V : P \to \{0, 1\}$ has $\hat{V}(\alpha) = \hat{V}(\beta)$ so that $\hat{V}(\beta) = \hat{V}(\alpha)$, so $\beta \equiv \alpha$.
- Transitive: if $\alpha \equiv \beta$ and $\beta \equiv \gamma$, then any valuation $V : P \to \{0, 1\}$ has $\hat{V}(\alpha) = \hat{V}(\beta)$ and $\hat{V}(\beta) = \hat{V}(\gamma)$ so that $\hat{V}(\alpha) = \hat{V}(\gamma)$, so $\alpha \equiv \gamma$. ∎

Remark 2.29. The above properties show that logical equivalence is an equivalence relation.

Corollary 2.30. Two formulae $\varphi, \psi \in \mathcal{L}(P)$ have φ equivalent to ψ if and only if $[\varphi] = [\psi]$. So this is also equivalent to $f^n_\varphi = f^n_\psi$.

Proof. We show the directions independently.

- Suppose $\varphi \equiv \psi$. Then, for any $\alpha \in [\varphi]$, we have $\varphi \equiv \alpha$, so $\alpha \equiv \varphi$ and $\varphi \equiv \psi$, so $\alpha \equiv \psi$, so $\alpha \in [\psi]$. By symmetry, any $\alpha \in [\psi]$, we have $\psi \equiv \alpha$ and $\psi \equiv \varphi$, so $\alpha \equiv \varphi$, so $\alpha \in [\varphi]$.
- Suppose $[\varphi] = [\psi]$. Then $\varphi \equiv \varphi$ implies $\varphi \in [\varphi] = [\psi]$, so $\varphi \equiv \psi$. ■

2.3.2 Local Finiteness

We might be interested in counting the total number of equivalence classes, but this might be infinite if the number of propositions is infinite. Namely, each proposition $p \in P$ gives a different equivalence class $[p]$, lower-bounding the total number of formulae.

However, there is a notion of local finiteness.

Proposition 2.31 (Local finiteness)**.** Fix $P = \{p_1, \dots, p_n\}$ a finite set. Then there are only finitely many equivalence classes in $\mathcal{L}(P)$. In other words, there are only finitely many non-equivalent formulae we can write with the propositions $\{p_1, \dots, p_n\}$.

Example 2.32. There are only four different possible truth functions in $\mathcal{L}(\{p\})$, as follows.

p	$p \vee \neg p$
1	1
0	1

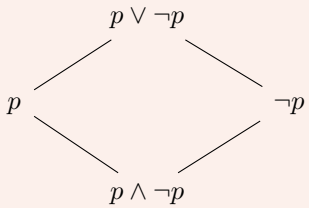
p	p
1	1
0	0

p	$\neg p$
1	0
0	1

p	$p \wedge \neg p$
1	0
0	0

Explicitly, each input $\{0, 1\}$ for $V(p)$ has only two options, giving 2^2 total possible truth tables, which we have manifest above.

Remark 2.33. We can tabulate equivalence classes in $\mathcal{L}(\{p\})$ in a Hasse diagram (moving up means logically implies) as follows.



For example, $p \wedge \neg p \vDash p \vDash p \vee \neg p$. As an aside, this is a boolean algebra.

It is a good exercise to make the Hasse diagram for $\mathcal{L}(\{p, q\})$.

Now let's prove our result.

Proof of Proposition 2.31. We can inject equivalence classes $[\varphi]$ in $\mathcal{L}(P)$ to their associated truth function f^n_φ. Namely, the function

$$[\varphi] \mapsto f^n_\varphi$$

is well-defined and one-to-one/injective. We have the following checks.

- Well-defined: if $[\varphi] = [\psi]$, then $\varphi \equiv \psi$, so $f^n_\varphi = f^n_\psi$.
- Injective: $f^n_\varphi = f^n_\psi$ implies $\varphi \equiv \psi$ implies $[\varphi] = [\psi]$.

This embedding implies that the number of equivalence classes is bounded above by the number of truth functions because each equivalence class yields a unique truth function.

Remark 2.34. The last step we just did is the Pigeonhole principle: if we have an injection $A \hookrightarrow B$, then the number of elements of A is at most the number of elements of B. This should feel intuitively obvious to prove, and in fact it makes a pretty good way to define the size of a set to begin with, so there is nothing to prove.

So to finish, we note that there are 2^{2^n} truth functions on n variables: there are only 2^n different ways to set the inputs to set true and false for each $p_\bullet$ because each $p_\bullet$ has two options. But then each of these inputs has 2 options to return true or false, so we have a total of 2^{2^n} different functions. ■

Corollary 2.35. Let $P = \{p_1, \dots, p_n\}$ be a finite set. There are at most 2^{2^n} different equivalence classes of formulae.

Proof. This essentially follows from the proof of Proposition 2.31, where we upper-bounded the number of equivalence classes by the number of truth functions, of which there are at most 2^{2^n}. ■

Remark 2.36. It will turn out that all truth functions are achievable from formulae, so the map $[\varphi] \mapsto f^n_\varphi$ will also be onto/surjective, so the number of logical equivalence classes will be exactly the number of truth functions, which is 2^{2^n}.

2.3.3 Completeness

Let's manifest Remark 2.36. Given a truth function, we need to find its formula.

Example 2.37. The truth table

x_1	x_2	$f(x_1, x_2)$
1	1	0
1	0	1
0	1	1
0	0	0

is $f^2_{\neg p \leftrightarrow q}$.

So here is the main claim.

Theorem 2.38. Any truth function (with finitely many inputs) can be written by a formula in our language.

Proof. Fix $f : \{0,1\}^n \to \{0,1\}$ some truth function with n inputs, and let $P = \{p_1, \dots, p_n\}$ be our propositions. We remark that if f is the zero function, then $f = f^n_{p_1 \wedge \neg p_1}$ will work.

Otherwise, f returns true somewhere. Let S be the set of all inputs which return true, and because f takes finitely many inputs, the input space of f is finite, so S is finite, so set $S = \{s_1, \dots, s_n\}$. Now, for each $s := (x_1, \dots, x_n) \in S$, we create the term

$$\varphi_s := \pm p_1 \wedge \cdots \wedge \pm p_n,$$

where we choose $+p_\bullet = p_\bullet$ when $x_\bullet = 1$ and $-p_\bullet = \neg p_\bullet$ when $x_\bullet = 0$. Then the formula

$$\varphi = \varphi_{s_1} \vee \cdots \varphi_{s_n}$$

will do the trick. Namely, it is true if and only if our input s lives in S because φ_s will be true for the input x if and only if $s = x$. ■

Remark 2.39. Because we showed that any formula is equivalent to one that only uses the connectives $\{\neg, \wedge\}$ or even $\{\uparrow\}$, any truth function can be written by a formula in $\{\neg, \wedge\}$ or even $\{\uparrow\}$.

2.4 February 16

Here we go.

2.4.1 Finishing Completeness

Today we are continuing the proof of Theorem 2.38.

Theorem 2.38. Any truth function (with finitely many inputs) can be written by a formula in our language.

Proof. Here is the idea: suppose that $f : \{0,1\}^n \to \{0,1\}$ is a truth function. Because it has finitely many inputs, the pre-image of $\{1\}$ is finite, so we enumerate them by

$$r_1, \ldots, r_m,$$

where the "row" r_k is the input $(b_{k1}, \ldots, b_{km})$. With this, we define

$$\varphi_{r_k} \coloneqq \pm p_1 \wedge \cdots \wedge p_m,$$

where we take $+p_\bullet = p_\bullet$ if and only if $b_{k\bullet} = 1$ and $-p_\bullet = \neg p_\bullet$ otherwise.

Definition 2.40 (Literal)**.** A formula of the form $p, \neg p \in \mathcal{L}(P)$ is called a *literal*.

The point is that φ_{r_k} is true if and only if the input is equal to $(b_1, \ldots, b_m)$. In particular, when we take

$$\varphi \coloneqq \varphi_{r_1} \vee \cdots \vee \varphi_{r_m},$$

we see that φ is true if and only if one of the $\varphi_{r_\bullet}$ is true if and only if one of the inputs is equal to $r_\bullet$. So φ defines the truth function f. ■

2.4.2 Normal Forms

We have the following corollary of Theorem 2.38.

Corollary 2.41 (Disjunctive normal form)**.** Fix P a finite set. Every truth function $P \to \{0,1\}$ can be represented by using $\neg, \vee, \wedge$ as a disjunction of conjunction of literals.

Recall that disjunction means $\vee$ (having one but not the other is legal), and conjunction means $\wedge$ (we must have both).

Proof. This is exactly what Theorem 2.38 does: it produces conjunctions of literals and then disjuncts them together. ■

Definition 2.42 (Disjunctive normal form)**.** A formula φ which is a disjunction of a conjunction of literals is said to be in *disjunctive normal form*.

Example 2.43. Disjunctive normal form is not unique. For example,

$$(p \wedge q \wedge r) \vee (p \wedge q \wedge \neg r) \vee (\neg p \wedge q \wedge r) \vee (\neg p \wedge \neg q \wedge r)$$

is equivalent to

$$(\neg p \wedge q \wedge r) \vee (\neg p \wedge \neg q \wedge r),$$

both of which are in disjunctive normal form.

There is also a dual notion to disjunctive normal form.

Definition 2.44 (Conjunctive normal form)**.** A formula φ which is a conjunction of a disjunction of literals is said to be in *disjunctive normal form*.

To put a formula into conjunctive normal form, we can put $\neg\varphi$ into disjunctive normal form as

$$\neg\varphi \equiv \bigvee_{k=1}^{m} (\ell_{k,1} \wedge \cdots \wedge \ell_{k,n}).$$

In other words, we look for all the 0s in the truth table. Then we negate to take

$$\varphi \equiv \bigwedge_{k=1}^{m} (\ell'_{k,1} \vee \cdots \vee \ell'_{k,n}),$$

which is our conjunctive normal form, where $\ell'_{k,\bullet}$ is the negated literal.

Example 2.45. Consider the truth table as follows.

p	q	r	f
1	1	1	1
1	1	0	1
1	0	1	0
1	0	0	0
0	1	1	1
0	1	0	0
0	0	1	1
0	0	0	0

Our zeroes are at

$$\neg\varphi \equiv (p \wedge \neg q \wedge r) \vee (p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge \neg r).$$

Negating and applying de Morgan's laws, our formula is

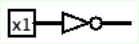
$$\varphi \equiv (\neg p \vee q \vee \neg r) \wedge (\neg p \vee q \vee r) \wedge (p \vee \neg q \vee r) \wedge (p \vee q \vee r).$$

We close by noting our proof of Theorem 2.38 finishes the exactness in the proof of Corollary 2.35. To be more explicit, we showed that there are at most 2^{2^n} different equivalent formulae because this is the number of truth functions, but Theorem 2.38 showed that all truth functions are possible. So there are indeed 2^{2^n} total different equivalence classes of formulae.

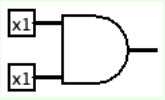
2.4.3 Our Gates

Let's start talking about some digital circuits. We start with some gates.

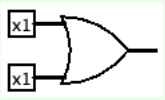
Definition 2.46 (NOT gate)**.** The *NOT gate* (or inverter) is a gate which takes low voltage to high voltage and vice versa.



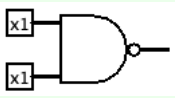
Definition 2.47 (AND gate)**.** The *AND gate* is a gate which takes high voltage if and only if both of its inputs are high voltage.



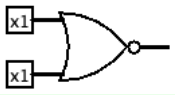
Definition 2.48 (OR gate)**.** The *OR gate* is a gate which takes high voltage if and only if at least one of its inputs are high voltage.



Definition 2.49 (NAND gate)**.** The *NAND gate* is a gate which takes high voltage if and only if at least one of its inputs are low voltage.



Definition 2.50 (NOR gate)**.** The *NOR gate* is a gate which behaves like negative of the OR gate.

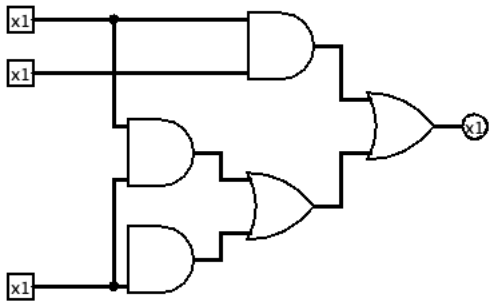


Definition 2.51 (XOR gate)**.** The *XOR gate* is a gate which has high voltage if and only if exactly one of the inputs is high-voltage.

Definition 2.52 (XNOR gate)**.** The *XNOR gate* is a gate which has high voltage if and only if its inputs have the same voltage.

2.4.4 Building Circuits

We can connect these circuits together in fun ways.



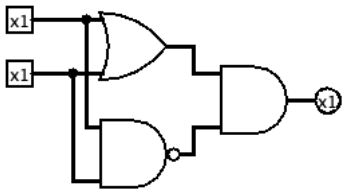
Reading off of this circuit corresponds gives the formula

$$(p \wedge q) \vee ((p \wedge r) \vee (q \wedge r)).$$

Let's see an example problem.

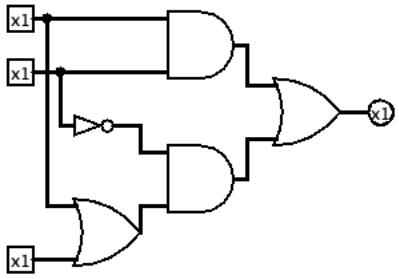
Exercise 2.53. We build a circuit with two switches such that either light switch turns the lightbulb on or off.

Proof. It is not too implausible to think that we can do this with a formula because we are essentially asking for the number of "on" light-switches to be odd to give the light-bulb on. Here is a circuit which works.



It corresponds to the formula $(p \vee q) \wedge \neg(p \wedge q)$. This is essentially an XOR gate. ■

It is possible for circuits to be inefficient. For example, consider the following circuit.



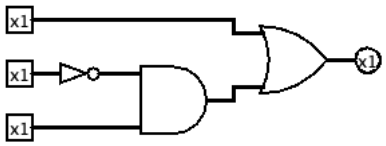
This can be simplified down. Indeed, this corresponds to the formula

$$(p \wedge q) \vee (\neg q \wedge (p \vee r)).$$

We can argue that this is equivalent to

$$p \vee (\neg q \wedge r)$$

by doing some simplification. Thus, we can build the following simpler circuit.



So this is an application of simplifying formulae.

Remark 2.54. We remark that any formula can be written using $\uparrow$, which corresponds to the fact that all circuit functions could be written in only NAND gates.

Remark 2.55. One can build the same theory of equivalence and so on of circuits by porting over the theory from formulae/truth functions to circuits.

2.5 February 18

Here we go.

2.5.1 Algorithms

Today we start talking about algorithms.

Definition 2.56 (Algorithm)**.** An *algorithm* is a procedure or set of rules or instructions that will tell us how to preform successive steps, which will complete a task after a finite number of steps.

This definition can be made more rigorous, but we will not bother for this class.

We have already seen some algorithms so far.

Example 2.57. To determine if a formula is satisfiable, we employ the following algorithm.

1. Fix a formula $\varphi \in \mathcal{L}(\{p_1, \dots, p_n\})$.
2. Fix some order of the valuations.
3. Choose the first valuation V.
4. Compute $\hat{V}(\varphi)$ by a recursive procedure.
5. If $\hat{V}(\varphi) = 1$, then φ is satisfiable.
6. Otherwise, choose the next valuation V and go back to step 4.
7. If we finish the truth table and never return satisfiable, then φ is unsatisfiable.

Remark 2.58. The above algorithm requires something like 2^n total valuations to check, making the worst-case (if unsatisfiable) requiring 2^n time. This makes this run in "exponential" time.

Remark 2.59. There are other algorithms for satisfiability. The textbook uses the semantic tableaux. We will shortly talk about resolution.

2.5.2 Another CNF Algorithm

For resolution, we will need to start in conjunctive normal form. Recall the definition, as follows.

Definition 2.60 (Conjunctive normal form)**.** A formula φ is in *conjunctive normal form* if and only if it is a conjunction of disjunctions.

We do already have an algorithm for this: find the disjunctive normal form $\neg\varphi$ and then negate and apply De Morgan's laws.

However, the disjunctive normal form requires us to in advance have a truth table, which needs a notion of semantics. Here is an algorithm which is more syntactic.

Proposition 2.61. Fix a formula φ. We put φ into conjunctive normal form.

1. Remove $\leftrightarrow$s: for each $\leftrightarrow$ starting from the left, replace each subformula of the form $(\alpha \leftrightarrow \beta)$ with $((\alpha \to \beta) \wedge (\beta \to \alpha))$.
2. Remove $\to$s: for each $\to$ starting from the left, replace each subformula $(\alpha \to \beta)$ with $(\neg\alpha \vee \beta)$.

At this point, the only legal connectives are $\wedge$, $\vee$, and $\neg$. In particular, if $\neg\psi$ is a subformula, then ψ is either of the form $(\alpha \wedge \beta)$ or $(\alpha \vee \beta)$ or $\neg\alpha$.

3. Drive $\neg$ inside: for each subformula of the form $\neg(\alpha \vee \beta)$ starting from the left, replace it with $(\neg\alpha \wedge \neg\beta)$.
4. Drive $\neg$ inside: for each subformula of the form $\neg(\alpha \wedge \beta)$ starting from the left, replace it with $(\neg\alpha \vee \neg\beta)$.
5. Drive $\neg$ inside: for each subformula of the form $\neg\neg\alpha$ starting from the left, replace it with α.

So now $\neg$ only applies to propositions, so we are only applying $\wedge$ and $\vee$ to literals. We need to normalize to get conjunctives of disjunctions; the only way we violate being in conjunctive of normal form is if we have a subformula which is a $\vee$ of non-literals, and one of these non-literals had better have $\wedge$ lest we just break apart into separate disjunctive terms.

6. Distribute: for each subformula of the form $(\alpha \vee (\beta \wedge \gamma))$ starting from the left, replace it with $((\alpha \vee \beta) \wedge (\alpha \vee \gamma))$.
7. Distribute: for each subformula of the form $((\alpha \wedge \beta) \vee \gamma)$ starting from the left, replace it with $(((\alpha\gamma) \wedge (\beta \vee \gamma))$.
8. Go back to step 6 until done.

We will not prove that this works, but there are remarks throughout the description. More formally, we are defining a recursive function $\mathrm{CNF} : \mathcal{L}(P) \to \mathcal{L}(P)$ with many cases. For completeness, here is the definition.

1. $\mathrm{CNF}(p) = p$ for each $p \in P$.
2. $\mathrm{CNF}(\varphi \leftrightarrow \psi) = (\neg\,\mathrm{CNF}(\alpha) \vee \mathrm{CNF}(\beta)) \wedge (\neg\,\mathrm{CNF}(\beta) \vee \mathrm{CNF}(\alpha))$.
3. $\mathrm{CNF}(\varphi \to \psi) = \neg\,\mathrm{CNF}(\alpha) \vee \mathrm{CNF}(\beta)$.
4. $\mathrm{CNF}(\varphi \wedge \psi) = \mathrm{CNF}(\varphi) \wedge \mathrm{CNF}(\psi)$.
5. $\mathrm{CNF}(\neg(\alpha \wedge \neg\beta)) = \neg\,\mathrm{CNF}(\alpha) \vee \neg\,\mathrm{CNF}(\beta)$.
6. $\mathrm{CNF}(\neg(\alpha \vee \neg\beta)) = \neg\,\mathrm{CNF}(\alpha) \wedge \neg\,\mathrm{CNF}(\beta)$.
7. $\mathrm{CNF}(\neg\neg\varphi) = \mathrm{CNF}(\varphi)$.
8. $\mathrm{CNF}(\alpha \vee (\beta \wedge \gamma)) = \mathrm{CNF}(\alpha \vee \beta) \wedge \mathrm{CNF}(\alpha \vee \gamma)$.

9. $\mathrm{CNF}((\alpha \wedge \beta) \vee \gamma) = \mathrm{CNF}(\alpha \vee \gamma) \wedge \mathrm{CNF}(\beta \vee \gamma)$.

It is not obvious that the above conditions fully determine CNF recursively (in fact, the last two steps make CNF not well-defined because we might have a subformula of the form $(\alpha \wedge \beta) \vee (\gamma \wedge \delta)$), but there is a small justification present in the definition of the algorithm, and it is not too hard to imagine how to fix this. We will not fix this formally out of laziness.

Remark 2.62. The above algorithm need not give the same conjunctive normal form as with the truth table approach. The above algorithm is purely syntactic.

Let's see some examples.

Exercise 2.63. We put $\varphi := \neg(p_1 \wedge \neg(\neg p_2 \wedge (p_3 \to p_4)))$ in conjunctive normal form.

Proof. We have the following computation. To start, we get rid of $\to$.

$$\varphi \equiv \neg(p_1 \wedge \neg(\neg p_2 \wedge (p_3 \to p_4)))$$

Next we push $\neg$s inside and cancel out double negations.

$$\begin{aligned}
\varphi &\equiv \neg(p_1 \wedge \neg(\neg p_2 \wedge (\neg p_3 \vee p_4))) \\
&\equiv \neg p_1 \vee \neg\neg(\neg p_2 \wedge (\neg p_3 \vee p_4)) \\
&\equiv \neg p_1 \vee \neg(\neg\neg p_2 \vee \neg(\neg p_3 \vee p_4)) \\
&\equiv \neg p_1 \vee (\neg\neg\neg p_2 \wedge \neg\neg(\neg p_3 \vee p_4)) \\
&\equiv \neg p_1 \vee (\neg\neg\neg p_2 \wedge \neg(\neg\neg p_3 \wedge \neg p_4)) \\
&\equiv \neg p_1 \vee (\neg\neg\neg p_2 \wedge (\neg\neg\neg p_3 \vee \neg\neg p_4)) \\
&\equiv \neg p_1 \vee (\neg p_2 \wedge (\neg\neg\neg p_3 \vee \neg\neg p_4)) \\
&\equiv \neg p_1 \vee (\neg p_2 \wedge (\neg p_3 \vee \neg\neg p_4)) \\
&\equiv \neg p_1 \vee (\neg p_2 \wedge (\neg p_3 \vee p_4)).
\end{aligned}$$

Now we distribute. We see

$$\begin{aligned}
\varphi &\equiv \neg p_1 \vee (\neg p_2 \wedge (\neg p_3 \vee p_4)) \\
&\equiv (\neg p_1 \vee \neg p_2) \wedge (\neg p_1 \vee (\neg p_3 \vee p_4))) \\
&\equiv \boxed{(\neg p_1 \vee \neg p_2) \wedge (\neg p_1 \vee \neg p_3 \vee p_4)},
\end{aligned}$$

which is what we wanted. ∎

Exercise 2.64. We convert $\varphi := (p_1 \wedge p_2) \vee (\neg p_1 \wedge (\neg p_2 \wedge p_3))$ to conjunctive normal form using the algorithm.

Proof. We only have to distribute right now, so we compute

$$\begin{aligned}
\varphi &= (p_1 \wedge p_2) \vee (\neg p_1 \wedge (\neg p_2 \wedge p_3)) \\
&\equiv ((p_1 \wedge p_2) \vee \neg p_1) \wedge ((p_1 \wedge p_2) \vee (\neg p_2 \wedge p_3)) \\
&\equiv ((p_1 \vee \neg p_1) \wedge (p_2 \vee \neg p_1)) \wedge ((p_1 \wedge p_2) \vee (\neg p_2 \wedge p_3)) \\
&\equiv ((p_1 \wedge p_2) \vee \neg p_1) \wedge ((p_1 \vee (\neg p_2 \wedge p_3) \wedge (p_2 \vee (\neg p_2 \wedge p_3)))) \\
&\equiv ((p_1 \wedge p_2) \vee \neg p_1) \wedge ((((p_1 \vee \neg p_2) \wedge (p_1 \vee p_3)) \wedge ((p_2 \vee \neg p_2) \wedge (p_2 \vee p_3)))),
\end{aligned}$$

which finishes. ∎

2.6 February 23

Today we are talking about resolution.

2.6.1 Resolution

The goal of resolution is to compute if a formula is satisfiable. So far we have a fairly semantic algorithm, which is the truth table algorithm: list out all possibilities and then check each.

The big-picture question is "How can we get a machine to reason?" For example, once we know about satisfiability, we can ask the machine if φ is valid by seeing if $\neg\varphi$ is satisfiable. So to get a machine to reason, we can ask if a formula φ follows from the premises $\varphi_1, \dots, \varphi_n$ if and only if

$$(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \varphi$$

is valid if and only if

$$\neg\big((\varphi_1 \wedge \cdots \wedge \varphi_n) \to \varphi\big)$$

is not satisfiable.

The story so far is that we already have a couple ways to convert a formula into conjunctive normal form: either run the negation of the DNF algorithm, or we had the algorithm from last class which was more syntactic and essentially massaged all the symbols in place (where the key step was to apply distribution of $\vee$ over $\wedge$ to group clauses together). At a high level, CNF looks like

$$(\ell_{1,1} \wedge \cdots \ell_{1,m_1}) \vee \cdots \vee (\ell_{n,1} \wedge \cdots \ell_{n,m_n}),$$

where the $\ell_\bullet$ are literals. To make this formula true, we need one literal in each clause to be true.

For resolution, the key idea is as follows.

Idea 2.65. Suppose φ contains some subformula of the form $(p \vee \psi_1) \wedge (\neg p \vee \psi_2)$. Then $\psi_1 \vee \psi_2$ follows.

Here's an example of this idea in action.

Exercise 2.66. We do resolution on the formula

$$\varphi \coloneqq (p_1 \vee p_3 \vee p_4) \wedge (\neg p_1 \vee \neg p_2 \vee p_4) \wedge (p_1 \vee \neg p_2 \vee \neg p_3).$$

Proof. Observe that the clause $(p_1 \vee p_3 \vee p_4)$ and $(\neg p_1 \vee \neg p_2 \vee p_4)$ contain both p_1 and $\neg p_1$. So if both of these clauses are to be true, we had better not be checking the p_1 box, so it follows that

$$\psi \coloneqq p_3 \vee p_4 \vee \neg p_2 \vee p_4.$$

Explicitly, we claim $\varphi \vDash \psi$. We have two cases.

- If $V(p_1) = 0$, then $\hat{V}(\varphi) = 1$ forces $\hat{V}((p_1 \vee p_3 \vee p_4)) = 1$ forces $\hat{V}((p_3 \vee p_4))$, so $\hat{V}(\psi) = 1$.
- If $V(p_1) = 1$, then $\hat{V}(\varphi) = 1$ forces $\hat{V}((\neg p_1 \vee \neg p_2 \vee p_4)) = 1$ forces $\hat{V}((\neg p_2 \vee p_4))$, so $\hat{V}(\psi) = 1$.

To make our lives easier, we will take ψ and remove the redundancy p_4 to get $\psi = p_3 \vee p_3 \vee \neg p_2$.

This idea gives the following definition.

Definition 2.67 (Resolvent)**.** Fix φ a formula. If we have clauses $C_1 = p \vee \psi_1$ and $C_2 = \neg p \vee \psi_2$ in φ (or some rearrangement of p sitting anywhere in C_1 and C_2), then we call $\psi_1 \vee \psi_2$ a *resolvent* of φ by p.

Remark 2.68. The argument given above generalizes to show that any resolvent ψ of φ is a logical consequence of φ.

Here are the other resolvents of φ.

- In $(\neg p_1 \vee \neg p_2 \vee p_4)$ and $(p_1 \vee \neg p_2 \vee \neg p_3)$, we see we have p_1 and $\neg p_1$, so we get the resolvent $p_1 \vee p_4 \vee \neg p_2$.
- In $(p_1 \vee p_3 \vee p_4)$ and $(p_1 \vee \neg p_2 \vee \neg p_3)$, we see we have p_3 as well as $\neg p_3$, so we get the resolvent $\neg p_2 \vee p_4 \vee \neg p_3$.

Because all of our resolvents are logical consequences, we see that

$$\begin{aligned}\varphi \equiv (p_1 \vee p_3 \vee p_4) \wedge (\neg p_1 \vee \neg p_2 \vee p_4) \wedge (p_1 \vee \neg p_2 \vee \neg p_3) \\ (p_3 \vee p_3 \vee \neg p_2) \wedge (p_1 \vee p_4 \vee \neg p_2) \wedge (\neg p_2 \vee p_4 \vee \neg p_3).\end{aligned}$$

We now do more resolution on this new formula. The clauses $(p_3 \vee p_3 \vee \neg p_2)$ and $(\neg p_2 \vee p_4 \vee \neg p_3)$ resolve to $p_4 \vee \neg p_2$ by p_3. But now, when we write

$$\begin{aligned}\varphi \equiv (p_1 \vee p_3 \vee p_4) \wedge (\neg p_1 \vee \neg p_2 \vee p_4) \wedge (p_1 \vee \neg p_2 \vee \neg p_3) \\ (p_3 \vee p_3 \vee \neg p_2) \wedge (p_1 \vee p_4 \vee \neg p_2) \wedge (\neg p_2 \vee p_4 \vee \neg p_3) \\ (p_4 \vee \neg p_2),\end{aligned}$$

we can check that there are no more resolvents to add. We call this formula $\operatorname{res}(\varphi)$, with the following definition.

Definition 2.69 (Resolution)**.** Fix a formula φ. Then we define the *resolution* of φ by $\operatorname{res}(\varphi)$ to be φ closed under adding in resolvents.

Anyways, this finishes the resolution algorithm by the following relevant theorem; in particular, we see quickly that φ is satisfiable. ■

And here is the relevant theorem.

Theorem 2.70. A formula φ is unsatisfiable if and only if some clause of $\operatorname{res}(\varphi)$ has both p and $\neg p$ as clauses. This conclusion is called an overt contradiction.

Proof. We omit this. The easy direction is that, if $\operatorname{res}(\varphi)$ has both p and $\neg p$ as clauses, then of course there is no way to satisfy $\operatorname{res}(\varphi)$. ■

Remark 2.71. One might complain that this algorithm is quite inefficient. This is essentially to make it simpler; there are many optimizations that one could make by trying to add in fewer or more useful resolvents.

Example 2.72. Exercise 2.66 provides a formula φ with $\operatorname{res}(\varphi)$ which has no overt contradiction and hence is satisfiable.

Le's see another example of resolution.

Exercise 2.73. We apply resolution to

$$\varphi \coloneqq (p_1 \vee p_3) \wedge (\neg p_1 \vee p_2) \wedge (p_1 \vee \neg p_3) \wedge (\neg p_1 \vee \neg p_2).$$

Proof. We optimize with the following steps.

- The clauses $(\neg p_1 \vee p_3)$ and $(p_1 \vee \neg p_3)$ gives p_1.
- From $(\neg p_1 \vee p_2)$ and $(\neg p_1 \vee \neg p_2)$, we get $\neg p_1$.

But earlier we had p_1, so we have a contradiction. ■

In truth, there are likely many resolvents we would have to add in before finding the two resolvents given above, but we omitted them for brevity.

2.6.2 Complexity

It is somewhat annoying that determining if a formula φ is satisfiable requires a lot of computation. Consider the following remarks.

- If φ is satisfiable, then technically the truth table method merely requires finding the correct line of the truth table instead of the whole truth table.
- However, if φ is unsatisfiable, then we have to compute everything.
- If φ is unsatisfiable, then technically we only need to compute resolvents until we actually get an overt contradiction.
- However, if φ is satisfiable, then we have to compute everything.

The bad news is that the worst-case performance of both resolution and creating a truth table is exponential. Essentially this is because we have to continue checking for new resolvents, and each step can potentially add lots and lots of resolvents.

Example 2.74. With 10 variables, this would require $2^{10} = 1024$ rows of a truth table.

Here is the question we are interested in: can we make satisfiability faster?

Question 2.75. Is there an algorithm which runs in time bounded by a polynomial in n which can determine if a formula with n connectives is satisfiable?

Question 2.75 is perhaps the most famous unsolved problem in all of theoretical computer science. It is called the P vs. NP problem. Sadly, most people do not think that such an algorithm exists.

2.7 February 28

And we are back on the road.

2.7.1 More on Complexity

Last time we were talking about complexity and in particular Question 2.75, complaining that determining satisfiability seems to have an exponential blow-up in steps as the number of atomic formulae increases. For example, the truth-table method to compute satisfiability requires on the order of 2^n steps; most researchers do not think there is a way to do this much faster.

So let's talk more about this P vs. NP problem.

Definition 2.76 (P)**.** A problem is in *P* if and only if there is a deterministic polynomial-time algorithm to solve the problem. Namely, the algorithm run-time is polynomial in the input size.

Example 2.77. Deciding if a number is even runs in essentially constant time: simply check if the last digit is in $\{0, 2, 4, 6, 8\}$.

Example 2.78. There is an algorithm which can determine if a positive integer is prime in polynomial time.

Definition 2.79 (NP)**.** A problem is in *NP* if and only if there is a non-deterministic polynomial-time algorithm to solve the problem.

Roughly, a non-deterministic algorithm is allowed to guess and be lucky. Here are some examples.

Example 2.80. To check satisfiability of some formula φ, we can just guess a valuation and compute if $\hat{V}(\varphi) = 1$. Doing this check only runs in polynomial (in fact, linear) time with respect to the length of φ. Because life is easy after making this guess, we call the algorithm "non-deterministic" polynomial-time, where the adjective non-deterministic refers to the guess.

Example 2.81. It is difficult to solve a Sudoku puzzle—and there is an exponential blow-up in solving Sudoku as the grid gets bigger—but checking that the Sudoku works is comparatively easy—and checking does not get much harder as the problems gets bigger. Thus, there is a non-deterministic polynomial-time algorithm (guess some filled-in grid and check).

If we wanted to make satisfiability more analogous to Sudoku, we can transform the Sudoku problem as follows: given a partially filled-in $n^2 \times n^2$ Sudoku grid, determine if we can fill in the grid in some valid way. In particular, this is a binary-output and scalable problem like with satisfiability.

Remark 2.82. If we could show that merely satisfiability is in P, this would in fact show that all problems of NP would live in P. In particular, any problem in NP can be "easily" translated (i.e., translated in polynomial-time) into a satisfiability question.

The above remark is why satisfiability alone is enough to answer Question 2.75.

2.7.2 Scheduling via Satisfiability

Suppose we are trying to schedule lectures, so we are given a list of possible rooms and some list of lectures and some list of professor availabilities. Here would be our constraints.

1. Each lecture should be scheduled in some room in some time slot.
2. No two lectures should happen in the same room at the same time.
3. Two lectures given by the same professor cannot be given at the same time.
4. A lecture should not be scheduled at a time when the professor giving the lecture is unavailable.

To turn this into a satisfiability problem, we build some notation. We set $s_{i,j,k}$ to mean that lecture i is to be scheduled in room j during time slot k.

Remark 2.83. Availabilities and who gives lectures is not in the control of the scheduler. The only thing the scheduler can change is when the lecture happens, so the only atomic formulae we will deal with are $s_{i,j,k}$.

We now translate our constraints into propositional logic.

1. "Each lecture should be scheduled in some room in some time slot" becomes
$$\bigwedge_i \left(\bigvee_{j,k} s_{i,j,k} \right).$$
Namely, for each lecture i, some $s_{i,j,k}$ must be true for some room j and time slot k.

2. "No two lectures should happen in the same room at the same time" becomes

$$\bigwedge_{i \neq i', j, k} \neg(s_{i,j,k} \wedge s_{i',j,k}).$$

   In other words, for any two distinct lectures i and i', they are not both in the same room j and the same time slot k.

3. "Two lectures given by the same professor cannot be given at the same time" becomes

$$\bigwedge_{\substack{i \neq i',\ i,i' \text{ have the same professor} \\ j,k}} \neg(s_{i,j,k} \wedge s_{i',j,k}).$$

4. "A lecture should not be scheduled at a time when the professor giving the lecture is unavailable" becomes

$$\bigwedge_{\substack{i,j,k \\ \text{prof for } i \text{ unavailable at } k}} \neg s_{i,j,k}.$$

   In other words, the professor for lecture i should not be lecturing when available.

So to solve scheduling, we just hand over these constraints to a satisfiability solver, and it'll do it.

Remark 2.84. Scheduling turns out to not be so bad in comparison to worst-case scenarios for satisfiability. In particular, satisfiability is not so bad when there are "few enough" letters and the formula we are testing is not so bad.

Remark 2.85. Professor Holliday would like someone to use a satisfiability solver to run wedding planning (e.g., for dinner placement).

Theme 3
Natural Deduction

3.1 March 2

Welcome back everybody.

3.1.1 Proof

We are moving from "semantic" theory to more "proof-based" theory.

Recall that we already had a way to determine if an argument $\varphi_1, \dots \varphi_n \vDash \psi$ is by truth table: we simply check that the formula

$$(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \psi$$

is a tautology by truth table. However, this is pretty annoying when there are lots of propositions floating around and in fact impossible when there are infinitely many variables. So we will introduce the idea of formal proof.

Here is an example of a formal proof for some simple fact.

Proposition 3.1. Fix a and b real numbers. If $0 < a < b$, then $a^2 < b^2$.

Proof. We start by assuming $0 < a < b$. Because $0 < a$, we can multiply both sides of $a < b$ by a, which gives

$$a^2 < ab.$$

Similarly, $0 < a < b$ implies $0 < b$, so we can multiply both sides of $a < b$ by b, so

$$ab < b^2.$$

It follows $a^2 < ab$ and $ab < b^2$, so we do get $a^2 < b^2$, which is what we wanted. ∎

Remark 3.2. In the above proof, we showed a statement of the form $\varphi \to \psi$ by assuming φ and proving ψ.

To formalize the above, we will diagram our deduction to make a "sub-proof." This is called a Fitch-style formal proof, and they are pretty.

1	⊢	
2	$0 < a < b$	
⋮	⋮	
n	$a^2 < b^2$	
$n+1$	$(0 < a < b) \to a^2 < b^2$	$\to$Intro, 2, n

More generally speaking, a Fitch-style proof with premises $\varphi_1, \ldots, \varphi_n$ and proving ψ looks like the following.

1	φ_1
⋮	⋮
n	φ_n
⋮	⋮
m	ψ

The vertical dot empty spaces are to be filled with sub-proofs or formulae. Our goal is to put everything into this formal structure.

Here is an example of a rule, which is what we used above.

Definition 3.3 ($\to$Introduction)**.** If we have a sub-proof with assumption α and concludes β, then we can pop out of the sub-proof and add $\alpha \to \beta$ with the rule $\to$ with the named lines.

This looks like the following.

	⊢		
i		φ	
⋮		⋮	
j		ψ	
⋮	⋮		
k	$\varphi \to \psi$		$\to$Intro, i, j

3.1.2 Elimination of Conditionals

We can use conditionals to prove other conditionals. For example, suppose we have proven the following lemma.

Lemma 3.4. Fix a and b real numbers. If $0 < a < b$, then $a^2 < b^2$.

Then we can prove the following.

Lemma 3.5. Fix c and d real numbers. If $0 < d < c$, then $(c-d)^2 < c^2$.

Proof. As before, if $0 < c - d < c$, then $(c-d)^2 < c^2$ by the lemma. However, $d < c$ implies $0 < c - d$, and $0 < d$ then gives $0 < c - d < c$. So it does follow $(c-d)^2 < c^2$. ■

This rule is proving the antecedent from the consequent.

Definition 3.6 ($\to$Elimination)**.** If we have shown $\alpha \to \beta$ in some line i and have shown α in some line j, then we can conclude β from $\to$ *elimination*, citing i and j.

This looks like the following.

$\vdots$	$\vdots$	
i	$\alpha \to \beta$	
$\vdots$	$\vdots$	
j	α	
$\vdots$	$\vdots$	
k	β	$\to$Elim, i, j

We can even reverse these as follows.

$\vdots$	$\vdots$	
i	α	
$\vdots$	$\vdots$	
j	$\alpha \to \beta$	
$\vdots$	$\vdots$	
k	β	$\to$Elim, j, i

The point is to get the computer to be able to check this.

Example 3.7. This sort of reasoning is pretty intuitive: if we read "If you received Form 1098T, you may be eligible for a tax credit or deduction" and then we have received Form 1098T, then we can conclude that we may eligible. Here is a formalization.

1		
2	received 1098T $\to$ may be eligible for tax credit	
3	received 1098T	
4	may be eligible for tax credit	$\to$Elim, 3, 2

3.1.3 Reiteration

Let's return to the proof of Lemma 3.5. It used both $\to$ introduction and elimination. It looked like the following.

1	⊢	
2	$0 < c - d < c \to (c-d)^2 < c^2$	
3	$\quad 0 < d < c$	
⋮	$\quad 0 < c - d < c$	$(*)$
n	$\quad 0 < c - d < c \to (c-d)^2 < c^2$	Reit, 2
$n+1$	$\quad (c-d)^2 < c^2$	$\to$Elim, 3, 2
$n+2$	$0 < d < c \to (c-d)^2 < c^2$	$\to$Intro, 3, $n+1$

The point of our discussion is our use of the reiteration rule on line 5. Logically speaking, we note that the sub-proof technically lives in a different world than the rest of the outer proof, so we have a rule that says we can pull from the outside world into a sub-proof.

Here is our formal definition.

Definition 3.8 (Open)**.** A subproof is *open* starting from its first assumption until the conclusion of the subproof.

Definition 3.9 (Directly)**.** A formula φ occurs *directly* in a subproof if and only if φ occurs in the main column of S.

Example 3.10. In $(*)$, the lines 2 through 6 are open in the subproof.

Definition 3.11 (Reiteration)**.** The *reiteration* tells us that we may add φ to any subproof currently in the subproof.

Non-Example 3.12. The following is an incorrect use of reiteration.

1	⊢	
2	$\quad$ We are going to the party.	
3	$\quad$ We will have a good time.	
4	We will have a good time.	Reit, 3

Namely, this line does not live in the sub-proof, so we cannot pull it out.

3.2 March 4

Today we continue with natural deduction.

3.2.1 More on Reiteration

Today we continue with our proof system, which we will use for the homework on Sunday. Recall that we have the following incorrect use of reiteration.

1	⊢	
2	$\quad \alpha$	
3	$\quad \beta$	
4	β	Reit, 3

Namely, β only worked on the assumption of α, so the above is invalid. Here is a similarly bad example.

1	⊢	
2	$\quad \alpha$	
$\vdots$	$\quad \vdots$	
n	$\quad \beta$	
$n+1$	$\quad \gamma$	
$n+2$	$\quad \beta$	Reit, n

Namely, β only worked under the assumption of α, not under the assumption of γ necessarily.

In contrast, here is a correct use of reiteration.

1	$\quad p$	
2	$\quad\quad q$	
3	$\quad\quad p$	Reit, 1
4	$\quad q \to p$	$\to$Intro, 1, 3
5	$p \to (q \to p)$	$\to$Intro, 1, 4

The important point is that reiteration inside the subsubproof.

3.2.2 Proofs with the Conditional

So far we have the following rules for $\to$.

- $\to$ Introduction.
- $\to$ Elimination.
- Reiteration.

Let's actually prove something.

Proposition 3.13. We have that $(p \to q) \to ((q \to r) \to (p \to r))$.

Remark 3.14. This should intuitively make sense: if p implies q and q implies r, then p should imply r.

Proof. We apply brute force. We would like to end with $(p \to q) \to ((q \to r) \to (p \to r))$, so we should start with a subproof. Repeating this intuition, we have the following.

1	$p \to q$	
2	$\quad q \to r$	
3	$\quad\quad p$	
4	$\quad\quad p \to q$	Reit, 1
5	$\quad\quad q$	$\to$Elim, 3, 4
6	$\quad\quad q \to r$	Reit, 2
7	$\quad\quad r$	$\to$Elim, 5, 6
8	$\quad p \to r$	$\to$Intro, 3, 7
9	$q \to p$	$\to$Intro, 2, 8
10	$(p \to q) \to ((q \to r) \to (p \to r))$	$\to$Intro, 1, 9

This finishes. ∎

As an aside, we shouldn't really have to reiterate $p \to q$ in order to cite it because it was still up there in the hypotheses. So in the future, we will permit allowing to use elimination more liberally.

Definition 3.15 (Liberal elimination)**.** In a proof, we can add ψ if we have both φ and $\varphi \to \psi$ in the main column of the proof or directly in some subproofs that are currently open.

Essentially the above is saying that we could apply a liberal elimination whenever we could apply reiteration followed by elimination.

Example 3.16. The following proof works faster for the previous proposition.

1	$p \to q$	
2	$\quad q \to r$	
3	$\quad\quad p$	
4	$\quad\quad q$	L$\to$Elim, 1, 3
5	$\quad\quad r$	L$\to$Elim, 2, 4
6	$\quad p \to r$	$\to$Intro, 3, 5
7	$q \to p$	$\to$Intro, 2, 6
8	$(p \to q) \to ((q \to r) \to (p \to r))$	$\to$Intro, 1, 7

Non-Example 3.17. The following argument is not valid.

1	α	
2	$\varphi \to \psi$	
3	φ	
4	ψ	L$\to$Elim, 2, 3

So far our proof system is pretty weak, but at least we have conditionals. As example of things that we don't have, we don't have and, or, quantifiers, etc., but we'll add them later.

It turns out that our current proof system has the following nice property.

Definition 3.18 (Normal)**.** A proof is called *normal* if all applications of elimination precede all applications of an introduction.

In particular, we have the following.

Proposition 3.19. Any statement with only $\to$ can be proven by a normal proof.

Proof. Omitted. ■

Remark 3.20. Of course, we can have normal proofs for statements only using $\to$. For example, we could just add the proof

1	s	
2	s	Reit, 1

to the beginning of some normal proof.

Here are some more nice properties.

Definition 3.21 (Sound)**.** A proof system is *sound* if and only if any proof involving hypotheses $\varphi_1, \ldots, \varphi_n$ and concluding ψ, we have $\varphi_1, \ldots, \varphi_n \vDash \psi$. In other words, we can only prove true things.

Proposition 3.22. The above proof system is sound.

Proof. Omitted. We can more or less see this from our individual rules. ■

Definition 3.23 (Complete)**.** A proof system is *complete* if and only if having $\varphi_1, \ldots, \varphi_n \vDash \psi$, then there exists a proof with hypotheses $\varphi_1, \ldots, \varphi_n$ and concluding ψ. In other words, we can prove anything true.

Sadly, our proof system is currently not complete.

Example 3.24. Our proof system cannot prove $((p \to q) \to p) \to p$, which one can check is actually a tautology. The way that we do this is by changing the semantics of $\to$ so that the proof system will still always prove things but such that $((p \to q) \to p) \to p$ is actually false. In particular, it turns out that this formula implies the law of the excluded middle.

3.2.3 Rules for Conjunction

As before, we are going to have introduction and elimination rules, which answer the questions how to prove and how to use a conjunction. Here is an example.

Proposition 3.25. We show $3 < \sqrt{11} < 4$.

Proof. Note $3 < \sqrt{11}$ because $9 < 11$ implies $3 = \sqrt{9} < \sqrt{11}$. Note $\sqrt{11} < 4$ because $11 < 16$ implies $\sqrt{11} < \sqrt{16} = 4$. With both, we conclude $3 < \sqrt{11} < 4$. ■

Formally, the above proof would like the following.

$\vdots$	$\vdots$	
i	$3 < \sqrt{11}$	
$\vdots$	$\vdots$	
j	$\sqrt{11} < 4$	
$\vdots$	$\vdots$	
k	$3 < \sqrt{11} \wedge \sqrt{11} < 4$	$\wedge$-Intro, i, j

And here is that rule

Definition 3.26 ($\wedge$ Introduction)**.** If we have both φ and ψ, then we can conclude $\varphi \wedge \psi$ and cite the relevant rules.

Visually, this looks like one of the following.

$\vdots$	$\vdots$		$\vdots$	$\vdots$	
i	α		i	α	
$\vdots$	$\vdots$		$\vdots$	$\vdots$	
j	β		j	β	
$\vdots$	$\vdots$		$\vdots$	$\vdots$	
k	$\alpha \wedge \beta$	$\wedge$-Intro, i, j	k	$\beta \wedge \alpha$	$\wedge$-Intro, i, j

And here is an example proof.

Proposition 3.27. We show that $(p \to q), (p \to r)$ can prove $p \to (q \wedge r)$.

Proof. We proceed by brute force.

1	$p \to q$	
2	$p \to r$	
3	$\quad p$	
4	$\quad q$	L$\to$Elim, 1, 3
5	$\quad r$	L$\to$Elim, 2, 3
6	$\quad q \wedge r$	$\wedge$Intro, 4, 5
7	$p \to (q \wedge r)$	$\to$Intro, 3, 6

■

As usual, there is a liberal version, which we won't write out.

To prove a conjunction, we simply deduce one of the conjuncts. There was an example here which I did not have time to write out, but here is the rule.

Definition 3.28 ($\wedge$ Elimination)**.** If we have both $\varphi \wedge \psi$, then we can conclude φ or ψ and cite the relevant rules.

This looks like one of the following.

$\vdots$	$\vdots$		$\vdots$	$\vdots$	
i	$\alpha \wedge \beta$		i	$\alpha \wedge \beta$	
$\vdots$	$\vdots$		$\vdots$	$\vdots$	
k	α	$\wedge$-Elim, i	k	β	$\wedge$-Elim, i

Let's see another example.

Proposition 3.29. We show that from $p \to q$ we can show $(p \wedge r) \to (q \wedge r)$.

Proof. Here is our proof, which is mostly just mechanical.

1	$p \to q$	
2	$\quad p \wedge r$	
3	$\quad p \to q$	Reit, 1
4	$\quad p$	$\wedge$Elim, 2
5	$\quad q$	L$\to$Elim, 3, 4
6	$\quad r$	$\wedge$Elim, 2
7	$\quad q \wedge r$	$\wedge$-Intro, 5, 6
8	$(p \wedge r) \to (q \wedge r)$	$\to$Intro, 2–7

This works. ■

3.3 March 7

We roll. As a note on the midterm, we will be given a take-home, 48-hour midterm. It is not intended to take that long; it should be approximately a shorter problem set.

3.3.1 Deduction Theorem

We are still talking about natural deduction proofs. So far we have talked about how to introduce and eliminate the connectives $\to$ and $\wedge$. For example, the following proof uses all of our rules.

1	$p \to q$	
2	$\quad p \wedge r$	
3	$\quad p \to q$	Reit, 1
4	$\quad p$	$\wedge$Elim, 2
5	$\quad q$	L$\to$Elim, 3, 4
6	$\quad r$	$\wedge$Elim, 2
7	$\quad q \wedge r$	$\wedge$-Intro, 5, 6
8	$(p \wedge r) \to (q \wedge r)$	$\to$Intro, 2–7

In our story so far, we saw that $\{\varphi_1, \dots, \varphi_n\} \vDash \psi$ if and only if the formula $(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \psi$ is a tautology.

There is also a proof-theoretic analog of this statement.

Theorem 3.30 (Deduction)**.** There is a proof with assumptions $\{\varphi_1, \dots, \varphi_n\}$ and conclusion ψ if and only if there is a proof with no assumption $(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \psi$.

Proof. In one direction, suppose that we can prove $\{\varphi_1, \dots, \varphi_n\}$ gives ψ. Well, simply open a subproof to place $\varphi_1 \wedge \cdots \wedge \varphi_n$ into the hypotheses and use $\wedge$ elimination to deduce $\varphi_1, \dots, \varphi_n$. Then the old proof to conclude ψ still works. Visually, we are applying the following process.

			1	$\varphi_1 \wedge \cdots \wedge \varphi_n$	
1	φ_1		2	φ_1	$\wedge$Intro, 1
$\vdots$	$\vdots$		$\vdots$	$\vdots$	
n	φ_n	$\Longrightarrow$	n	φ_n	$\wedge$Intro, 1
$\vdots$	$\vdots$		$\vdots$	P	
m	ψ		m	ψ	
P			$m+1$	$(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \psi$	

In the other direction, $\{\varphi_1, \dots, \varphi_n\}$ as assumptions lets us conclude $(\varphi_1 \wedge \cdots \wedge \varphi_n)$, from which $\to$ elimination lets us conclude ψ. ∎

3.3.2 Biconditional

The biconditional $\varphi \leftrightarrow \psi$ simply means $(\varphi \to \psi) \wedge (\psi \to \varphi)$. As such, we introduce $\varphi \leftrightarrow \psi$ by having $\varphi \to \psi$ and $\psi \to \varphi$ and using $\wedge$ introduction. Formally, we have the following.

Definition 3.31 ($\leftrightarrow$Introduction)**.** If we have a subproof hypothesizing φ and proving ψ and another subproof hypothesizing ψ and proving φ, then we may conclude that $\varphi \leftrightarrow \psi$.

For example, we might write this as follows.

i	| φ	
$\vdots$	| $\vdots$	
j	| ψ	
$\vdots$	$\vdots$	
k	| ψ	
$\vdots$	| $\vdots$	
ℓ	| φ	
$\vdots$	$\vdots$	
n	$\varphi \leftrightarrow \psi$	$\leftrightarrow$Intro, i–j, k–ℓ

Here are some examples.

Exercise 3.32. We show $q \leftrightarrow q$.

Proof. The proof just uses the reiteration rule.

1	| q	
2	| q	Reit, 1
3	| q	
4	| q	Reit, 3
5	$q \leftrightarrow q$	$\leftrightarrow$Intro, 1–2, 3–4

This works. ■

Exercise 3.33. We show $(\varphi \to \psi) \wedge (\psi \to \varphi)$ lets us conclude $\varphi \leftrightarrow \psi$.

Proof.

1	$(\varphi \to \psi) \wedge (\psi \to \varphi)$	
2	$\quad \varphi$	
3	$\quad \varphi \to \psi$	$\wedge$Elim, 1
4	$\quad \psi$	$\to$Elim, 2, 3
5	$\quad \psi$	
6	$\quad \psi \to \varphi$	$\wedge$Elim, 1
7	$\quad \varphi$	$\to$Elim, 5, 6
8	$\varphi \leftrightarrow \psi$	$\leftrightarrow$Intro, 2–4, 5–7

■

Now, to use $\leftrightarrow$, we have the following.

$\vdots$	$\vdots$	
i	$\varphi \leftrightarrow \psi$	
$\vdots$	$\vdots$	
j	φ	
$\vdots$	$\vdots$	
n	ψ	$\leftrightarrow$Elim, i, j

Essentially, having one of them lets you conclude the other. As usual, this can be reduced to $(\varphi \to \psi) \wedge (\psi \to \varphi)$. We codify this as follows.

Definition 3.34 ($\leftrightarrow$Elimination)**.** If we have $\varphi \leftrightarrow \psi$ and φ, then we can conclude ψ.

3.3.3 Negation Introduction

The idea to introduce $\neg$, we have to use proof by contradiction: to prove $\neg\varphi$, we assume φ and derive a contradiction.

Let's see the most famous example of a proof by contradiction.

Proposition 3.35. We show $\sqrt{2}$ is irrational.

Proof. Suppose for the sake of contradiction that $\sqrt{2}$ is rational. Well, then we can write $\sqrt{2} = \frac{a}{b}$ for integers a and b such that $b \neq 0$ and a and b cannot be reduced. Squaring, we see that $2 = \frac{a^2}{b^2}$, so

$$a^2 = 2b^2.$$

Now, a^2 is even, so a is even (which we will more formally justify later in the course), so we can write $a = 2k$. Plugging in, we see that

$$4k^2 = (2k)^2 = a^2 = 2b^2,$$

so $b^2 = 2k^2$. But now b^2 is even, so b is even! Thus, the fraction $\frac{a}{b}$ has even numerator and denominator, so it could have been reduced, which is our contradiction. ■

At a high level, the proof looked like the following.

1	$\sqrt{2} \in \mathbb{Q}$	
⋮	⋮	
n	$\frac{a}{b}$ is reducible $\wedge$ $\frac{a}{b}$ is not reducible	
$n+1$	$\neg(\sqrt{2} \in \mathbb{Q})$	$\neg$Intro, 1–n

This is our introduction rule.

Definition 3.36 ($\neg$Intro)**.** If we have a subproof starting with φ and concluding $\beta \wedge \neg\beta$ for some formula β, then we can conclude $\neg\varphi$.

Remark 3.37. Many texts write $\beta \wedge \neg\beta$ as $\bot$ to mean "false."

In a diagram, such a proof would look as follows.

⋮	⋮	
i	φ	
⋮	⋮	
j	$\psi \wedge \neg\psi$	
⋮	⋮	
n	$\neg\varphi$	$\neg$Intro, i–j

Here is an example.

Exercise 3.38. From $p \to q$ we may conclude $\neg q \to \neg p$.

Proof. We simply use $\to$ introduction and $\neg$ introduction for fun and profit.

1	$p \to q$	
2	$\neg q$	
3	p	
4	q	$\to$Elim, 1, 3
5	$\neg q$	Reit, 2
6	$q \wedge \neg q$	$\wedge$Intro, 4, 5
7	$\neg p$	$\neg$Intro, 3–6
8	$\neg q \to \neg p$	$\to$Intro, 5–7

This works. ■

Remark 3.39. By the deduction theorem, this also proves $(p \to q) \to (\neg q \to \neg p)$.

In fact, one can try to prove that $\neg q \to \neg p$ can prove $p \to q$, but for this we need to know that $\neg\neg p \leftrightarrow p$, which we have to be careful about.

Now, there are two ways to talk about "negation elimination." Here is one way.

Definition 3.40 (¬Elimination)**.** If we are given $\neg\varphi$, then we may infer $\varphi \to \psi$ for any ψ.

Essentially, the idea is that the antecedent is false, so the conditional is true.

Proposition 3.41. We show that the empty set $\varnothing$ is a subset of a set A.

Proof. Fix A any set. Now, for any x, we have $x \notin \varnothing$. So (vacuously) $x \in \varnothing$ implies $x \in A$ by our elimination rule. This finishes. ■

We can represent this proof as follows.

1	$\neg(x \in \varnothing)$	
2	$(x \in \varnothing) \to (x \in A)$	¬Elim, 1

3.4 March 9

Here we go.

3.4.1 Negation Elimination

There are two forms of negation elimination. Here is one way.

Definition 3.40 (¬Elimination)**.** If we are given $\neg\varphi$, then we may infer $\varphi \to \psi$ for any ψ.

Diagrammatically, this looks like the following.

$\vdots$	$\vdots$	
n	$\neg\alpha$	
$\vdots$	$\vdots$	
$\alpha \to \beta$ $\vdots$ $+1$	b	¬Elim, n

Here is another form of negation elimination.

Definition 3.42 (Ex Falso Quodlibet)**.** From a contradiction that $\alpha \wedge \neg\alpha$, any formula β can be introduced.

This looks like the following.

$\vdots$	$\vdots$	
n	$\alpha \wedge \neg\alpha$	
$\vdots$	$\vdots$	
β $\vdots$ $+1$	b	EFQ, n

In theory, we should only introduce a contradiction in a subproof.

Remark 3.43. Any application of ex falso quodlibet can be replaced with $\wedge$ elimination, $\neg$ elimination, and $\to$ elimination. Conversely, $\neg$ elimination be can be replaced with $\to$ introduction, reiteration, $\wedge$ introduction, and ex falso quodlibet. We will not prove this.

To manifest this remark, we show how to get $\neg$ elimination from the others.

$\vdots$	$\vdots$		
n	$\neg\alpha$		
$\vdots$	$\vdots$		
m		α	
$m+1$		$\neg\alpha$	Reit, n
$m+2$		$\alpha \wedge \neg\alpha$	$\wedge$Intro, m, $m+1$
$m+3$		β	EFQ, $m+2$
$m+4$	$\alpha \to \beta$		$\to$Intro, m, $m+3$

Let's see an example.

Exercise 3.44. We show that $\neg p$ and $q \to p$ can prove $q \to r$.

Proof. We proceed as follows.

1	$\neg p$		
2	$q \to p$		
3		q	
4		$q \to p$	Reit, 2
5		p	$\to$Elim, 3, 4
6		$\neg p$	Reit, 1
7		$p \wedge \neg p$	$\wedge$Intro, 5, 6
8		r	EFQ, 7
9	$q \to r$		$\to$Intro, 3–8

This proof is valid. ∎

Remark 3.45. Our proof system is still not complete. For example, $((p \to q) \to p) \to p$ is still a valid formula but not provable in this proof system.

3.4.2 Reductio Ad Absurdum

To complete our proof system, we need to introduce one more rule. Let's review the following proof.

Proposition 3.41. We show that the empty set $\varnothing$ is a subset of a set A.

We showed this by assuming φ, deriving a contradiction, and then concluded $\neg\varphi$. In contrast, reductio ad absurdum assumes $\neg\varphi$, derives a contradiction, and then gives φ. Namely, the number of $\neg$s is decreasing.

Let's give an example.

Proposition 3.46. There are irrational numbers a and b such that a^b is rational.

Proof. Assume for the sake of contradiction that there are no such irrational numbers. Now, we know that $\sqrt{2}$ is irrational, so $\sqrt{2}^{\sqrt{2}}$ is irrational as well (from a contraposition). But then

$$\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = (\sqrt{2})^{\sqrt{2}\cdot\sqrt{2}} = (\sqrt{2})^2 = 2$$

is also irrational, which is a contradiction. This contradicts the assumption. So instead, there must be irrational numbers a and b such that a^b is rational. ■

Remark 3.47. This proof is nonconstructive: we don't actually have examples of irrational numbers a and b such that $a^b \in \mathbb{Q}$.

To be explicit, either $(\sqrt{2})^{\sqrt{2}}$ is our example or $\left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}}$ is our example, and we don't know which.

To see that this proof is different from what we had before, note the proof above looks like the following.

1		$\neg$(there are $a, b \notin \mathbb{Q}$ with $a, b \in \mathbb{Q}$
$\vdots$		$\vdots$
n		$\bot$
$n+1$	there are $a, b \notin \mathbb{Q}$ with $a, b \in \mathbb{Q}$	
$n+2$	(there are $a, b \notin \mathbb{Q}$ with $a, b \in \mathbb{Q}$) $\wedge$ $\neg$(there are $a, b \notin \mathbb{Q}$ with $a, b \in \mathbb{Q}$)	$\wedge$Intro, $n+1$, 1

Indeed, we can see that we stripped off a $\neg$, which is more powerful than typical $\neg$ introduction.

Formally, here is our description.

Definition 3.48 (Reductio ad absurdum)**.** If we assume $\neg\varphi$ (in a subproof) and derive a contradiction, then we can prove φ.

Warning 3.49. Reductio ad absurdum is different from $\neg$ introduction: the number of negations goes down from $\neg\varphi$ to φ, but $\neg$ introduction increases the number of negations.

Here's an example.

Exercise 3.50. From $\neg q \to \neg p$, we prove $p \to q$.

Proof. We proceed as follows.

1	$\neg q \to \neg p$	
2	$\quad p$	
3	$\quad\quad \neg q$	
4	$\quad\quad \neg q \to \neg p$	Reit, 1
5	$\quad\quad \neg p$	$\to$Elim, 4, 3
6	$\quad\quad p$	Reit, 2
7	$\quad\quad p \wedge \neg p$	$\wedge$Intro, 2, 5
8	$\quad q$	RAA, 3–7
9	$p \to q$	$\to$Intro, 2–8

Notably, by $\neg$ introduction, we would still be able to conclude $\neg\neg q$ just not the formula q. ■

3.4.3 Contraposition

Some inferences are not primitive, but they are sufficiently common to have their own names.

Definition 3.51 (Contraposition)**.** If we can prove $\neg q \to \neg p$, then we can prove $p \to q$ by *contraposition*.

This is not a primitive, but we can always apply this schematic.

Lemma 3.52. Fix n an integer. Then n is odd (i.e., of the form $2k+1$) if and only if n is not even (i.e., of the form $2k$).

Proof. Omitted; we will discuss this more next class. ■

Let's show the following.

Lemma 3.53. Fix n an integer. If n^2 is even, then n is even.

Proof. We proceed by contraposition: suppose that n is not even, and we show n^2 is not even. Well, because n is not even, we get to say $n = 2k+1$, so

$$n^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 2\left(2k^2 + 2k\right) + 1,$$

so we see that n^2 is odd. It follows n^2 is not even, so we are done. ■

Remark 3.54. Now that we've added in reductio ad absurdum, our proof system is complete: any valid formula is provable.

To manifest this remark, let's show Pierce's law that $((p \to q) \to p) \to p$, which is not provable without reductio ad absurdum even though it only uses $\to$.

Exercise 3.55. We show $((p \to q) \to p) \to p$.

Proof. We proceed as follows.

1	$(p \to q) \to p$	
2	$\quad \neg p$	
3	$\quad p \to q$	$\neg$Elim, 2
4	$\quad (p \to q) \to p$	Reit, 1
5	$\quad p$	$\to$ELim, 3, 4
6	$\quad p \wedge \neg p$	$\wedge$Intro, 5, 2
7	p	RAA, 2–6
8	$((p \to q) \to p) \to p$	$\to$Intro, 1–7

This works. ∎

3.5 March 11

We continue our journey in natural deduction.

3.5.1 Disjunction Introduction

The story so far as that we have given introduction and elimination rules for $\to$ and $\wedge$ and $\leftrightarrow$ and $\neg$. It remains to talk about $\vee$. We have also introduced reductio ad absurdum, which made our proof system complete.

Remark 3.56. A computer could theoretically use this system to prove statements. One silly way is to just list out all possible sequences of steps in a Fitch-style proof and just check until we find a well-formed proof of the conclusion.

Here is our way to introduce disjunctions.

Definition 3.57 ($\vee$Introduction)**.** Give φ, we can introduce $\varphi \vee \psi$ or introduce $\psi \vee \varphi$.

More formally, either of the following are valid uses of $\vee$ introduction.

$\vdots$	$\vdots$			$\vdots$	$\vdots$		
i	φ			i	ψ		
$\vdots$	$\vdots$			$\vdots$	$\vdots$		
j	$\varphi \vee \psi$	$\vee$Intro, i		j	$\varphi \vee \psi$	$\vee$Intro, i	

A priori, it might look like we are only even weakening in this step, so let's see an example where this is useful.

Lemma 3.52. Fix n an integer. Then n is odd (i.e., of the form $2k+1$) if and only if n is not even (i.e., of the form $2k$).

Proof. Given a nonnegative integer n, we need to show that $n = 2k$ for some integer k or that $n = 2k+1$ for some integer k. We prove this by induction on n.

- For our base case, we note that $n = 0 = 2 \cdot 0$, so n is even and in particular even or odd.
  Note here is where we applied weakening! After all, we need to go back to our inductive step somehow.
- We will do the inductive step later when we talk about disjunctive elimination. ■

Diagrammatically, the base case of the proof looks like the following.

$\vdots$	$\vdots$	
i	0 is even	
$i+1$	$(0 \text{ is even}) \vee (0 \text{ is odd})$	$\vee$Intro, i

So yes, we are weakening, but we had to.

There are other ways—non-constructive ways!—to prove disjunctions without $\vee$ introduction by reductio ad absurdum. Here's an example.

Exercise 3.58. We deduce $p \vee \neg p$.

Proof. We will never be able to prove p or $\neg p$ on its own, but we will still be able to derive $p \vee \neg p$. Instead, we use reductio ad absurdum, proceeding as follows; the key trick is that our contradiction will be $(p \vee \neg p) \wedge \neg(p \vee \neg p)$, from which we can work backwards, trying to prove $(p \vee \neg p)$ and derive contradiction.

1	$\neg(p \vee \neg p)$	
2	$\quad p$	
3	$\quad p \vee \neg p$	$\vee$Intro, 2
4	$\quad \neg(p \vee \neg p)$	Reit, 4
5	$\quad (p \vee \neg p) \wedge \neg(p \vee \neg p)$	$\wedge$Intro, 3, 4
6	$\neg p$	$\neg$Intro, 2, 5
7	$p \vee \neg p$	$\vee$Intro, 6
8	$(p \vee \neg p) \wedge \neg(p \vee \neg p)$	$\wedge$Intro, 7, 1
9	$p \vee \neg p$	RAA, 1–8

Observe that we could show $\neg\neg(p \vee \neg p)$ by simply using $\neg$ introduction at the end instead of reductio ad absurdum. ■

Remark 3.59. Intuitionist logic bans reductio ad absurdum, essentially because it is annoying that proving $p \vee \neg p$ doesn't actually tell you which one which is true; in fact, it is impossible to prove p or $\neg p$ from no assumptions because these formulae need not be valid, and we know that our proof system is sound.

However, removing reductio ad absurdum is nice in some sense: for example, without reductio ad absurdum, any proof of $\varphi \vee \psi$ must actually prove φ or prove ψ.

Remark 3.60. Any theorem φ provable with reductio ad absurdum, one can prove $\neg\neg\varphi$ without reductio ad absurdum.

3.5.2 Proof by Cases

So how do we use $\alpha \vee \beta$? After all, we don't know which one is true. But never fear—if we have some φ with both $\alpha \to \varphi$ and $\beta \to \varphi$, then surely we can conclude φ. This is called "proof by cases," where we try α and then try β, and we see that both of them go the conclusion.

Let's see an example.

Lemma 3.61. Let n be a nonnegative integer. Then the remainder when n^2 is divided by 4 is either 0 or 1.

Proof. By Lemma 3.52, we know that n is even or odd. So we proceed by cases.

- Suppose that n is even so that $n = 2k$. Then $n^2 = (2k)^2 = 4k^2$, so the remainder when n is divided by 4 is 0, so we can introduce a disjunction and assert that the remainder is 0 or 1.
- Suppose that n is odd so that $n = 2k + 1$. Then

$$n^2 = (2k+1)^2 = 4k^2 + 4k + 1 = 4\left(k^2 + k\right) + 1,$$

  so the remainder when n is divided by 4 is 0, so we can introduce a disjunction and assert that the remainder is 0 or 1.

Thus, both cases were able to derive the conclusion, so we are done. ■

In a Fitch style proof, this looks like the following.

1	$(n$ is even$) \vee (n$ is odd$)$
2	$\quad n$ is even
$\vdots$	$\quad \vdots$
i	$\quad n^2$ has remainder 0
$i+1$	$\quad (n^2$ has remainder $0) \vee (n^2$ has remainder $1)$
$i+2$	$\quad n$ is odd
$\vdots$	$\quad \vdots$
j	$\quad n^2$ has remainder 1
$j+1$	$\quad (n^2$ has remainder $0) \vee (n^2$ has remainder $1)$
$j+2$	$(n^2$ has remainder $0) \vee (n^2$ has remainder $1)$

Formally, we have the following.

Definition 3.62 ($\vee$Elimination)**.** Given $\alpha \vee \beta$ and two subproofs deducing φ from α and φ from β, then we can deduce φ.

Diagrammatically, this looks like the following.

a	$\alpha \vee \beta$	
b	$\quad \alpha$	
$\vdots$	$\quad \vdots$	
c	$\quad \varphi$	
$\vdots$	$\vdots$	
d	$\quad \beta$	
$\vdots$	$\quad \vdots$	
e	$\quad \varphi$	
$\vdots$	$\vdots$	
n	φ	$\vee$Elim, a, b–c, d–e

We now close by discussing the inductive step for Lemma 3.52.

Proof of the inductive step in Lemma 3.52. We know that n is even or odd, so we have the following two cases.

- If n is even, then set $n = 2k$, so $n + 1 = 2k + 1$, so n is odd, so n is even or odd.
- If n is odd, then set $n = 2k + 1$, so $n + 1 = (2k + 1)$, so n is even, so n is even or odd.

It follows that $n + 1$ is even or odd, finishing our inductive step. ■

3.6 March 14

Here we go.

3.6.1 Finishing Disjunction Elimination

Today we finish natural deduction by talking about syllogistic logic. Recall the following definition.

Definition 3.62 ($\vee$Elimination)**.** Given $\alpha \vee \beta$ and two subproofs deducing φ from α and φ from β, then we can deduce φ.

Here is an example from propositional logic.

Exercise 3.63. We derive $p \vee q$ from $\neg p \to q$.

Proof. We want to do a proof by cases. Recall that we have shown $p \vee \neg p$ (from reductio ad absurdum in Exercise 3.58), so we will just start from there; formally, we should copy and paste the whole proof in, but

we won't bother. The point is that the conclusion has $p \vee q$, so we need to get a p there somehow.

1	$\neg p \to q$	
2	$p \vee \neg p$	
3	$\quad p$	
4	$\quad p \vee q$	$\vee$Intro, 3
5	$\quad \neg p$	
6	$\quad \neg p \to q$	Reit, 1
7	$\quad q$	$\to$Elim, 5, 6
8	$\quad p \vee q$	$\vee$Intro, 8
9	$p \vee q$	$\vee$Elim, 2, 3–4, 5–8

Again, we see that $\vee$ introduction is crucial. ■

To review, we see that we have now built a full proof system for our propositional logic. With reductio ad absurdum, this proof system is sound (it only proves true statements) and complete (it can prove any true statement).

Remark 3.64. More precisely, a formula φ can be called *provable* if and only if there is a proof of φ via our proof system. Then it is a powerful theorem that provability is equivalent to validity. For example, a computer can (somewhat) quickly check a proof, which is nice.

3.6.2 Contradiction

We are now essentially done with propositional logic, but we'll add in a small bonus. The point here is that many of our rules (namely, $\neg$ introduction, ex falso quodlibet, and reductio ad absurdum) have contradictions of the form $\varphi \vee \neg\varphi$ play a crucial role.

It will be psychologically easier to introduce the following notion.

Definition 3.65 (Falsum, bottom)**.** We introduce the symbol $\perp$ to our propositional language $\mathcal{L}(P)$ by deciding to live in $P \cup \{\perp\}$ as our set of atomic formulae.

We will have $\perp$ "stand in for" any contradiction. Explicitly, in order to extend our valuations to $\perp$, we will require that

$$V(\perp) = 0$$

for any valuation $V : P \to \{0, 1\}$. In other words, $\perp$ is always false.

To use our $\perp$ symbol, we now need introduction and elimination rules.

Definition 3.66 ($\perp$Introducion)**.** Given α and $\neg\alpha$ in a partial proof, we may introduce $\perp$.

Diagrammatically, this looks like the following.

$$\begin{array}{l|l|l} \vdots & \vdots & \\ i & \alpha & \\ \vdots & \vdots & \\ j & \neg\alpha & \\ \vdots & \vdots & \\ k & \bot & \bot\text{Intro, } i, j \end{array}$$

Namely, we can introduce $\bot$ from any explicit contradiction.

Here is our elimination rule.

Definition 3.67 ($\bot$Elimination)**.** Given $\bot$, we may introduce any formula ψ.

Diagrammatically, this looks like the following.

$$\begin{array}{l|l|l} \vdots & \vdots & \\ i & \bot & \\ \vdots & \vdots & \\ j & \psi & \bot\text{Elim, } i \end{array}$$

This is essentially ex falso quodlibet, where we took a contradiction $\varphi \wedge \neg\varphi$ and derived any formula.

Now, the symbol $\bot$ will let us remove the explicit contradictions from our proof system. Our construction of $\bot$ elimination was essentially ex falso quodlibet, which means that we have to talk about $\neg$ introduction and reductio ad absurdum. The point is to take our definitions and replace the explicit contradictions with $\bot$. Here are our rules.

Definition 3.68 ($\neg$Introduction)**.** If a subproof with hypothesis φ derives $\bot$, we can deduce $\neg\varphi$.

Definition 3.69 (Reductio ad absurdum)**.** If a subproof with hypothesis $\neg\varphi$ derives $\bot$, we can deduce φ.

This looks like the following.

$$\begin{array}{l|l|l} \vdots & \vdots & \\ i & \left|\begin{array}{l} \neg\varphi \\ \hline \vdots \\ \bot \end{array}\right. & \\ \vdots & \vdots & \\ k & \varphi & \text{RAA, } i\text{–}j \end{array}$$

Let's see an example.

Exercise 3.70. If we have $p \vee q$ and $\neg p \to \neg q$, then we can deduce p.

Proof. Here is the first version of the proof, without $\bot$.

1	$p \vee q$	
2	$\neg p \to \neg q$	
3	$\quad \neg p$	
4	$\quad \neg q$	L$\to$Elim, 3, 2
5	$\quad\quad p$	
6	$\quad\quad p \wedge \neg p$	L$\wedge$Intro, 5, 3
7	$\quad\quad q$	
8	$\quad\quad q \wedge \neg q$	L$\wedge$Intro, 7, 4
9	$\quad\quad p \wedge \neg p$	EFQ, 8
10	$\quad p \wedge \neg p$	$\vee$Elim, 1, 5–6, 7–9
11	p	RAA, 3–10

What is inelegant here is that we have to move between contradictions $q \wedge \neg q$ to $p \wedge \neg p$. To remove this inelegance, we can use $\bot$.

1	$p \vee q$	
2	$\neg p \to \neg q$	
3	$\quad \neg p$	
4	$\quad \neg q$	L$\to$Elim, 3, 2
5	$\quad\quad p$	
6	$\quad\quad \bot$	L$\bot$Intro, 5, 3
7	$\quad\quad q$	
8	$\quad\quad \bot$	L$\bot$Intro, 7, 4
9	$\quad \bot$	$\vee$Elim, 1, 5–6, 7–8
10	p	RAA, 3–9

So we got to save one line, but also our proof looks a little cleaner. ■

This finishes contradiction. The midterm will only use content from our discussion on propositional logic.

Part II

Predicate Logic

Theme 4
Monadic Predicate Logic

4.1 March 16

Today we start moving towards predicate logic.

4.1.1 Examples of Syllogisms

To smooth over our movement to predicate logic, we will talk about syllogistic logic. Let's start with some examples.

Example 4.1. Here is a valid syllogism.

1. All sophomores are students.
2. All students are invited to the game.
3. Therefore, all sophomores are invited to the game.

Example 4.2. More generally, here is a valid syllogism.

1. All A are B.
2. All B are C.
3. Therefore, all A are C.

Remark 4.3. Notably, we are using capital letters A, B, and C because they are properties, not statements.

Example 4.4. Here is another valid syllogism.

1. All those invited to the game are invited to the after party.
2. No faculty are invited to the after party.
3. Therefore, no faculty are invited to the game.

Example 4.5. More generally, here is a valid syllogism.

1. All A are B.
2. No C are B.
3. Therefore, no C is A.

Non-Example 4.6. Here is an invalid syllogism.

1. All students are invited to the game.
2. No faculty are students.
3. Therefore, no faculty are invited to the game.

Explicitly, it is possible for a faculty to be invited to the game while not violating our premises.

4.1.2 Models

Syllogistic logic has its own language, inductively created just as for propositional logic.

Definition 4.7 (Syllogistic formula)**.** Fix Pred a set of predicates. Then for any three predicates A, B, C, the following three are all predicates.

- $\mathrm{All}\ A\ \mathrm{are}\ B$.
- $\mathrm{Some}\ A\ \mathrm{are}\ B$.
- $\mathrm{No}\ A\ \mathrm{are}\ B$.
- $\mathrm{Not\ all}\ A\ \mathrm{are}\ B$.

Remark 4.8. We have not allowed for our predicates to have propositional properties; for example, we cannot say that $\mathrm{All}\ A\ \mathrm{are}\ B\ \mathrm{or}\ C$ yet. This will come in first-order logic.

Remark 4.9. The point of dealing with syllogistic logic right now is that we get to just focus on the quantifiers without dealing with propositional logic.

Note that we are calling these predicates, not propositions or formulae.

Warning 4.10. A predicate is a property, not a statement with a truth value.

More explicitly, a predicate's meaning is really made of the set of sophomores. That's where all of its meaning come from: what is a sophomore, and what isn't.

In particular, to determine the truth value of some predicated like $\mathrm{All}\ A\ \mathrm{are}\ B$, we need to know about A and B themselves. For this, we introduce the idea of a model. We have already seen models, in some sense.

Definition 4.11 (Model)**.** A *model* for a propositional language $\mathcal{L}(P)$ is a valuation $V : P \to \{0, 1\}$.

Namely, a model is how we are able to determine truth values.

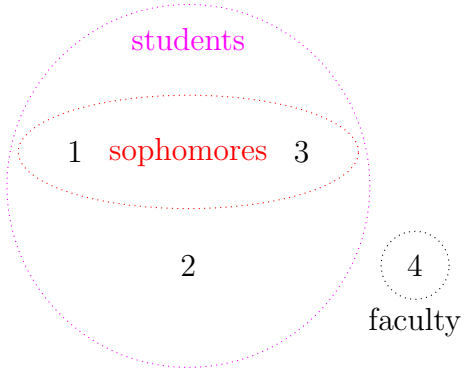
For syllogistic logic, our notion of model changes a little.

Definition 4.12 (Model)**.** A *model* $\mathcal{M} = (D, I)$ for our syllogistic language consists of a nonempty set D and a function $I : \text{Pred} \to \mathcal{P}(D)$ that sends each predicate A to a subset $I(A) \subseteq D$. Then $I(A)$ is the *extension* or *interpretation* of A in $\mathcal{M}$.

Example 4.13. Fix our base predicates consisting of sophomores, students, faculty, and invitees to the game. Then with four people $D = \{1, 2, 3, 4\}$, we could have the following interpretations.

- $I(\text{Sophomore}) = \{1, 3\}$.
- $I(\text{Students}) = \{1, 2, 3\}$.
- $I(\text{Faculty}) = \{4\}$.
- $I(\text{Invitees}) = \{1, 2, 3, 4\}$.

One could visualize this more graphically as a diagram with blobs and such, as follows; here everyone is an invitee, so we have not drawn this in.



We also have a notion of truth in a model, which we will go ahead and define perhaps insultingly carefully.

Definition 4.14 (Truth)**.** Fix a formula φ in syllogistic logic. Then the formula φ is *true* in the model $\mathcal{M} = (D, I)$, denoted $\mathcal{M} \vDash \varphi$ if and only if $\mathcal{M}$ makes φ true, in the following sense, defined inductively.

- $\mathcal{M} \vDash \text{All } A \text{ are } B$ if and only if $I(A) \subseteq I(B)$.
- $\mathcal{M} \vDash \text{Some } A \text{ are } B$ if and only if $I(A) \cap I(B) \neq \varnothing$.
- $\mathcal{M} \vDash \text{No } A \text{ are } B$ if and only if $I(A) \cap I(B) = \varnothing$.
- $\mathcal{M} \vDash \text{Not all } A \text{ are } B$ if and only if $I(A) \not\subseteq I(B)$.

Remark 4.15. For $\mathcal{M} \vDash \text{Not all } A \text{ are } B$ to be true, we actually need to have something in A.

Example 4.16. We return to our model from Example 4.13.

- $\mathcal{M} \vDash \text{All sophomores are students}$ because $I(\text{sophomore}) \subseteq I(\text{students})$.
- $\mathcal{M} \vDash \text{Some invitees to the game are sophomores}$ because 1 is both a sophomore and an invitee.
- $\mathcal{M} \vDash \text{No students are faculty}$ because $I(\text{students}) \cap I(\text{faculty}) = \{1, 2, 3\} \cap \{4\} = \varnothing$.
- $\mathcal{M} \vDash \text{Not all invitees to the game are students}$ because 4 is not a student but is an invitee.

Remark 4.17. We are allowed to have the model be false, in the real-world sense, but that is still a valid model. For example, we could have sophomores who are not students, even though this does not make sense in the real world.

4.1.3 Validity

From a notion of truth, we can bring in a notion of validity.

Definition 4.18 (Valid)**.** An argument consisting of premises $\{\varphi_1, \dots, \varphi_n\}$ and conclusion ψ is *valid* if and only if every model $\mathcal{M}$ which satisfy the premises $\varphi_1, \dots, \varphi_n$ also satisfy ψ.

Remark 4.19. This is really the same notion of validity as in propositional logic, except that models in propositional logic are valuations.

Example 4.20. The following argument is valid, for any predicates A, B, C.

1. All A are B.
2. All B are C.
3. Therefore, All A are C.

Indeed, fix any model $\mathcal{M} = (D, I)$ which satisfies the premises. Then the premises give $I(A) \subseteq I(B)$ and $I(B) \subseteq I(C)$, so $I(A) \subseteq I(C)$, so the conclusion also holds.

Non-Example 4.21. The following argument is invalid.

1. All A are B.
2. No C are A.
3. No C are B.

Indeed, use the previous model with A being students and B being invitees and C being faculty. We have the following checks.

1. $\mathcal{M} \vDash$ All students are invitees to the game.
2. $\mathcal{M} \vDash$ No faculty are students.
3. $\mathcal{M} \nvDash$ No faculty are invitees to the game because 4 is invited to the game but also a faculty.

Remark 4.22. As with propositional logic, validity in syllogistic logic is decidable, which is essentially because we only have to check smallish models for validity.

Remark 4.23. It is possible to define a sound and complete proof system for our syllogistic logic, just as with propositional logic.

We close by saying that we can extend syllogisms with our propositional logic to make a more complex language, to allow statements like

$$\neg(\text{Some } A \text{ are } (B \vee C)).$$

We know how to add connectives to our language, so the main point is figuring out how to add these connectives to our notion of validity, which we know how to do from propositional logic.

4.2 March 18

Welcome to Friday before spring break.

4.2.1 Finishing Syllogistic Logic

To close up syllogistic logic, we introduce propositional logic to syllogistic logic, to set up our discussion of predicate logic.

Definition 4.24 (Syllogistic propositional formula)**.** A *syllogistic propositional formula* consists of syllogistic propositional formulae used as propositions, on which we can use the propositional connectives $\{\neg, \wedge, \vee, \to, \leftrightarrow\}$.

And here is our notion of truth.

Definition 4.25 (Model)**.** A model $\mathcal{M} = (D, I)$ appends truth to syllogistic propositional formulae by behaving with connectives as follows.

- $\mathcal{M} \vDash \varphi \wedge \psi$ if and only if $\mathcal{M} \vDash \varphi$ and $\mathcal{M} \vDash \psi$.
- $\mathcal{M} \vDash \varphi \vee \psi$ if and only if $\mathcal{M} \vDash \varphi$ or $\mathcal{M} \vDash \psi$.
- $\mathcal{M} \vDash \neg\varphi$ if and only if $\mathcal{M} \nvDash \varphi$.
- $\mathcal{M} \vDash \varphi \to \psi$ if and only if $\mathcal{M} \vDash \varphi$ implies $\mathcal{M} \vDash \psi$.
- $\mathcal{M} \vDash \varphi \leftrightarrow \psi$ if and only if $\mathcal{M} \vDash \varphi$ exactly when $\mathcal{M} \vDash \psi$.

These notions of models help us introduce validity and so on, in the same way.

4.2.2 Monadic Predicate Logic

Today, we are going to move from our syllogistic notation

$$\text{All } A \text{ are } B$$

to

$$\forall x(A(x) \to B(x)).$$

The point is that we can leverage what we know about the conditional $\to$ so that we merely have to add in the semantics for a new symbol $\forall$ in order to include propositional logic to are system.

Similarly,

$$\text{No } A \text{ is } B$$

becomes

$$\neg\exists x(A(x) \wedge B(x)).$$

Let's be more rigorous now. Let's introduce our formulae.

Definition 4.26 (Mondadic predicate formula)**.** Fix $\text{Pred} \coloneqq \{P_1, P_2, \ldots\}$ a set of unary predicates and some variables $\text{Var} \coloneqq \{x_1, x_2, \ldots\}$. Then our formulae are created as follows.

- If $P \in \text{Pred}$ and $x \in \text{Var}$, then $P(x)$ is a formula.
- If φ and ψ are formulae and $x \in \text{Var}$, then $\neg\varphi$ and $\varphi \wedge \psi$ and $\varphi \vee \psi$ and $\varphi \to \psi$ and $\varphi \leftrightarrow \psi$ and $\forall x\varphi$ and $\exists x\varphi$ are all formulae.
- There are no other formulae.

Remark 4.27. Our formulae are called monadic because we are only using unary predicates.

Remark 4.28. We can rigorize the notion that there are no other formulae in the typical "inductive" way by intersecting all subsets of expressions which satisfy the above.

Example 4.29. The expression $\neg\exists x P(x)$ is a formula.

Variables can come in two types.

Example 4.30. In the formula $\forall x P(x) \to P(y)$, the variable x in $P(x)$ is bound to $\forall x$. The $P(y)$ thus has a free variable y.

Example 4.31. In the formula $\forall x P(x) \to P(x)$, the first $P(x)$ has a bound x, but the second $P(x)$ has a free x because the $\forall x$ does not have "scope" beyond the main connective $\to$. Notably, if we wanted to bound the second $P(x)$, we would have written $\forall x(P(x) \to P(x))$.

Here is our definition.

Definition 4.32 (Bound, free)**.** A variable $x \in \mathrm{Var}$ in a formula φ is *bound* if it belongs to a quantifier $\forall$ or $\exists$. Otherwise, the variable x is *free*.

A more rigorous definition can be given by construction sequences, for example. The point is that we need to check the scope of every quantifier.

Free and bounded variables give us the following definition.

Definition 4.33 (Open, closed formulae)**.** A formula φ is *open* if and only if some variable in φ is free. Otherwise, φ is *closed* or a *sentence*.

Example 4.34. The formulae $\forall x P(x) \to P(y)$ and $\forall x P(x) \to P(y)$ are open formulae.

Example 4.35. The formulae $\forall x(P(x) \to P(x))$ and

$$\exists x(P(x) \to \forall y Q(y))$$

are both closed.

The point is that open formulae need to be told what their free variables are before we can determine their truth value. However, closed formulae have all their variables appropriately bound, so the truth value will not require setting variables.

4.2.3 Models for Monadic Predicate Logic

Models for monadic predicate logic are exactly the same as for syllogistic logic. Here is our definition, again.

Definition 4.36 (Model)**.** A *model* $\mathcal{M} = (D, I)$ for our pure monadic language consists of a nonempty set D and a function $I : \mathrm{Pred} \to \mathcal{P}(D)$ that sends each predicate A to a subset $I(A) \subseteq D$. Then $I(A)$ is the *extension* or *interpretation* of A in $\mathcal{M}$.

However, for free variables, we will also want access to variable assignments.

Definition 4.37 (Variable assignment)**.** Given a model $\mathcal{M} = (D, I)$ a *variable assignment* is a function $g : \mathrm{Var} \to D$. Intuitively, we are setting our variables equal to objects in our set/domain D.

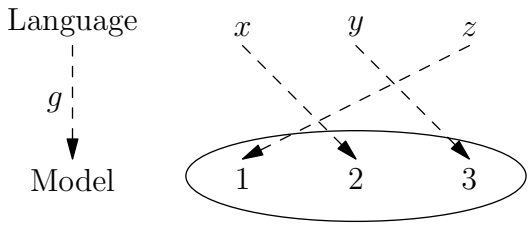
Remark 4.38. The variable assignment g does not have to be neither surjective nor injective.

Remark 4.39. The point of a variable assignment is that we cannot determine the truth of

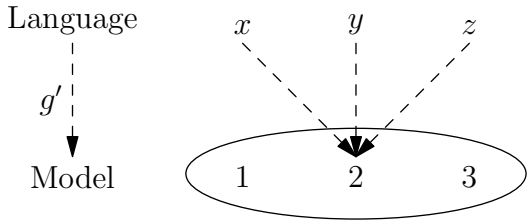
$$P(x)$$

on its own, without knowing what x is.

Here is the image for some assignment.



Of course, there are lots of ways to do this assignment. Here is a pathological one.



There are also ways to modify a variable assignment.

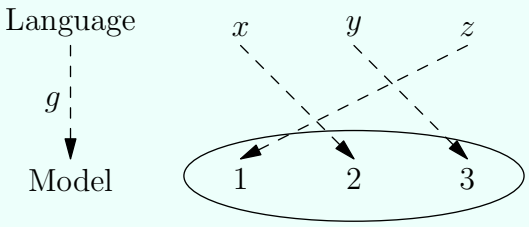
Definition 4.40 (Variable assignment modification)**.** Fix a model $\mathcal{M} = (D, I)$ and a variable assignment $g : \mathrm{Var} \to D$. Then we set the *modification* $g[x \coloneqq d] : \mathrm{Var} \to D$ by

$$g[x \coloneqq d](y) = \begin{cases} g(y) & y \neq x, \\ d & y = x. \end{cases}$$

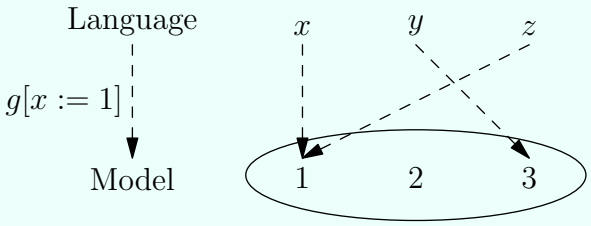
Intuitively, we set $g[x \coloneqq d]$ to be g, except for the requirement that we definitely send x to d.

Remark 4.41. We can stack these and write something like $g[x_1 = d_1][x_2 = d_2]$ to force more than one.

Example 4.42. The variable assignment



becomes

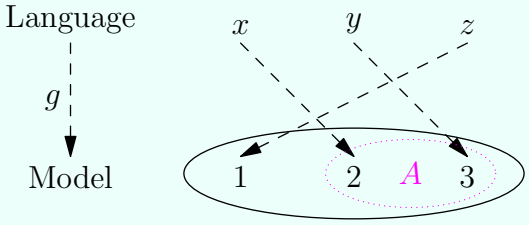


With variable assignments, here is our notion of semantics, without quantifiers. The point is that we only have to introduce what do with the variables themselves.

Definition 4.43 (Truth)**.** Fix a model $\mathcal{M} = (D, I)$ and a variable assignment g. Then we define the truth value of a formula φ (i.e., φ is *true in the model* $\mathcal{M}$ *under* g) recursively, as follows.

- For atomic formulae, $\mathcal{M} \vDash_g P(x)$ if and only if $g(x) \in I(P)$.
- We build truth for connectives in the same way as for propositional logic.

Example 4.44. Consider the following variable assignment.



Here, $\mathcal{M} \vDash_g A(y)$ and $\mathcal{M} \vDash_g A(z)$, but $\mathcal{M} \nvDash_g A(x)$.

We now continue Definition 4.43 to add in quantifiers.

Definition 4.45 (Truth)**.** Fix a model $\mathcal{M} = (D, I)$ and a variable assignment g. Then we define the truth value of a formula φ (i.e., φ is *true in the model* $\mathcal{M}$ *under* g) recursively, as follows.

- For atomic formulae, $\mathcal{M} \vDash_g P(x)$ if and only if $g(x) \in I(P)$.
- We build truth for connectives in the same way as for propositional logic.
- $\mathcal{M} \vDash_g \forall x \varphi$ if and only if, for all $d \in D$, we have $\mathcal{M} \vDash_{g[x:=d]} \varphi$.
- $\mathcal{M} \vDash_g \exists x \varphi$ if and only if there is some $d \in D$ for which $\mathcal{M} \vDash_{g[x:=d]} \varphi$.

Example 4.46. Fix $D := \{1, 2, 3\}$ and $\mathcal{M} := (D, I)$ some model. Then $\mathcal{M} \vDash_g (P(x) \vee Q(x))$ if and only if all the following hold.

(i) $\mathcal{M} \vDash_{g[x:=1]} P(x) \vee Q(x)$, which is equivalent to $1 \in I(P)$ or $1 \in I(Q)$.

(ii) $\mathcal{M} \vDash_{g[x:=2]} P(x) \vee Q(x)$, which is equivalent to $2 \in I(P)$ or $2 \in I(Q)$.

(iii) $\mathcal{M} \vDash_{g[x:=3]} P(x) \vee Q(x)$, which is equivalent to $3 \in I(P)$ or $3 \in I(Q)$.

Example 4.47. In Example 4.13, the statement

$$\mathcal{M} \vDash_g \forall x(\text{Student}(x) \vee \text{Faculty}(x))$$

is true basically by eye-balling it.

4.3 March 28

We continue our discussion of monadic logic. There is no homework this week due to the midterm.

4.3.1 The Universal Quantifier

To review, recall the following example for our monadic logic.

Example 4.47. In Example 4.13, the statement

$$\mathcal{M} \vDash_g \forall x(\text{Student}(x) \vee \text{Faculty}(x))$$

is true basically by eye-balling it.

In particular, we had the following notion of modeling for $\forall$.

Definition 4.48 (Truth for $\forall$)**.** Fix a model $\mathcal{M} = (D, I)$ and variable assignment $g : \text{Var} \to D$. Then, for some formula φ, then $\mathcal{M} \vDash_g \forall x \varphi x$ if and only if, for all variable assignments g of the open variables in φ and all $a \in D$, we have

$$\mathcal{M} \vDash_{g[x:=a]} \varphi.$$

Example 4.49. In the context of Example 4.13, asking if "x is a student" is true does not make sense while φ has open variables. For example,

$$\mathcal{M} \vDash_g \text{Student}(x) \vee \text{Faculty}(x)$$

because we can check each $x \in \{1, 2, 3, 4\}$ by hand.

- We see $\mathcal{M} \vDash_{g[x:=1]} \text{student}(x) \vee \text{Faculty}(x)$ because $\text{Student}(1)$ is true.
- We see $\mathcal{M} \vDash_{g[x:=2]} \text{student}(x) \vee \text{Faculty}(x)$ because $\text{Student}(2)$ is true.
- We see $\mathcal{M} \vDash_{g[x:=3]} \text{student}(x) \vee \text{Faculty}(x)$ because $\text{Student}(3)$ is true.
- We see $\mathcal{M} \vDash_{g[x:=4]} \text{student}(x) \vee \text{Faculty}(x)$ because $\text{Faculty}(4)$ is true.

Remark 4.50. A claim like "All A are B" can be written down as

$$\forall x(A(x) \to B(x)).$$

Namely, each x such that $A(x)$ must also have $B(x)$.

4.3.2 The Existential Quantifier

And here is what we do for $\exists$.

Definition 4.51 (Truth for $\exists$)**.** Fix a model $\mathcal{M} = (D, I)$ and variable assignment $g : \mathrm{Var} \to D$. Then, for some formula φ, then $\mathcal{M} \vDash_g \forall x \varphi x$ if and only if, for all variable assignments g of the open variables in φ and some $a \in D$, we have

$$\mathcal{M} \vDash_{g[x:=a]} \varphi.$$

Then a is a *witness* to this statement.

Remark 4.52. It is possible to have more than one witness.

The only difference is that we require this to be true for some $a \in D$, which is notably different from requiring all $a \in D$.

Example 4.53. We have $\mathcal{M} \vDash_g \exists(P(x) \wedge \neg Q(x))$ if and only if there is some $d \in D$ such that $\mathcal{M} \vDash_{g[x:=d]} P(x) \wedge \neg Q(x)$.

Concretely, if $D = \{1, 2, 3\}$, then $\mathcal{M} \vDash_g \exists(P(x) \wedge \neg Q(x))$ holds if and only if any one of the following is true.

(i) $\mathcal{M} \vDash_{g[x:=1]} P(x) \wedge \neg Q(x)$, which means $P(1)$ and not $Q(1)$.

(ii) $\mathcal{M} \vDash_{g[x:=2]} P(x) \wedge \neg Q(x)$, which means $P(2)$ and not $Q(2)$.

(iii) $\mathcal{M} \vDash_{g[x:=3]} P(x) \wedge \neg Q(x)$, which means $P(3)$ and not $Q(3)$.

Example 4.54. Working in Example 4.13, we have

$$\mathcal{M} \vDash_g \exists x(\mathrm{Student}(x) \wedge \neg\, \mathrm{Sophomore}(x)).$$

Indeed,

$$\mathcal{M} \vDash (\mathrm{Student}(2) \wedge \neg\, \mathrm{Sophomore}(2)).$$

Namely, 2 is a student and not a sophomore.

Remark 4.55. Intuitively, we can think about $\forall$ as some giant $\wedge$ over all variables. Dually, we can think about $\exists$ as some giant $\vee$ over all variables. In particular, if our model D has finitely many elements, then $\forall x P(x)$

$$\bigwedge_{d \in D} P(d).$$

What are we missing? Well, we would like properties that depend on more than one variable, and we would also like a notion of equality. These will be a feature of first-order logic that we do not have in pure monadic logic.

4.3.3 Validity

Note that, if φ is a closed formula (i.e., has no free variables), then φ is true independent of the chosen variable assignment g. In other words, if we have a model $\mathcal{M}$ and two variable assignments g and g', then

$$\mathcal{M} \vDash_g \varphi \iff \mathcal{M} \vDash_{g'} \psi.$$

Simply speaking, whatever g does doesn't matter because we are going to set everything in a $\forall$ or $\exists$ modification.

In fact, we have the following.

Proposition 4.56. Fix a model $\mathcal{M}$ and formula φ. Then if g and g' agree on all free variables in φ, then we still have

$$\mathcal{M} \vDash_g \varphi \iff \mathcal{M} \vDash_{g'} \varphi.$$

Proof. Once again, the reasoning is that we can determine the truth of φ as soon as we know what happens to its free variables. ■

As with propositional logic, we have a notion of validity.

Definition 4.57 (Valid)**.** An argument with a set Γ of premises and conclusion ψ is *valid* if and only if every model $\mathcal{M}$ and variable assignment g such that $\mathcal{M} \vDash_g \varphi$ for each $\varphi \in \Gamma$, we have

$$\mathcal{M} \vDash_g \psi$$

as well. We notate this $\Gamma \vDash \psi$ and say that ψ is a semantic consequence of Γ.

As before, we will still say that a formula φ is valid if and only if $\varnothing \vDash \varphi$.

Example 4.58. The following argument is valid.

1. $\forall x(A(x) \wedge B(x))$.
2. Therefore, $\forall x A(x)$.

For more examples, the following statements are valid. We will not prove them rigorously, though the reader might want to be convinced that they could if they wanted to.

- Quantifier exchange: $\forall x P(x) \leftrightarrow \neg\exists x \neg P(x)$ and $\exists P(x) \leftrightarrow \neg\forall x \neg P(x)$.
  For example, someone had a high-five if and only if nobody did not have a high-five.
- Universal instantiation: $\forall P(x) \to P(y)$. Namely, if $P(x)$ is always true, then surely $P(y)$ is true.
- Existential generalization: $P(y) \to \exists x P(x)$. Namely, if we have found an apple, then we know that there exists an apple.
- Distribution: $\forall x(P(x) \to Q(x)) \to (\forall x P(x) \to \forall x Q(x))$. In words, if we know that everyone with P has Q, and we know that everybody has P, then we also know that everyone has Q.
- Vacuous quantification: $P(x) \to \forall y P(x)$. Namely, the $\forall y$ does not help us evaluate $P(x)$.

Remark 4.59. There is an algorithm that will decide in finite time if a formula in pure monadic logic is valid. The main key to the algorithm is to run through all possible "versions" of interpretation instead of working through all models and variable assignments, of which there are infinitely many. The reason we are able to do this is that our formula only has finitely many predicates to care about.

To manifest the above remark, we state the following lemma.

Lemma 4.60. If a formula φ is not valid, then there is a model $\mathcal{M} = (D, I)$ with $\#D = 2^k$, where k is the number of predicate symbols appearing in φ.

Example 4.61. The formula

$$\varphi \coloneqq \forall x A(x) \to \forall x (A(x) \wedge B(x))$$

can be falsified by a model with domain $\{1, 2, 3, 4\}$. Namely, we set $\mathcal{I}$ by defining I as $I(A) = \{1, 2\}$ and $I(B) = \{1, 3\}$ so that $A(2) \to (A(2) \wedge B(2))$ is false, thus falsifying φ.

Corollary 4.62. There is an algorithm for deciding if a formula in pure monadic logic

Proof. Fix φ a formula with k predicates. In the hard direction, suppose that φ is not valid. Then by the lemma, we know that there is a model $\mathcal{M} = (D, I)$ where $\#D = 2^k$. Now, I is a function $\mathrm{Pred} \to \mathcal{P}(D)$, of which there are

$$(\#\mathcal{P}(D))^{\#\,\mathrm{Pred}} = \left(2^k\right)^k = 2^{k^2}$$

total models to check, which gives our algorithm. Conversely, if φ is true, then the above process will not find any model falsifying φ, so we still return true. ■

This algorithm might be slow, but at least it works.

4.4 March 30

We continue moving towards first-order logic.

4.4.1 Decidability

Last time we introduced validity for pure monadic first-order logic and gave a rough sketch for why there is a finite-time algorithm to determine if a formula in pure monadic first-order logic is valid.

Remark 4.63. Roughly speaking, the reason why we can work with only finite models is that two objects with the exact same properties (up to interpretation in the predicates in φ) can be identified. Then, because φ has only k predicates, we only have 2^k ways to assign these properties, so this is the number of elements we need for the domain of our model.

However, it will turn out that there is no algorithm to determine if an algorithm in first-order logic is valid or not. This is the price of a powerful language.

4.4.2 Constants

We will now add "names" to our language which cannot change on variable assignments. Quickly, recall that we defined our language of pure monadic logic by inducting upwards from connective, predicates, and variables.

We now define a different kind of "term," which will be constants.

Example 4.64. In the sentence "For any integer n," the letter "n" is a variable.

Example 4.65. We will never say "For any integer 0," the meaning of 0 is fixed, so it is a constant.

As such, we will add in constant symbols

$$\mathrm{Const} = \{c_1, c_2, \ldots\}$$

which is disjoint from our sets of variables Var and of predicates Pred. Otherwise, constants behave like variables.

Definition 4.66 (Mondadic predicate formula)**.** Fix $\mathrm{Pred} \coloneqq \{P_1, P_2, \ldots\}$ a set of unary predicates and some variables $\mathrm{Var} \coloneqq \{x_1, x_2, \ldots\}$ and some constants $\mathrm{Const} \coloneqq \{c_1, c_2, \ldots\}$. Then our formulae are created as follows.

- If $P \in \mathrm{Pred}$ and $x \in \mathrm{Var}$, then $P(x)$ is a formula.
- Similarly, if $P \in \mathrm{Pred}$ and $c \in \mathrm{Const}$, then $P(c)$ is a formula.
- If φ and ψ are formulae and $x \in \mathrm{Var}$, then $\neg\varphi$ and $\varphi \wedge \psi$ and $\varphi \vee \psi$ and $\varphi \to \psi$ and $\varphi \leftrightarrow \psi$ and $\forall x \varphi$ and $\exists x \varphi$ are all formulae.
- There are no other formulae.

We also have to update our models.

Definition 4.67 (Model)**.** A *model* $\mathcal{M} = (D, I)$ for pure monadic logic consists of a nonempty set D and a function $I : \mathrm{Pred} \to \mathcal{P}(D)$ that has the following data.

- Each predicate A to a subset $I(A) \subseteq D$.
- We also have to assign $I(c) \in D$ for each $c \in \mathrm{Const}$.

Then for validity, we also have to add in our constants.

Definition 4.68 (Truth)**.** Fix a model $\mathcal{M} = (D, I)$ and a variable assignment g. Then we define the truth value of a formula φ (i.e., φ is *true in the model* $\mathcal{M}$ *under* g) recursively, as follows.

- For atomic formulae, $\mathcal{M} \vDash_g P(x)$ for $x \in \mathrm{Var}$ if and only if $g(x) \in I(P)$.
- Similarly, $\mathcal{M} \vDash_g P(x)$ for $x \in \mathrm{Const}$ if and only if $I(c) \in I(P)$.
- We build truth for connectives in the same way as for propositional logic.
- $\mathcal{M} \vDash_g \forall x \varphi$ if and only if, for all $d \in D$, we have $\mathcal{M} \vDash_{g[x:=d]} \varphi$.
- $\mathcal{M} \vDash_g \exists x \varphi$ if and only if there is some $d \in D$ for which $\mathcal{M} \vDash_{g[x:=d]} \varphi$.

Example 4.69. To the model in Example 4.13, we add a constant kate, and we set $I(\mathrm{kate}) = 4$. As such, we see that $\mathcal{M} \vDash_g \mathrm{Faculty}(\mathrm{kate})$ because

$$I(\mathrm{kate}) = 4 \in I(\mathrm{Faculty}) = \{4\}.$$

Similarly, we see that $\mathcal{M} \nvDash_g \mathrm{Student}(\mathrm{kate})$ because

$$I(\mathrm{kate}) = 4 \notin I(\mathrm{Student}) = \{1, 2, 3\}.$$

Note that constants let us turn quantifiers into propositional logic: fixing $D = \{1, 2, \ldots, n\}$, and suppose that our model has constants $\{c_1, \ldots, c_n\}$ which our model sends $I(c_k) = k$. Then

$$\begin{aligned} \mathcal{M} \vDash_g \forall x P(x) &\iff \mathcal{M} \vDash_g P(c_1) \wedge \cdots \wedge P(c_n) \\ \mathcal{M} \vDash_g \exists x P(x) &\iff \mathcal{M} \vDash_g P(c_1) \vee \cdots \vee P(c_n). \end{aligned}$$

This idea works as long as the domain D is finite; otherwise, we will need an infinite formula to correctly simulate.

4.4.3 Function Symbols

We will now introduce functions. The important part of a function is that we have unique outputs.

Example 4.70. There is a function biomom taking a person to their biological mother. This is a unary function: it takes one input. In fact,

$$\mathrm{biomom}(\mathrm{biomom}(\mathrm{kate}))$$

is a notion which makes sense.

Remark 4.71. Our predicates will be upper-case, and our functions will be lower-case.

Now, with functions, we need to be careful about what we call a term. We define it inductively.

Definition 4.72 (Term)**.** Fix a language with variables in Var and predicates Pred and constants Const and functions Func. We now define a *term* as follows.

- Variables are terms.
- Constants are terms.
- If $f \in \mathrm{Func}$ and t is a term, then $f(t)$ is a term.
- Nothing else is a term.

Example 4.73. If biomom is a function and $x \in \mathrm{Var}$, then $\mathrm{biomom}(x)$ is a term, so $\mathrm{biomom}(\mathrm{biomom}(x))$ is also a term.

We can now appropriately extend Definition 4.66 to talk about formulae as follows.

Definition 4.74 (Mondadic predicate formula)**.** Fix $\mathrm{Pred} \coloneqq \{P_1, P_2, \ldots\}$ a set of unary predicates and some variables $\mathrm{Var} \coloneqq \{x_1, x_2, \ldots\}$ and some constants $\mathrm{Const} \coloneqq \{c_1, c_2, \ldots\}$. Then our formulae are created as follows.

- If $P \in \mathrm{Pred}$ and $x \in \mathrm{Term}$, then $P(x)$ is a formula.
- If φ and ψ are formulae and $x \in \mathrm{Var}$, then $\neg\varphi$ and $\varphi \wedge \psi$ and $\varphi \vee \psi$ and $\varphi \to \psi$ and $\varphi \leftrightarrow \psi$ and $\forall x\varphi$ and $\exists x\varphi$ are all formulae.
- There are no other formulae.

Example 4.75. If biomom is a function and $\mathrm{kate} \in \mathrm{Const}$, the $\mathrm{biomom}(\mathrm{biomom}(x))$ is a term, so

$$\mathrm{LeftHanded}(\mathrm{biomom}(\mathrm{biomom}(\mathrm{kate}))$$

might be some formula.

We now need to give functions some semantics. So we update our definition of model.

Definition 4.76 (Model)**.** A *model* $\mathcal{M} = (D, I)$ for pure monadic logic consists of a nonempty set D and a function $I : \mathrm{Pred} \to \mathcal{P}(D)$ that has the following data.

- Each predicate A to a subset $I(A) \subseteq D$.
- We also have to assign $I(c) \in D$ for each $c \in \mathrm{Const}$.
- A unary function $f \in \mathrm{Func}$ gets sent to a unary function $I(f) : D \to D$.

We won't define precisely what it means to be a function, but it should be intuitive.

Example 4.77. In a model $\mathcal{M} = (D, I)$ where $D = \mathbb{N}$, we can define $I(\mathrm{biomom})$ to be the successor function $n \mapsto n + 1$. For fun, we will also set $I(\mathrm{kate}) = 0$, and add a predicate

$$I(\mathrm{LeftHanded}) = \{3, 4\}.$$

To keep track, here is everything that we've introduced.

- Predicates Pred belong to the language.
- Constants Const belong to the language.
- Functions Func belong to the language.
- More generally, terms Term belong to the language.
- However, the interpretation of a model tells us what to do with all this data.

To keep track of this, here is some notation.

Definition 4.78 (Denotation)**.** Fix a model $\mathcal{M} = (D, I)$. Then, given a variable assignment g, we set

$$[\![t]\!]_I^g \in D$$

to be the term t relative to the interpretation t and variable assignment g. This is defined inductively as follows.

- For $x \in \mathrm{Var}$, we set $[\![x]\!]_I^g \coloneqq g(x)$.
- For $c \in \mathrm{Const}$, we set $[\![c]\!]_I^g \coloneqq I(c)$.
- Lastly, if $f \in \mathrm{Func}$ and $t \in \mathrm{Term}$, we set $[\![f(t)]\!]_I^g$ to be the interpreted function $I(f)$ applied to the term $[\![t]\!]_I^g$. I.e.,
$$[\![f(t)]\!]_I^g \coloneqq I(f)([\![t]\!]_I^g).$$

Example 4.79. From Example 4.77, we can compute

$$[\![\mathrm{biomom}(\mathrm{kate})]\!]_I^g \coloneqq I(\mathrm{biomom})([\![\mathrm{kate}]\!]_I^g) = I(\mathrm{biomom})(I(\mathrm{kate})) = I(\mathrm{biomom})(0) = 1.$$

And now we update our semantics, which generalizes.

Definition 4.80 (Truth)**.** Fix a model $\mathcal{M} = (D, I)$ and a variable assignment g. Then we define the truth value of a formula φ (i.e., φ is *true in the model* $\mathcal{M}$ *under* g) recursively, as follows.

- For atomic formulae, $\mathcal{M} \vDash_g P(x)$ for $x \in \mathrm{Var}$ if and only if $[\![t]\!]_I^g \in I(P)$.
- We build truth for connectives in the same way as for propositional logic.
- $\mathcal{M} \vDash_g \forall x \varphi$ if and only if, for all $d \in D$, we have $\mathcal{M} \vDash_{g[x:=d]} \varphi$.
- $\mathcal{M} \vDash_g \exists x \varphi$ if and only if there is some $d \in D$ for which $\mathcal{M} \vDash_{g[x:=d]} \varphi$.

Example 4.81. From Example 4.77, we see that

$$\mathcal{M} \nvDash_g \mathrm{LeftHanded}(\mathrm{biomom}(\mathrm{kate}))$$

because $1 \notin I(\mathrm{LeftHanded})$.

4.5 April 1

We continue moving towards first-order logic.

4.5.1 Validity

Continuing our story with function symbols, we still have the same notion of validity.

Definition 4.82 (Valid)**.** A formula φ is *valid* if and only if, for all models $\mathcal{M} = (D, I)$, we have that $\mathcal{M} \vDash_g \varphi$ for all variable assignments in D.

It turns out that, even with our unary function symbols, it is possible to give an algorithm for the truth. Here is the key result.

Theorem 4.83. Given a formula φ with only unary function symbols, there exists a formula ψ which is valid if and only if φ is valid, but now ψ has no function symbols at all.

Thus, we can reduce validity to the pure monadic case without function symbols.

4.5.2 Identity

We are interested in talking about identity in a rigorous way; namely, we want a notion of two objects being (literally) the same. To start, we need to introduce the predicate for equality.

Definition 4.84 (Mondadic predicate formula)**.** Fix $\mathrm{Pred} := \{P_1, P_2, \ldots\}$ a set of unary predicates and some variables $\mathrm{Var} := \{x_1, x_2, \ldots\}$ and some constants $\mathrm{Const} := \{c_1, c_2, \ldots\}$. Then our formulae are created as follows.

- If $s, t \in \mathrm{Term}$, then $s \doteq t$ is a formula.
- If $P \in \mathrm{Pred}$ and $x \in \mathrm{Term}$, then $P(x)$ is a formula.
- If φ and ψ are formulae and $x \in \mathrm{Var}$, then $\neg\varphi$ and $\varphi \wedge \psi$ and $\varphi \vee \psi$ and $\varphi \to \psi$ and $\varphi \leftrightarrow \psi$ and $\forall x \varphi$ and $\exists x \varphi$ are all formulae.
- There are no other formulae.

Note that we are writing $\doteq$ to distinguish the language's identity with our identity outside.

Notation 4.85. We may write $\neg(s \doteq t)$ for $s \neq t$.

And here are our semantics.

Definition 4.86 (Truth)**.** Fix a model $\mathcal{M} = (D, I)$ and a variable assignment g. Then we define the truth value of a formula φ (i.e., φ is *true in the model $\mathcal{M}$ under g*) recursively, as follows.

- For atomic formulae $P(x)$, $\mathcal{M} \vDash_g P(t)$ for $t \in \text{Term}$ if and only if $[\![t]\!]_I^g \in I(P)$.
- For atomic formulae $s \doteq t$, we have $\mathcal{M} \vDash_g s \doteq t$ if and only if $[\![s]\!]_I^g = [\![t]\!]_I^g$.
- We build truth for connectives in the same way as for propositional logic.
- $\mathcal{M} \vDash_g \forall x \varphi$ if and only if, for all $d \in D$, we have $\mathcal{M} \vDash_{g[x:=d]} \varphi$.
- $\mathcal{M} \vDash_g \exists x \varphi$ if and only if there is some $d \in D$ for which $\mathcal{M} \vDash_{g[x:=d]} \varphi$.

Note that we are not getting free relation about how the predicate $\doteq$ is interpreted by our model's I function.

Example 4.87. Continuing from Example 4.77, we set a constant $\text{bev} \in \text{Const}$ so that $I(\text{bev}) = 1$. Then we can compute

$$\mathcal{M} \vDash_g \text{bev} \doteq \text{biomom}(\text{kate}).$$

Indeed, $[\![\text{bev}]\!]_I^g = I(\text{bev}) = 1$ and $[\![\text{biomom}(\text{kate})]\!] = I(\text{biomom})(I(\text{kate})) = 1$ as well.

As an aside, we are able to talk about distinctness with our notion of identity, so we can count the number of objects in our model as well.

Example 4.88. In a model with a one-object domain, we cannot have

$$\exists x_1 \exists x_2 (x_1 \neq x_2).$$

Similarly, we can write

$$\exists x_1 \cdots \exists x_n \bigwedge_{\substack{1 \le i,j \le n \\ i \neq j}} (x_i \neq x_j)$$

to say that there are at least n elements. Negating asserts that there at most $n-1$ objects, so we can actually pin down a finite number of objects. For example,

$$\neg\big(\exists x_1 \exists x_2 \exists x_3 (x_1 \neq x_2 \wedge x_1 \neq x_3 \wedge x_2 \neq x_3)\big) \wedge (\exists x_1 \exists x_2 (x_1 \neg \doteq x_2)$$

asserts that there are exactly 2 objects.

Example 4.89. If we want to classify objects with a property P, we can write something like

$$\exists x_1 \exists x_2 (P(x_1) \wedge P(x_2) \wedge x_1 \neq x_2).$$

As such, we can more or less completely describe a finite model by asserting the existence and number of objects with particular properties.

We take a moment to list some tautologies.

Proposition 4.90. The following are true, for any terms s and t.

- Reflexivity: $t \doteq t$.
- Symmetry: $s \doteq t \to t \doteq s$.
- Transitivity: $(a \doteq b) \wedge (b \doteq c) \to (a \doteq c)$.
- Substitution: $s \doteq t \to (P(s) \leftrightarrow P(t))$.

Proof. Plug into the definitions. For example, for a model $\mathcal{M} = (D, I)$, we must have

$$[\![t]\!]_I^g = [\![t]\!]_I^g,$$

which we could show by an induction if we so chose, but it is also just true in our metalanguage. The rest are similar. ■

4.5.3 Decidability, Again

Our notion of validity is the same as usual. It turns out that we still have validity, which causes the following result.

Lemma 4.91. If a formula φ (with no function symbols but perhaps with identity) can be falsified in a model, then φ can be falsified in a model with domain

$$\left\{1, 2, \ldots, r \cdot 2^k\right\},$$

where r is the number of variables and k is the number of predicates.

Remark 4.92. We need to depend on the number of variables and not just the number of predicates because our remarks on counting allow us to force the model to have an arbitrary number of elements without changing the number of predicates.

If we add in one unary function symbol, then we are still decidable, which is its own theorem. But it turns out that we have two unary function symbols, then there is no algorithm for deciding validity (!).

4.5.4 Substitution for Terms

We have a little bookkeeping to do before we can continue. We showed earlier that

$$\forall x P(x) \to P(y)$$

is valid, for any variable x and term y. However, this only holds for our unary predicates P, but we want something like this to hold for arbitrary formulae φ. For example, we want to be able to say

$$\forall (P(x) \wedge Q(x)) \to (P(y) \wedge Q(y)).$$

As of now, we don't have an intelligent way to talk about these formulae with multiple variables and perhaps lots of complexity: we need a notion of substituting a term into a formula.

Example 4.93. Working in Example 4.77, we note that we can start with the term

$$\mathrm{biomom}(x)$$

and substitute the term $\mathrm{biomom}(\mathrm{kate})$ for the variable x to get

$$\mathrm{biomom}(\mathrm{biomom}(\mathrm{kate})).$$

We now codify this idea into a definition.

Definition 4.94 (Substitution, terms)**.** Fix terms s and t and a variable x. Then we define the *substitution* function $(-)_t^x : \mathrm{Term} \to \mathrm{Term}$ inductively, as follows.

- Do nothing on constants: if $c \in \mathrm{Const}$, then $c_t^x \coloneqq c$.
- Do nothing on wrong variables: if $y \in \mathrm{Var}$ and $y \neq x$, then $y_t^x = y$.
- Substitute: we take $x_t^x = t$, which is our substitution.
- Recuse: if $f \in \mathrm{Func}$ and $u \in \mathrm{Term}$, then $f(u)_t^x = f(u_t^x)$.

Intuitively, we are modifying our term s to replace xs with ts.

Theme 5
Syntax and Semantics

5.1 April 4

Welcome back everyone.

5.1.1 Substitution for Formulae

We return to the notion of substitution. Here is our definition.

Definition 4.94 (Substitution, terms)**.** Fix terms s and t and a variable x. Then we define the *substitution* function $(-)_t^x : \mathrm{Term} \to \mathrm{Term}$ inductively, as follows.

- Do nothing on constants: if $c \in \mathrm{Const}$, then $c_t^x \coloneqq c$.
- Do nothing on wrong variables: if $y \in \mathrm{Var}$ and $y \neq x$, then $y_t^x = y$.
- Substitute: we take $x_t^x = t$, which is our substitution.
- Recuse: if $f \in \mathrm{Func}$ and $u \in \mathrm{Term}$, then $f(u)_t^x = f(u_t^x)$.

Example 5.1. We compute

$$\mathrm{biomom}(x)^x_{\mathrm{biomom(kate)}} = \mathrm{biomom}\left(x^x_{\mathrm{biomom(kate)}}\right) = \mathrm{biomom}(\mathrm{biomom}(\mathrm{kate})).$$

We can actually extend substitution to formulae.

Example 5.2. In the formula

$$\mathrm{LeftHanded}(\mathrm{biomom}(x)),$$

we can still substitute $\mathrm{biomom}(\mathrm{kate})$ for x.

However, some care is required here: we should only substitute into free variables. For example, looking at

$$P(x) \wedge \exists x Q(x),$$

To substitute kate for x, we should not write $P(\mathrm{kate})\exists x Q(\mathrm{kate})$ or even worse $P(\mathrm{kate})\exists \mathrm{kate} Q(\mathrm{kate})$. The former is poorly behaved, and the latter is not even grammatical. To see that the former is not necessarily true, we note that the formula

$$\forall x(P(x) \wedge \exists x Q(x)) \to (P(\mathrm{kate}) \wedge \exists x Q(\mathrm{kate}))$$

need not be true, which is not what we want out of our substitution (i.e., "universal instantiation"). In particular, $\exists x Q(\text{kate})$ requires $Q(\text{kate})$.

So to be rigorous, here is our definition.

Definition 5.3 (Substitution, formulae)**.** Fix terms s and t and a variable x. Then we define the *substitution* function $(-)_t^x : \mathcal{L} \to \mathcal{L}$ inductively, as follows.

- Atomic formulae: $P(s)_t^x = P\left(s_t^x\right)$ for terms s and predicates P.
- Connectives: $(\neg\varphi)_t^x = \neg(\varphi_t^x)$ and $(\varphi \# \psi)_t^x = \varphi_t^x \# \psi_t^x$ for any $\# \in \{\wedge, \vee, \to, \leftrightarrow\}$.
- Quantifiers: if $y \in \text{Var} \setminus \{x\}$, then $(\forall y\varphi)_t^x = \forall y \varphi_t^x$ and $(\exists y\varphi)_t^x = \exists y \varphi_t^x$.
- Quantifiers: $(\forall x\varphi)_t^x = \forall x\varphi$ and $(\exists x\varphi)_t^x = \exists x\varphi$.

Note that we are not doing anything in the last case because the x has already been bound in a formula of the form $\forall x\varphi$ or $\exists x\varphi$.

Example 5.4. We compute

$$\begin{aligned}(\forall y P(x) \wedge \exists x Q(x))_{\text{kate}}^x &= (\forall y P(x))_{\text{kate}}^x \wedge (\exists x Q(x))_{\text{kate}}^x \\ &= \forall y P(x)_{\text{kate}}^x \wedge \exists x Q(x) \\ &= \forall y P(x_{\text{kate}}^x) \wedge \exists x Q(x) \\ &= \forall y P(\text{kate}) \wedge \exists x Q(x).\end{aligned}$$

5.1.2 Universal Instantiation

We are almost ready for universal instantiation. However, we note the following.

Example 5.5. Work in a model with more than two objects in the domain. Then we can write

$$\forall x \exists y (x \neq y)$$

by simply picking any x and then finding a distinct y. However, if we run $\exists y(x \neq y)$ and substitute y for x, then we might hope to deduce

$$\exists y (y \neq y).$$

The issue with the above example is that we are substituting in a variable already seen in the formula, which is creating problems. Namely, substituting in the variable y does not keep it free because it is bound within the formula.

There are a few ways to fix this. For one, we choose a variable outside the ones used in φ; we could also just substitute in constants for variables exclusively. To unify these, we have the notion of substitutable.

Definition 5.6 (Substitutable)**.** A term t is *substitutable* for x in φ if and only if no variable in t becomes bound in φ after substituting t for x.

As usual, one could give a recursive definition of this, but we will not bother.

Example 5.7. We cannot substitute $\text{biomom}(y)$ for x in the formula $\exists y(x \neq y)$ because y has already been bound.

And here, finally, here is our statement.

Theorem 5.8 (Universal instantiation)**.** Fix a formula φ, a variable x, and a term t so that t is substitutable for x in φ. Then

$$\forall x\varphi \to \varphi_t^x$$

is valid.

Intuitively, if φ is true for all x, then φ is true for t specifically.

Corollary 5.9. Fix terms s and t and a predicate P. Then, for a formula φ with a variable x such that s and t are substitutable for tx in φ, we have

$$(s \doteq t) \to (\varphi_s^x \leftrightarrow \varphi_t^x).$$

This wraps up our story on the basic parts of monadic predicate logic. We will now go beyond monadic predicate logic.

Example 5.10. In monadic predicate logic, there is no way to stay that "Kate is taller than Sue."

5.1.3 Predicates of Higher Arity

We close class by introducing predicates with more inputs, which will dramatically increase our expressive power.

Example 5.11. Here are some predicates.

- We can say a loves b with $\mathrm{Loves}(a, b)$.
- We can say a is greater than b with $a \dot{>} b$. Note we are writing $\dot{>}$ to distinguish the language's version of $>$ with English's version (i.e., the metalanguage's) version of $>$.
- We can say a, b, c are collinear with $\mathrm{Coll}(a, b, c)$.

Now here is our new definition of formulae: we just update the atomic formulae using predicates.

Definition 5.12 (Predicate formula)**.** Fix $\mathrm{Pred}^n := \{P_1^n, P_2^n, \ldots\}$ a set of n-ary predicates (for any positive integer n) and some variables $\mathrm{Var} := \{x_1, x_2, \ldots\}$ and some constants $\mathrm{Const} := \{c_1, c_2, \ldots\}$. Then our formulae are created as follows.

- If $s, t \in \mathrm{Term}$, then $s \doteq t$ is a formula.
- If $P \in \mathrm{Pred}$ and $(x_1, \ldots, x_n) \in \mathrm{Term}$, then $P(x_1, \ldots, x_n)$ is a formula.
- If φ and ψ are formulae and $x \in \mathrm{Var}$, then $\neg\varphi$ and $\varphi \wedge \psi$ and $\varphi \vee \psi$ and $\varphi \to \psi$ and $\varphi \leftrightarrow \psi$ and $\forall x\varphi$ and $\exists x\varphi$ are all formulae.
- There are no other formulae.

As usual, here is an example.

Example 5.13. Fix f a unary function and R a binary predicate. Then

$$\forall x \exists y R(x, f(y))$$

is a good formula. Namely, $f(y)$ is a term, and $R(x, f(y))$ is atomic, so then we can write $\exists y R(x, f(y))$ and then $\forall x \exists y R(x, f(y))$.

We can also update the semantics of our models in the usual way. The point is that, in a model $\mathcal{M} = (D, I)$, we can set $I(P^n)$ to be an n-ary predicate P^n to be some subset of D^n.

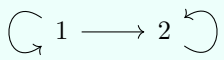
5.2 April 6

We continue our leap towards first-order logic.

5.2.1 Binary Relations

Namely, today we continue talking about the relations between two objects. Namely, we need to update what our models now look like because models need to keep track of what our higher-arity predicates behave.

Example 5.14. We could represent a relation $I(R)$ (where $R \in \text{Pred}^2$) in a model $\mathcal{M} = (\{1, 2\}, I)$ as follows.



Concretely, we might imagine $R(t_1, t_2)$ is the relation "t_1 follows t_2," so the above diagram asserts that "1 follows 2." In total, we can write down

$$I(R) = \{\langle 1, 1\rangle, \langle 1, 2\rangle, \langle 2, 2\rangle\}.$$

Remark 5.15. The order of the elements matters. Namely, $\langle 1, 2\rangle \in I(R)$ above does not force $\langle 2, 1\rangle \in I(R)$. Namely, $\langle 1, 2\rangle \neq \langle 2, 1\rangle$. We are not using parentheses here to avoid confusion with the parentheses used in the language itself. We are not using curly brace (i.e., $\{\}$) because sets do not care about order.

It takes some getting used to, but it is important to realize that we can gather all the data of a binary relation R on a set of objects D as simply stating the $\langle a, b\rangle \in D^2$ with $\langle a, b\rangle$ related by R. So we can think of $I(R)$ as the subset D^2.

Codifying, we have the following.

Definition 5.16 (Model)**.** A *model* $\mathcal{M} = (D, I)$ consists of a nonempty set D and a function I which restricts as $I : \text{Pred} \to \mathcal{P}(D)$ and $I : \text{Pred}^2 \to \mathcal{P}\left(D^2\right)$.

We can now also update our semantics.

Definition 5.17 (Truth)**.** Fix a model $\mathcal{M} = (D, I)$ and a variable assignment g. Then we define the truth value of a formula φ (i.e., φ is *true in the model* $\mathcal{M}$ *under* g) recursively, as follows.

- For atomic formulae $P(t)$, $\mathcal{M} \vDash_g P(t)$ for $t \in \text{Term}$ if and only if $[\![t]\!]_I^g \in I(P)$.
- For atomic formulae $P(t_1, t_2)$, $\mathcal{M} \vDash_g P(t_1, t_2)$ for $t_1, t_2 \in \text{Term}$ if and only if $\langle [\![t_1]\!]_I^g, [\![t_2]\!]_I^g\rangle \in I(P)$.
- Lastly, for atomic formulae $s \doteq t$, we have $\mathcal{M} \vDash_g s \doteq t$ if and only if $[\![s]\!]_I^g = [\![t]\!]_I^g$.
- We build truth for connectives in the same way as for propositional logic.
- $\mathcal{M} \vDash_g \forall x\varphi$ if and only if, for all $d \in D$, we have $\mathcal{M} \vDash_{g[x:=d]} \varphi$.
- $\mathcal{M} \vDash_g \exists x\varphi$ if and only if there is some $d \in D$ for which $\mathcal{M} \vDash_{g[x:=d]} \varphi$.

Example 5.18. Working in Example 5.14, we check

$$\mathcal{M} \vDash_{g[x:=1]} R(x, c),$$

where $I(c) = 2$. Well, we can compute

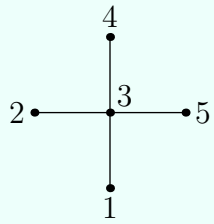
$$\left\langle [\![x]\!]_I^{g[x:=1]}, [\![c]\!]_I^{g[x:=1]} \right\rangle = \langle 1, 2\rangle \in R,$$

so we are done.

5.2.2 Predicates of Higher Arity

Continuing, we can also talk about ternary relations $B(a, b, c)$, which should be interpreted as a subset of D^3.

Example 5.19. Let $B(a, b, c)$ be the predicate that "b is between a and c." Well, consider the following model $\mathcal{M}$.



Then we can compute what $I(B)$ should be off of the diagram.

Then for semantics, to check $\mathcal{M}_g B(t_1, t_2, t_3)$ for a ternary relation B, we need to check

$$\langle [\![t_1]\!]_I^g, [\![t_2]\!]_I^g, , [\![t_3]\!]_I^g \rangle \in I(B).$$

As an example, we have the following.

Example 5.20. We work in the context of Example 5.19. We check

$$\mathcal{M} \nvDash_{g[x:=2,y:=3,z:=4]} B(x, y, z)$$

Well, we see that

$$\left\langle [\![x]\!]_I^{g[x:=2,y:=3,z:=4]}, [\![y]\!]_I^{g[x:=2,y:=3,z:=4]}, [\![z]\!]_I^{g[x:=2,y:=3,z:=4]} \right\rangle = \langle 2, 3, 4\rangle \notin I(B).$$

In the most general case, we have the following.

Definition 5.21 (Model)**.** A *model* $\mathcal{M} = (D, I)$ consists of a nonempty set D and a function I which consists of the following data.

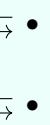
- For each n-ary predicate $P \in \mathrm{Pred}^n$, we have $I(P) \subseteq D^n$.
- For each $f \in \mathrm{Func}$, we have $I(f) : D \to D$.
- Lastly, for each $c \in \mathrm{Const}$, we have $I(c) \in D$.

And here are our semantics.

Definition 5.22 (Truth)**.** Fix a model $\mathcal{M} = (D, I)$ and a variable assignment g. Then we define the truth value of a formula φ (i.e., φ is *true in the model* $\mathcal{M}$ *under* g) recursively, as follows.

- For atomic formulae $P(t_1, \dots, t_n)$, $\mathcal{M} \vDash_g P(t)$ for $P \in \mathrm{Pred}^n$ and $t_1, \dots, t_n \in \mathrm{Term}$ if and only if
$$\langle [\![t_1]\!]_I^g, \dots, [\![t_n]\!]_I^g \rangle \in I(P).$$
- For atomic formulae $s \doteq t$, we have $\mathcal{M} \vDash_g s \doteq t$ if and only if $[\![s]\!]_I^g = [\![t]\!]_I^g$.
- We build truth for connectives in the same way as for propositional logic.
- $\mathcal{M} \vDash_g \forall x \varphi$ if and only if, for all $d \in D$, we have $\mathcal{M} \vDash_{g[x:=d]} \varphi$.
- $\mathcal{M} \vDash_g \exists x \varphi$ if and only if there is some $d \in D$ for which $\mathcal{M} \vDash_{g[x:=d]} \varphi$.

Example 5.23. In monadic predicate logic, the statement
$$\forall x \exists y \varphi \leftrightarrow \exists y \forall x \varphi$$
is valid. However, the statement
$$\forall x \exists y L(x, y) \to \exists y \forall x L(x, y)$$
is not valid. For example, letting $L(x, y)$ be "x loves y," then the above is saying that everyone loves someone should imply that someone is loved by everyone. But this does not make sense. To be rigorous, we would have to build a model where the statement is false. Here is one such.



Indeed, all elements love someone else, but there is no one element loved by everyone.

Remark 5.24. As a note on decidability of validity again, there is no algorithm to decide if a formula in "pure predicate logic" is valid. Namely, we can restrict our language to only have functions, identity, and a single predicate of arity at least 2, and then validity is not decidable. For the proof, take Philosophy 140B.

5.2.3 Functions of Higher Arity

The last part we need to get to predicate logic is to add in functions of higher arity.

Example 5.25. The following are functions.

- The function $(a, b) \mapsto a + b$ is binary.
- The function $(a, b) \mapsto a \times b$ is binary.
- The function $(a, b, m) \mapsto (a + b) \pmod m$ is ternary.

To start, we need to add in these functions to our language. For this, we just need to modify the definition of a term.

Definition 5.26 (Term)**.** Fix a language with variables in Var and n-ary predicates Pred^n (for any n) and constants Const and n-ary functions Func^n (for any n). We now define a *term* as follows.

- Variables are terms.
- Constants are terms.
- If $f \in \mathrm{Func}^n$ and $t_1, \dots, t_n$ are term, then $f(t_1, \dots, t_n)$ is a term.
- Nothing else is a term.

Example 5.27. Let $\dot{0}$ be a constant symbol and $\dot{s}$ be a unary function symbol and $\dot{+}$ a binary function symbol. Then

$$\dot{+}(\dot{+}(\dot{s}(\dot{0}), \dot{s}(\dot{0})), \dot{0})$$

is a term. In infix notation, we can write this as $(\dot{s}(\dot{0})\dot{+}\dot{s}(\dot{0}))\dot{+}\dot{0}$.

Remark 5.28. Notably, associativity is not automatic for these terms. In particular, $f(f(x, y), z)$ and $f(x, f(y, z))$ are different terms: the parentheses matter.

As such, we need to add in how to interpret an n-ary function in a model. It happens that this should be a function $D^n \to D$, which we will discuss more next class.

5.3 April 8

Welcome back everyone.

5.3.1 Semantics for Functions

Today we continue adding in our functions of higher arity. Namely, we have added these functions to our set of terms Term, which automatically add them to our syntax because formulae are built from terms. Thus, we have left to add functions of higher arity to our semantics. As such, we need to discuss how to interpret such a function.

Example 5.29. For a binary function $f \in \mathrm{Func}^2$, we would like to view this as a binary function $D \times D \to D$. For example, we should view $\dot{+}$ as a function which takes two (say) natural numbers $a, b \in \mathbb{N}$ and outputs $a + b \in \mathbb{N}$. The way to keep track of these inputs and outputs is by working with ordered pairs in $D \times D$.

Remark 5.30. Technically speaking, we will often think of a function $f : A \to B$ set-theoretically as a set of pairs (a, b) where $a \in A$ and $b \in B$, and each $a \in A$ has exactly one pair (a', b') with $a = a'$. We will not labor on this technical point, but it is correct.

For ternary functions $f \in \mathrm{Func}^3$, we should similarly interpret this as a function $D^3 \to D$. More generally, we have the following.

Definition 5.31 (Model)**.** A *model* $\mathcal{M} = (D, I)$ consists of a nonempty set D and a function I which consists of the following data.

- For each n-ary predicate $P \in \mathrm{Pred}^n$, we have $I(P) \subseteq D^n$.
- For each n-ary function $f \in \mathrm{Func}^n$, we have $I(f) : D^n \to D$.
- Lastly, for each $c \in \mathrm{Const}$, we have $I(c) \in D$.

As such, we need to update our denotation to accommodate these functions.

Definition 5.32 (Denotation)**.** Fix a model $\mathcal{M} = (D, I)$. Then, given a variable assignment g, we set

$$[\![t]\!]_I^g \in D$$

to be the term t relative to the interpretation t and variable assignment g. This is defined inductively as follows.

- For $x \in \mathrm{Var}$, we set $[\![x]\!]_I^g \coloneqq g(x)$.
- For $c \in \mathrm{Const}$, we set $[\![c]\!]_I^g \coloneqq I(c)$.
- Lastly, if $f \in \mathrm{Func}^n$ and $t_1, \dots, t_n \in \mathrm{Term}$, we set $[\![f(t_1, \dots, t_n)]\!]_I^g$ to be the interpreted function $I(f)$ applied to the terms $[\![t_\bullet]\!]_I^g$. I.e.,

$$[\![f(t_1, \dots, t_n)]\!]_I^g \coloneqq I(f)([\![t_1]\!]_I^g, \dots, [\![t_n]\!]_I^g).$$

Note that our semantics from Definition 5.22 follow directly because we defined these via terms.

Example 5.33. We work in a language with a constant symbol $\dot{0}$, a unary function symbol $\dot{s}$, and a binary function symbol $\dot{+}$. Our model $\mathcal{M} = (\mathbb{N}, I)$ is interpreted with

$$I(\dot{0}) = 0, \quad I(\dot{s})(n) = n + 1, \quad \text{and} \quad I(\dot{+})(a, b) = a + b.$$

From here, we can show that our model $\mathcal{M}$ satisfies the following.

- $\forall x (x \dot{+} \dot{0} \doteq x)$.
- $\forall x \forall (x \dot{+} \dot{s}(y) \doteq \dot{s}(x \dot{+} y))$.

It turns out that these can define addition, in a suitable way.

At a high level, the reason why we've added all these function symbols is that we want to express things beyond what we've been talking about with unary function symbols. For example, unary functions are not expressive enough to talk about addition.

Remark 5.34. Having a single binary function symbol and identity (even without any predicates!) is enough to make validity undecidable again.

5.3.2 Universal Elimination

We will now talk about how to prove things. In propositional logic, we could get by via truth tables. However, our models are more complicated now (namely, we cannot run through all of them in finite time), so we will be forced to use natural deduction if we want to show that something is valid.

Most of our rules are carried over directly from propositional logic (namely, we maintain all rules for our connectives), but we will need to talk about how to eliminate and introduce $\forall$ and $\exists$. Let's see an example.

Example 5.35. It is valid to take $\forall x P(x)$ and deduce $P(t)$ for any term $t \in \mathrm{Term}$.

Lemma 5.36. For all real numbers x and y, if $0 < x < y$, then $x^2 < y^2$.

Proposition 5.37. We have $9 < \pi^2$.

Proof. Note $0 < 3 < \pi$, so Lemma 5.36 forces $9 = 3^2 < \pi^2$, so we are done. ∎

The point is that we applied Lemma 5.36 to take a statement with the "for all" and apply it directly to 3 and π. In Fitch style, this looks like the following.

1	$\forall x \forall y \left(0 < x < y \to x^2 < y^2\right)$	
2	$0 < 3 < \pi \to x^2 < y^2$	$\forall$Elim, 1
3	$0 < 3 < \pi$	Math
4	$3^2 < \pi^2$	$\to$Elim, 2, 3
5	$3^2 \doteq 9$	Math
6	$9 < \pi^2$	$=$Elim, 4, 5

Note our use of the rules "$\forall$Elim" and "$=$Elim." These are the ones we are currently paying attention to.

Here is our $\forall$ elimination.

Definition 5.38 ($\forall$ Elimination)**.** If we have $\forall x \varphi$ such that the term $t \in \mathrm{Term}$ is substitutable for x in φ, then we can deduce φ_t^x on a new line in the same column, citing $\forall$ elimination.

In essence, this is just using Theorem 5.8 to get $\forall x \varphi \to \varphi_t^x$ and then using $\to$ elimination.

Remark 5.39. Another way to see this is $\forall x P(x)$ is roughly saying

$$P(c_1) \wedge P(c_2) \wedge \cdots$$

and then deriving some individual $P(c_\bullet)$ via $\wedge$ elimination.

Non-Example 5.40. We remark that the substitutability condition is necessary. For example, the following is not a valid proof.

1	$\forall x \exists y (x \dot{\neq} y)$	
2	$\exists y (y \dot{\neq} y)$	Bad $\forall$Elim, 1

The issue is that y is not substitutable for x in $\forall x \exists y (x \dot{\neq} y)$.

Here is another example.

Exercise 5.41. We show $\forall x P(x) \to \neg \forall x \neg P(x)$.

Proof. We proceed as follows.

1	$\forall x P(x)$	
2	$\quad \forall x \neg P(x)$	
3	$\quad \neg P(c)$	$\forall$Elim, 2
4	$\quad \forall x P(x)$	Reit, 1
5	$\quad P(c)$	$\forall$Elim, 4
6	$\quad P(c) \wedge \neg P(c)$	$\wedge$Elim, 5, 3
7	$\neg \forall x \neg P(x)$	$\neg$Elim, 2–6
8	$\forall x P(x) \to \neg \forall x \neg P(x)$	$\to$Intro, 1–7

This finishes. ■

5.4 April 11

Welcome back everyone.

5.4.1 Universal Introduction

We continue with our discussion of natural deduction for first-order logic. Last time we discussed the elimination rule for $\forall$, which was like the elimination rule for $\wedge$.

Today we start with a discussion of introducing $\forall$, which will be like $\wedge$ introduction. For example, if there will only finitely many elements of the domain D, then we could check $\forall x \varphi$ by checking φ on $x := d$ for each of the finitely many $d \in D$. However, if the domain is infinite, we need to be careful; the idea is to prove for an "arbitrary" element d of the domain.

Idea 5.42. If we can prove something for an arbitrary object, then we can prove it for all objects.

Let's see an example.

Lemma 5.43. For any real numbers a and b, if $0 < a < b$, then $a^2 < b^2$.

Proposition 5.44. For any real numbers a and b, if $0 < a < b$, then $(a-b)^2 < a^2$.

Proof. Let c and d be any real numbers. Notably, we are allowing c and d be whatever they want to be. To start, we assume the hypothesis so that $0 < d < c$. Now, we see that $0 < c - d < c$, so it follows

$$(c-d)^2 < c^2$$

from Lemma 5.43, using $\forall$ elimination. But c and d were arbitrary real numbers, so any reals numbers a and b have, if $0 < a < b$, then $(a-b)^2 < a^2$. ■

Example 5.45. Let P be an arbitrary patient in a hospital. Because they are in the hospital, they have an email address logged in the database, so in particular, P has an email address. As such,

Let's write out the above proof in Fitch style proofs.

1	$\boxed{c, d}$	
2	$0 < d < c$	
3	$0 < c - d < c$	Basic math
4	$\forall x \forall y \left(0 < x < y \to x^2 < y^2\right)$	Lemma 5.43
5	$0 < c - d < d \to (c-d)^2 < d^2$	$\forall$Elim, 4
6	$(c-d)^2 < c^2$	$\to$Elim, 3, 5
7	$(0 < d < c) \to (c-d)^2 < c^2$	$\to$Intro, 2–6
8	$\forall u \forall v \left((0 < v < u) \to (u-v)^2 < u^2\right)$	$\forall$Intro, 1–7

The point of the boxes at the top is that they are newly introduced constants. As such, we make the slightly formal point that we will permit adding in some finitely many new constants $c_1, \dots, c_n$ without changing our semantics.

Remark 5.46. Technically, we are not allowed to introduce both $\forall$ at the same time, but the fix is just to introduce the quantifiers one at a time.

Anyway, here is our rule.

Definition 5.47 ($\forall$ Introduction)**.** Suppose we have the following in a given subproof.

- c is a constant not in the base language $\mathcal{L}$.
- We have a proof starting with $\boxed{c}$ and ending with φ_c^x.
- The constant c is nowhere outside the subproof.

Then we may write down $\forall x \varphi$ outside the current subproof.

Diagrammatically, this looks like the following.

1	$\boxed{c}$	
$\vdots$	$\vdots$	
n	φ_c^x	
$n+1$	$\forall x \varphi$	$\forall$Intro, 1–n

We have some extra hypotheses ensuring that c is really an "arbitrary" object, but that is all they are. The main point is that the subproof above is a "template" of sorts for any object in the domain.

Let's do another example.

Exercise 5.48. We show $\forall x (P(x) \to Q(x)) \to (\forall x P(x) \to \forall x Q(x))$.

Proof. We proceed as follows.

1	$\forall x(P(x) \to Q(x))$	
2	$\quad \forall x P(x)$	
3	$\quad\quad \boxed{c}$	
4	$\quad\quad \forall x P(x)$	Reit, 2
5	$\quad\quad P(c)$	$\forall$Elim, 4
6	$\quad\quad \forall x(P(x) \to Q(x))$	Reit, 1
7	$\quad\quad P(c) \to Q(c)$	$\forall$Elim, 6
8	$\quad\quad Q(c)$	$\to$Elim, 5, 7
9	$\quad \forall x Q(x)$	$\forall$Intro, 3–8
10	$\forall x P(x) \to \forall x Q(x)$	$\to$Intro, 2–9
11	$\forall x(P(x) \to Q(x)) \to (\forall x P(x) \to \forall x Q(x))$	$\to$Intro, 1–10

The main point is to keep unwinding our quantifiers down until we get to $\forall x Q(x)$, from which point we know that we should introduce a constant c. Then, once we are dealing with no quantifiers, we are essentially down to working in propositional logic (and eliminating quantifiers). ■

Remark 5.49. As a quick check-in, one can still show that our proof system is "sound": if ψ has a proof from the assumptions $\varphi_1, \dots, \varphi_n$, then ψ is still a semantic consequence of the formulae $\varphi_1, \dots, \varphi_n$. The converse is called completeness.

If we add in the axiom $\exists x\varphi \leftrightarrow \neg\forall x\neg\varphi$ (i.e., if φ is true for some x, then it is not the case that φ fails for all x), then we have a fully sound and complete axiom for first-order logic.

Nevertheless, we go on to give introduction and elimination rules for $\exists$. While the axiom $\exists x\varphi \leftrightarrow \neg\forall x\neg\varphi$ is valid, people do not in general think about $\exists$ like this, and we would like our proof system to more faithfully mirror how people think.

5.4.2 Existential Introduction

We start with existential introduction, which is in some sense dual to universal elimination.

Recall from Proposition 3.46 that we were able to conjure the existence of irrational numbers a and b such that a^b is rational, but we did not know what the a and b were. In contrast, here is a constructive proof.

Proposition 3.46. There are irrational numbers a and b such that a^b is rational.

Proof. The point is to use logarithms. Set $r \coloneqq \log_3 2$ so that $2^r = 3$.[1] Further, recall that we showed $\sqrt{2}$ is irrational (constructively via $\neg$ introduction in Proposition 3.35).

Quickly, we show that $2\log_3 2$ is irrational. Well, if we could write $2\log_3 2 = \frac{m}{n}$ for positive integers m and n, then $\log_3 2 = \frac{m}{2n}$, so $2^{m/2n} = 3$ by substitution, so

$$2^m = 3^{2n}.$$

[1] Here is one way to see this: the set $\{r \in \mathbb{R} : 2^r < 3\}$ is upper-bounded, and so it will have a maximum. The maximum is the number that we want.

Now, $m \neq 0$, so 2^m is even, but 3^{2m} is odd, so we have our contradiction.

So we can conclude that $2\log_3 2$ is irrational, So to finish, we observe that

$$\left(\sqrt{2}\right)^{2\log_3 2} = \left(\left(\sqrt{2}\right)^2\right)^{\log_3 2} = 2^{\log_3 2} = 3,$$

so we have two irrational numbers $a = \sqrt{2}$ and $b = \log_3 2$ such that $a^b = 3$. ■

Diagrammatically, the point of the above proof is that we had the following.

$\vdots$	$\vdots$
i	$\neg\,\mathrm{Rat}(\sqrt{2})$
$\vdots$	$\vdots$
j	$\neg\,\mathrm{Rat}(\log_3 2)$
$\vdots$	$\vdots$
k	$\mathrm{Rat}(\sqrt{2}^{\log_3 2})$
$k+1$	$\neg\,\mathrm{Rat}(\sqrt{2}) \wedge \neg\,\mathrm{Rat}(\log_3 2) \wedge \mathrm{Rat}(2^{\log_3 2})$
$k+2$	$\exists x \exists y\big(\neg\,\mathrm{Rat}(x) \wedge \neg\,\mathrm{Rat}(y) \wedge \mathrm{Rat}(x^y)\big)$

The moral of the story is that we can prove an existential by going and constructing the objects.

5.5 April 13

Welcome back, everyone.

5.5.1 Existential Introduction

We continue talking about existential introduction. Here is our rule.

Definition 5.50 (∃ Introduction)**.** Suppose we have a formula φ_t^x where t is a term substitutable for x in φ. Then we can deduce $\exists x\varphi$ in the same subproof and cite existential introduction.

Example 5.51. Set $\varphi \coloneqq P(x)$. If we can deduce $P(c) = \varphi_c^x$ for some constant c, then we can deduce $\exists x P(x)$.

Remark 5.52. This makes sense given that we are viewing $\exists x P(x)$ as an infinite conjunction: given any $P(c)$, we can kind of deduce

$$\bigvee_{c \in \mathrm{Const}} P(c),$$

which is approximately $\exists x P(x)$.

We very quickly note that we can remove applications of ∃ introduction with instead using ∀ elimination and ¬ elimination with the axiom

$$\exists x\varphi \leftrightarrow \neg\forall\neg\varphi.$$

Concretely, we can turn the proof

1	φ_t^x	
2	$\exists x \varphi$	$\exists$Intro, 1

into the proof as follows.

1	φ_t^x	
2	$\quad \forall x \neg \varphi$	
3	$\quad \neg \varphi_t^x$	$\forall$Elim, 2
4	$\quad \varphi_t^x$	Reit, 1
5	$\quad \varphi_t^x \wedge \neg \varphi_t^x$	$\wedge$Intro, 3, 4
6	$\neg \forall x \neg \varphi$	$\neg$Intro, 2–5

And here is an example.

Exercise 5.53. We show $\forall x P(x) \to \exists x P(x)$.

Proof. Importantly, we are assuming that we are working in a nonempty domain. Anyway, we have the following proof.

1	$\quad \forall x P(x)$	
2	$\quad P(y)$	$\forall$Elim, 1
3	$\quad \exists x P(x)$	$\exists$Intro, 2
4	$\forall x P(x) \to \exists x P(x)$	$\to$Intro, 1–3

This finishes. ∎

Remark 5.54. In particular, the above proof makes our proof system unsound if we allow empty domains. There is a notion of "free logic" which fixes this.

5.5.2 Identity Elimination

As an intermission before existential elimination, we discuss identity. Here is an example proof.

Proposition 5.55. We prove $3 < \sqrt{11} < 4$.

Proof. We start by showing $3 < \sqrt{11}$. Well, note that $9 < 11$, so taking square roots gives $\sqrt{9} < \sqrt{11}$. However, $3 = \sqrt{9}$, so we may say $3 < \sqrt{11}$.

Next we show $\sqrt{11} < 4$. Well, note that $11 < 16$, so taking square roots gives $\sqrt{11} < \sqrt{16}$. However, $4 = \sqrt{16}$, so we may say $\sqrt{11} < \sqrt{16}$. ∎

The point we are emphasizing is our movement from $\sqrt{9} < \sqrt{11}$ to $3 < \sqrt{11}$. This is an elimination rule; diagrammatically, it looks like the following.

$$\begin{array}{l|ll} \vdots & \vdots & \\ i & \sqrt{9} < \sqrt{11} & \\ i+1 & 3 = \sqrt{9} & \\ i+2 & 3 < \sqrt{11} & \doteq\text{Elim}, i+1, i \end{array}$$

Similar works for 4.

Here is our rule.

Definition 5.56 ($\doteq$Elim)**.** Suppose we have a correct partial proof with the following constraints.

- $t_1 \doteq t_2$ and φ are in the same column.
- All variables of t_1 are free in φ (to prevent binding).
- t_2 is substitutable for t_1 in φ.
- φ' is obtained by replacing some (but perhaps not all!) occurrences of t_1 by t_2 in φ.

Then we can deduce φ' and cite identity elimination.

Diagrammatically, this looks like the following.

$$\begin{array}{l|ll} \vdots & \vdots & \\ i & t_1 \doteq t_2 & \\ \vdots & \vdots & \\ j & \varphi & \\ \vdots & \vdots & \\ k & \varphi' & \doteq\text{Elim}, i, j \end{array}$$

Example 5.57. Given $\sqrt{9} = 3$ and $(\sqrt{9} < \sqrt{11}) \wedge (\sqrt{9} = \sqrt{9})$, we can deduce $(3 < \sqrt{11}) \wedge (\sqrt{9} = \sqrt{9})$.

Here is an example.

Example 5.58 (Transitivity)**.** We can show transitivity, as follows.

$$\begin{array}{l|ll} 1 & a \doteq b & \\ 2 & b \doteq c & \\ \hline 3 & a \doteq c & \doteq\text{Elim}, 1, 2 \end{array}$$

Namely, the formula "$a \doteq c$" is the formula "$b \doteq c$" where we have replaced occurrences of b with a.

Non-Example 5.59. The following is not good, for binding reasons.

1	$x \doteq y$	
2	$\forall x(x \doteq x)$	
3	$\forall x(x \doteq y)$	$\doteq$Elim, 1, 2

Notably, the variable x is not free in $\forall x(x \doteq x)$—it already has a meaning! So we can't really replace for it.

Non-Example 5.60. The following is also not good, for substitutability reasons.

1	$x \doteq y$	
2	$\forall y P(x)$	
3	$\forall y P(y)$	$\doteq$Elim, 1, 2

Again, the issue is that y is not substitutable for x in $\forall y P(x)$ because this would make the new variable bound to $\forall y$.

Lastly, we mention our introduction rule for identity.

Definition 5.61 ($\doteq$Intro)**.** Given any term t, we get to write down $t \doteq t$ and cite $\doteq$ introduction.

Essentially, the idea is that the reflexive property of identity ought to be true, so it will be our introduction.

Here is a last example.

Exercise 5.62 (Symmetry)**.** We derive $b \doteq a$ from $a \doteq b$.

Proof. We have the following.

1	$a \doteq b$	
2	$a \doteq a$	$\doteq$Intro
3	$b \doteq a$	$\doteq$Elim, 1, 2

Notably, we replaced some but not all instances of a in $a \doteq a$ with a b, citing $a \doteq b$. ■

5.5.3 Existential Elimination

This is our last rule for natural deduction!

Recall that, to use a disjunction $\alpha \vee \beta$, we show $\alpha \to \varphi$ and show $\beta \to \varphi$ so that we can conclude φ in

either case. Diagrammatically, this looks like the following.

a	$\alpha \vee \beta$	
b	$\quad \alpha$	
c	$\quad \varphi$	
d	$\quad \beta$	
e	$\quad \varphi$	
n	φ	$\vee$Elim, a, b–c, d–e

Here is an example of a proof.

Definition 5.63 (Divides)**.** We say that an integer a divides an integer b if and only if there exists an integer c such that $b = ac$.

Proposition 5.64. For all integers x, y, z, if $x \mid y$ and $y \mid z$, then $x \mid z$.

Proof. As usual, pick up some specific integers a, b, c and suppose $a \mid b$ and $b \mid c$. As such, $a \mid b$ promises an integer m such that $b = am$; similarly, $b \mid c$ promises an integer n such that $c = bn$. But then we can write

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

so we deduce $a \mid c$. Notably, we are using some various symmetry and transitivity of identity laws to deduce $c = a(mn)$. To finish, we recall that a, b, c were arbitrary, so we are done. ■

Notably, everything above is justified, except perhaps our associativity step above. The main point of the proof is to "undo" the definition of divides so that we could find specific integers m and n such that $b = am$ and $c = bn$. Then our conclusion $a \mid c$ was able to throw out our auxiliary variables, so we get to actually conclude $a \mid c$.

Diagrammatically, this looks like the following.

1	$\quad \boxed{a}\,\boxed{b}\,\boxed{c}$	
2	$\quad\quad (\exists v(xv = y) \wedge \exists v(yv = z)$	
3	$\quad\quad \exists v(xv = y)$	
4	$\quad\quad \exists v(yv = z)$	
5	$\quad\quad \exists v(xv = z)$	
6	$\quad\quad\quad am = b \quad \boxed{m}$	
7	$\quad\quad\quad bn = c \quad \boxed{n}$	
8	$\quad\quad\quad a(mn) = c$	Basic math
9	$\quad (\exists v(xv = y) \wedge \exists v(yv = z) \to \exists v(xv = z))$	
10	$\forall x \forall y \forall z((\exists v(xv = y) \wedge \exists v(yv = z) \to \exists v(xv = z)))$	

The point is at lines 6 and 7, where we have used our existential quantifiers. We have not been entirely formal; next class we will actually give our rule.

5.6 April 15

Welcome back, everyone.

5.6.1 Existential Elimination

Last class we were talking about existential elimination. The idea is that a statement

$$\exists x \varphi$$

allows us to conjure a "witness" constant n for x satisfying φ and then proceed with the proof. Here is the formal rule.

Definition 5.65 ($\exists$ Elimination)**.** Suppose we have a statement $\exists x \varphi$. If granted a constant c (not in the original language $\mathcal{L}$) with a subproof starting with φ_c^x and ending with ψ (which does not contain c!), then we can add ψ to the main line outside the subproof and cite $\exists$ elimination.

Diagrammatically, this looks like the following.

$\vdots$	$\vdots$	
j	$\exists x \varphi$	
$j+1$	$\varphi_c^x \quad \boxed{c}$	
$\vdots$	$\vdots$	
k	ψ	
$k+1$	ψ	$\exists$Elim, j, $j+1$–k

And let's see an example.

Exercise 5.66. We derive $\exists x P(x) \vee \exists x Q(x)$ from $\exists x (P(x) \vee Q(x))$.

Proof. We have the following.

1	$\exists x (P(x) \vee Q(x))$	
2	$P(c) \vee Q(c) \quad \boxed{c}$	
3	$P(c)$	
4	$\exists x P(x)$	$\exists$Intro, 3
5	$\exists x P(x) \vee \exists x Q(x)$	$\vee$Intro, 4
6	$Q(c)$	
7	$\exists x P(x)$	$\exists$Intro, 6
8	$\exists x P(x) \vee \exists x Q(x)$	$\vee$Intro, 7
9	$\exists x P(x) \vee \exists x Q(x)$	$\vee$Elim, 2, 3–5, 6–8
10	$\exists x P(x) \vee \exists x Q(x)$	$\exists$Elim, 1, 2–9

This finishes. ∎

Remark 5.67. Again, this is like $\vee$ elimination as a proof-by-cases for an arbitrary number of cases. To see this, we note we can derive ψ from

$$P(c_1) \vee \cdots \vee P(c_n)$$

by doing a proof that $P(c_i)$ would imply ψ for each case, and from here we can use $\vee$ elimination. The ability to do all cases at once for some "arbitrary" c (to show ψ from $P(c)$) is what we've done above. In some sense, we are doing "all" of the cases at the same time, via some "template" proof.

Remark 5.68. As such, the above example is more or less like $\vee$ distributing over $\vee$, using the intuition that $\exists$ is a very large $\vee$.

Remark 5.69. We also have $(\forall x P(x) \wedge \forall x Q(x)) \leftrightarrow (\forall x (P(x) \wedge Q(x))$.

This finishes our discussion of natural deduction.

Theme 6
Arithmetic and Set Theory

6.1 April 15

We now transition into discussing arithmetic, using first-order logic. The idea is to add a few axioms to be able to discuss actual mathematics.

6.1.1 Peano Arithmetic

Let's see an example of an arithmetic proof, to start off.

Lemma 6.1. Let n be a nonnegative integer. Then n is even or odd.

Proof. We proceed by induction on n. As our base case, we note that $n = 0$ has $0 = 2 \cdot 0$, so 0 is even, so 0 is even or odd.

We now show the inductive step. Suppose that n is even or odd, and we will show $n + 1$ is even or odd. We have two cases.

- If n is even, then write $n = 2k$. Thus, $n + 1 = 2k + 1$, so $n + 1$ is odd, so $n + 1$ is even or odd.
- If n is odd, then write $n = 2k + 1$. Thus, $n + 1 = 2(k + 1)$, so $n + 1$ is even, so $n + 1$ is even or odd.

In all cases we were able to conclude that $n + 1$ is even or odd, so we are done. ■

The main point is that we are able to formalize all parts of this proof except for the induction. So we will build an induction.

For our arithmetic, we will build a language with the following tools.

Definition 6.2 (Peano arithmetic language)**.** The language of Peano arithmetic has the following data.

- No predicate symbols.
- One constant symbol $\dot{0}$, sometimes written 0.
- One unary function symbol S.
- Two binary function symbols $\dot{+}$ and $\dot{\times}$, sometimes written $+$ and $\times$.

And here are some conventions.

- $S(0)$ will be denoted 1, $S(S(0))$ will be denoted 2, and so on.

- We might also omit $\times$ and write (for example, $2k$) for $2 \times k$.
- We'll write $=$ for $\doteq$ for brevity.

For our arithmetic (which will be called Peano arithmetic) is that we will start with some basic axioms to avoid an infinite regress. Here they are.

Axiom 6.3 (Peano arithmetic)**.** Peano arithmetic takes the following axioms. We begin by pinning down S.

(S1) $\forall x(\neg S(x) = 0)$. In other words, 0 is not the successor (of any natural number).

(S2) $\forall x \forall y((S(x) = S(y)) \to (x = y))$. In other words, taking successors is one-to-one.

Next we talk about how $+$ behaves, more or less recursively.

(A1) $\forall x(x + 0 = x)$. In other words, adding 0 doesn't do anything.

(A2) $\forall x \forall y(x + S(y) = S(x + y))$. This codifies the associativity $x + (y + 1) = (x + y) + 1$ and defines $+$.

Here is multiplication, again recursively.

(M1) $\forall x(x \times 0 = 0)$. In other words, multiplying by 0 gets 0.

(M2) $\forall x \forall y(x \times S(y)) = (x \times y) + x$. Expanding out, we are roughly saying that $x \times (y + 1) = x \times y + x$, codifying the notion that multiplication is repeated addition.

Lastly, we need induction.

(IND) For any formula φ, we have $(\varphi^x_0 \wedge \forall x(\varphi \to \varphi^x_{S(x)}) \to \forall x \varphi$.

In other words, if we can show φ "holds" for 0 and that φ "holding" for x implies that φ "holding" for $x + 1$, then we may deduce $\forall x \varphi$.

Notably, we needed to talk about function symbols, constant symbols, and quantifiers in our discussion of first-order logic to be able to codify the above.

Remark 6.4. Even though our axiom list is infinite (namely, induction gives an axiom for each formula φ), our axiom list is at least computable: all the axioms provided by induction have a very specific form.

Let's close with an example.

Exercise 6.5. We show $\forall x(x + 1 = S(x))$.

Proof. We have the following.

1	$\forall x(x + 0 = x)$	
2	$\forall x(x + S(y) = S(x + y))$	
3	$\boxed{c}$	
4	$c + 0 = c$	L$\forall$Elim, 1
5	$c + 1 = S(c + 0)$	L$\forall$Elim, 2
6	$c + 1 = S(c)$	$=$Elim, 4, 5
7	$\forall x(x + 1 = S(x))$	$\forall$Intro, 3–6

This finishes. Notably, we didn't even need induction. ■

Remark 6.6. Notably, we can now proof $\forall$ statements by either $\forall$ introduction or by induction.

Remark 6.7. Without any axioms, the formula "$\forall x(x+1 = S(x))$" need not be true. Our axioms are now giving our semantics.

6.2 April 18

We take a pause from lecturing to discuss concepts from the problem set.

6.2.1 Concept Review

Here are some notes from the class's discussion.

- To say "all cats are cute," write $\forall x(\mathrm{Cat}(x) \to \mathrm{Cute}(x))$. Namely, if x is a cat, then x is cute; if x is not a cat, we don't assert anything.
- To say "there is a cute cat," write $\exists x(\mathrm{Cat}(x) \wedge \mathrm{Cute}(x))$.
- To show that a formula is not valid, we merely have to give a single model. For example,

  $$\vDash \forall x \forall y \forall z\big((R(x,y) \wedge R(y,z)) \to R(x,z)\big)$$

  is not valid. To see this, take the model $\mathcal{M} = (D, I)$ with $D = \{0,1,2\}$. To make the above conditional false, we would like to make the antecedent $R(x,y) \wedge R(y,z)$ true and $R(x,z)$ false. As such, we set

  $$I(R) \coloneqq \{\langle 0,1\rangle, \langle 1,2\rangle\}$$

  which has

  $$\mathcal{M} \nvDash_g \forall x \forall y \forall z\big((R(x,y) \wedge R(y,z)) \to R(x,z)\big).$$

  Smaller counterexamples do exist: for example, take $\mathcal{M} = (\{0,1\}, I)$ where $I(R) = \{\langle 0,1\rangle, \langle 1,0\rangle\}$.
- In general, to translate a sentence with a quantified object, we need a quantified variable. For example, to translate "Some dogs are not liked by all cats," we need a quantifier for the existence of dogs and a quantifier for all cats. Concretely, we have

  $$\exists x(\mathrm{Dog}(x) \wedge \neg \forall y(\mathrm{Cat}(y) \to \mathrm{Loves}(y,x))).$$

  As a warning, many of the problems on the problem set are ambiguous, so do not be too worried about choosing the exact correct interpretation as long as it is reasonable.
- For complicated sentences, it is often helpful to "chunk" parts of the sentence and work on them individually. For example, in "Every cat that is loved by all cats doesn't love any dog," we might start by thinking

  $$\forall x\big((\mathrm{Cat}(x) \wedge x \text{ is loved by all cats})) \to x \text{ loves no dog}\big).$$

  Then we can break down each piece as follows

  $$\forall x\big((\mathrm{Cat}(x) \wedge \forall y(\mathrm{Cat}(y) \to \mathrm{Loves}(y,x))) \to \neg \exists z(D(z) \wedge \mathrm{Loves}(x,z))\big).$$

- In general, it is very awkward to give implications inside an existential quantifier. If we must, the way to think about this is (for example) as

  $$\exists x(P(x) \to Q(x)) \equiv \exists x(\neg P(x) \vee Q(x)) \equiv \exists x \neg P(x) \vee \exists x Q(x).$$

  Namely, to say $\exists x(P(x) \to Q(x))$ can be cleanly split.

- Another trick to translate is to translate first into English as an intermediate step (to some easier statement) and then finish. For example, to translate "Nothing has A except those that have B," we can be convinced that this means "Everything has either not A or has B." So this translates as

$$\forall x(\neg A(x) \vee B(x)).$$

  Equivalently, we can write this as $\forall x(A(x) \to B(x))$ or even $\neg\exists x(A(x) \wedge \neg B(x))$.

  So for the last one,

> No cat loves a cat who is loved by a dog, except for the cats who love cats

  can be turned into

> All cats x either do not love a cat who is loved by a dog or x loves cats.

  It might be hard to formalize "x loves cats," but we do our best.

- We might want to add quantifiers like "most" to our first-order logic so that we can say things like "most cats are cute."

6.3 April 20

We continue discussing problem set 10.

6.3.1 Concept Review

Here we go.

- Some sentences are ambiguous. For example, "There is no largest integer divisible by 3" has two different readings, as follows.
  - There is no largest integer, which happens to be divisible by 3.
  - Among the integers divisible by 3, there is no largest.

  If this concerns you on the problem set, just mark it.

- Prenex form of a formula requires that all quantifiers appear in the front. For example, $\forall x\exists yP(x,y)$ is in prenex normal form.

6.4 April 22

6.4.1 Induction

Let's give an example proof using induction.

Proposition 6.8. We show that $\forall x\forall y\forall z((x+y)+z = x+(y+z))$.

Proof. We proceed as follows.

1	$\boxed{a}$	
2	$\quad\boxed{b}$	
3	$\quad\forall x(x+0=x)$	
4	$\quad(a+b)+0=a+b$	$\forall$Elim, 3
5	$\quad b+0=b$	$\forall$Elim, 3
6	$\quad(a+b)+0=a+(b+0)$	$=$Elim, 4, 5
7	$\quad\boxed{c}$	
8	$\quad\quad(a+b)+c=a+(b+c)$	
9	$\quad\quad\forall x\forall y(x+S(y))=S(x+y)$	(A2)
10	$\quad\quad\forall y(x+S(c)=S(x+c))$	$\forall$Elim, 9
11	$\quad\quad b+S(c)=S(b+c)$	$\forall$Elim, 9
12	$\quad\quad a+(b+S(c))=a+(b+S(c))$	$=$Intro
13	$\quad\quad a+(b+S(c))=a+S(b+c)$	$=$Elim, 11
14	$\quad\quad a+S(b+c)=S(a+(b+c))$	$\forall$Elim, 9
15	$\quad\quad(a+b)+S(c)=S((a+b)+c)$	$\forall$Elim, 9
16	$\quad\quad(a+b)+S(c)=a+(b+S(c)$	Many $=$Elim
17	$\quad(a+b)+c=a+(b+c)\to(a+b)+S(c)=a+(b+S(c))$	$\to$Intro, 8–16
18	$\quad\forall z((a+b)+z=a+(b+z)\to(a+b)+S(z)=a+(b+S(z)))$	
19	$\quad\varphi_0^z\wedge\forall z(\varphi\to\varphi_{S(z)}^z)$ for $\varphi:=(a+b)+z=a+(b+z)$	$\wedge$Intro, 6, 18
20	$\quad(\varphi_0^z\wedge\forall z(\varphi\to\varphi_{S(z)}^z))\to\varphi$	(IND)
21	$\quad\forall z((a+b)+z=a+(b+z))$	$\to$Elim, 19, 20
22	$\forall y\forall z((a+y)+z=a+(y+z))$	$\forall$Intro, 2–21
23	$\forall x\forall y\forall z((x+y)+z=x+(y+z))$	$\forall$Intro, 1–22

Notably, we have split the proof into a base case, which comes down to shifting $x+0=x$ around, and an inductive step, which is the business with the c. The point of the inductive step is the application of (A2). ■

Lemma 6.9. Any natural number n is either even or odd.

Proof. We induct on n. For $n=0$, we note $n=2\cdot 0$ is even. For the inductive step, take n to be either even or odd.

- If n is even, then $n=2k$, so $n+1=2k+1$ is odd.

- If n is odd, then $n = 2k + 1$, so $n + 1 = 2(k + 1)$ is odd.

These computations finish the proof. ■

And here is our formalization.

Exercise 6.10. We show that $\forall n(\exists k(n = 2k) \vee \exists k(n = 2 \times k + 1))$.

Proof. We proceed as follows.

1	$2 \times 0 = 0$	$\forall x(x \times 0 = 0)$
2	$0 = 2 \times 0$	Use symmetry
3	φ_0^n	$\vee$Intro, 2
4	$\quad \varphi$	
5	$\quad\quad \exists k(n = 2k)$	
6	$\quad\quad\quad n = 2c \quad \boxed{c}$	
7	$\quad\quad\quad n + 1 = 2c + 1$	$=$Elim, 6
8	$\quad\quad\quad S(n) = n + 1$	Shown in class
9	$\quad\quad\quad S(n) = 2c + 1$	$=$Elim, 8, 7
10	$\quad\quad\quad \exists k(S(n) = 2k + 1)$	$\exists$Intro, 9
11	$\quad\quad\quad \varphi_{S(n)}^n$	$\exists$Elim, 10
12	$\quad\quad \varphi_{S(n)}^n$	$\vee$Elim, 11
13	$\quad\quad n = 2k + 1$	
14	$\quad\quad\quad n = 2c + 1 \quad \boxed{c}$	
15	$\quad\quad\quad n + 1 = n + 1$	$=$Intro
16	$\quad\quad\quad n + 1 = (2c + 1) + 1$	$=$Elim, 13, 15
17	$\quad\quad\quad n + 1 = 2c + 2$	Association
18	$\quad\quad\quad n + 1 = 2(c + 1)$	Distribution, 17
19	$\quad\quad \exists k(n + 1 = 2(k + 1))$	$\exists$Elim, 14–18
20	$\quad\quad \varphi_{S(n)}^n$	$\vee$Intro, 19
21	$\quad \varphi_{S(n)}^n$	$\vee$Elim, 4, 5–12, 13–20
22	$\varphi \to \varphi_{S(n)}^n$	$\to$Intro, 4–21
23	$\varphi_0^n \wedge (\varphi \to \varphi_{S(n)}^n)$	$\wedge$Elim, 3, 22
24	$(\varphi_0^n \wedge (\varphi \to \varphi_{S(n)}^n)) \to \varphi$	(IND)
25	$\forall n \varphi$	$\to$Elim, 23, 24

The point is to induct on the formula φ defined as $\exists k(n = 2k) \vee \exists k(n = 2k+1)$. Notably, we are applying our previous results of association and $\forall x(S(x) = x + 1)$, which we showed earlier. We have also abbreviated the $\exists$ out of φ because of space reasons. ■

6.5 April 25

Welcome back, everyone.

6.5.1 Limits of Peano Arithmetic

It turns out that not everything is possible in Peano arithmetic. Nonetheless, here are some things that we can define.

- Evenness and oddness can be defined.
- The less than or equal to relation $\leq$ can be defined as $x \leq y$ as $\exists z(x + z = y)$.
- Primality can be defined as $\mathrm{Prime}(x)$, meaning
$$x \neq 1 \wedge \forall y \forall z(x = yz \to (y = 1 \vee z = 1)).$$
- It also turns out that we can define exponentiation in the following sense: we can make a formula $\mathrm{Exp}(x, y, z)$ which is true if and only if $x^y = z$. However, this is hard. More frequently, we just add the axioms
$$\forall x \left(x^0 = S(1)\right) \quad \text{and} \quad \forall x \forall y \left(x^{S(y)} = x^y \cdot x\right)$$
to our arithmetic.

One can also formalize many results into Peano arithmetic, such as the following.

- One can prove that there are infinitely many prime numbers.
- Some believe that the proof of Fermat's last theorem
$$\forall x \forall y \forall z \left(2 < n \to x^n + y^n = z^n\right)$$
can be formalized in Peano arithmetic. However, this would take a while.

To see our limits, we need some terminology.

Definition 6.11 (Negation complete)**.** A set of axioms Σ is *negation complete* if and only if, for every sentence φ, we either have $\Sigma \vdash \varphi$ or $\Sigma \vdash \neg\varphi$. In other words, Σ can prove φ or can prove $\neg\varphi$. Otherwise, Σ is *negation incomplete*.

And here is our corresponding result.

Theorem 6.12 (Gödel)**.** If Peano arithmetic is consistent (i.e., one cannot derive contradiction), then Peano arithmetic is negation incomplete.

Remark 6.13. Note that there is no notion of truth in this explanation: we are not claiming that Peano arithmetic is unable to prove true statements. What we are saying is that there is a formula φ such that Peano arithmetic cannot prove φ nor $\neg\varphi$.

The statement of Theorem 6.12 might appear weak: perhaps we need more axioms. However, there is a stronger notion.

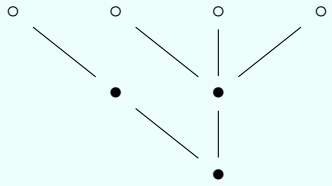
Remark 6.14. Most people believe that Peano arithmetic is consistent (e.g., there are proofs of its consistency from other principles which we believe), so we will just assume this.

6.5.2 Incompleteness of Peano Arithmetic

Let's give an example to exhibit Theorem 6.12. We will need to discuss the hydra game.

Definition 6.15 (Hydra)**.** A *hydra* is a finite rooted tree.

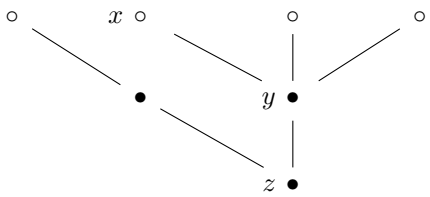
Example 6.16. The following is a hydra.



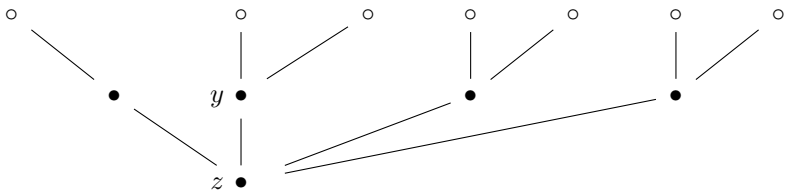
Now, here is the game we will play with the hydra.

1. Every round, we choose some leaf x other than the root, and we chop off its connection to its parent y below.
2. Then the hydra responds.
   - If y is the root, nothing happens.
   - If there is a node z below y, then the hydra regrows n copies of the subtree from y (with x chopped off), all connected to z.

For example, we might chop off x as follows, on round $n = 2$.



This gives the following.



Now, we win this game if and only if we eventually chop off the root. The issue, of course, is that winning the game will take a while because the hydra gets very big as the step counter n increases.

So the question is if we can create an algorithm to slay the hydra.

Example 6.17. We might try to simply chop off the leftmost head of the hydra.

Example 6.18. We might try to chop off a random node furthest from the root.

However, it is actually not that hard to beat the hydra.

Theorem 6.19 (Kirby and Paris)**.** Any algorithm to play the hydra game will always win against any hydra.

The point is that Theorem 6.19 can be formalized in Peano arithmetic; roughly speaking, the point is that we can encode algorithms, finite trees, and games all with nonnegative integers. To believe this, computers essentially do this already with binary.

However, we have the following concrete incompleteness.

Theorem 6.20 (Kirby and Paris)**.** Peano arithmetic, if consistent, cannot prove Theorem 6.19.

Remark 6.21. Even though we cannot prove Theorem 6.19, it is still true (say, in ZFC).

Remark 6.22. It also turns out that consistency of Peano arithmetic is not provable from inside Peano arithmetic. However, this consistency can be encoded into Peano arithmetic (via integers, as usual), so this is another witness to incompleteness.

Again, one might complain that Peano arithmetic is simply too weak, and we just need to add some axioms to fix it. However, Gödel disagrees.

Theorem 6.23 (Gödel)**.** Let T be any set of sentence containing all sentences that Peano arithmetic can prove. Further, suppose that there is an algorithm to determine if a sentence is in T. Then T is negation incomplete.

Remark 6.24. Do note that the axioms for Peano arithmetic are algorithmically recognizable—simply check the fixed sentences other than induction by hand and then check the "format" for the induction schema.

Remark 6.25. The algorithmic requirement might seem awkward, but it is necessary, for otherwise we could just set T to be the set of sentences which are modeled by $\mathbb{N}$. It follows that there is no algorithm to determine if a sentence is modeled by $\mathbb{N}$.

Thus, Peano arithmetic cannot be fixed by merely adding axioms.

6.5.3 Set Theory Appetizer

To be able to talk about "real" mathematics, we will end the semester by discussing set theory, which is the actual foundation for mathematics. Roughly speaking, we are claiming the following.

- All mathematical statements can be translated into statements about sets.
- Proving statements in mathematics is equivalent to proving them in our set theory.

6.6 April 27

Welcome back, everyone.

6.6.1 Numbers as Sets

We now turn to set theory.

Remark 6.26. It is possible to think about all objects in mathematics as sets.

Let's manifest the above remark.

Exercise 6.27. We encode the natural numbers as sets.

Proof. There are a few ways to do this; here is one.

- Set 0 to be $\varnothing$.
- Set 1 to be $\{0\} = \{\varnothing\}$.
- Continuing, set 2 to be $\{0, 1\} = \{\varnothing, \{\varnothing\}\}$.
- In general, define $S(n)$ recursively to be $\{0, 1, \dots, n\}$.

With this in mind, we will now define $\mathbb{N}$ to be the smallest set (defined inductively) containing $\varnothing$ and closed under the map

$$S \colon x \mapsto x \cup \{x\}.$$

In particular, $S \colon \mathbb{N} \to \mathbb{N}$ is $S(x) \coloneqq x \cup \{x\}$. Our Peano axioms "make" $\mathbb{N}$. Then $+$ and $\times$ are defined recursively to make true

$$\begin{cases} n + 0 \coloneqq n, \\ n + S(m) \coloneqq S(n + m), \end{cases} \quad \text{and} \quad \begin{cases} n \times 0 \coloneqq 0, \\ n \times m \coloneqq (n \times m) + n. \end{cases}$$

This is, approximately speaking, the idea behind von Neumann universe. ∎

Remark 6.28. To define addition and multiplication recursively, as we have done here, we need to know that these recursive definitions force a unique function; this is by the Recursion theorem, which we will not say more about.

As an aside, addition lets us define a notion of subtraction:

$$c = a - b \iff a = b + c.$$

However, this might not be possible for all natural numbers. So next we construct the integers $\mathbb{Z}$.

Exercise 6.29. We encode ordered pairs as sets.

Proof. We simply define

$$\langle x, y \rangle \coloneqq \{\{x\}, \{x, y\}\}.$$

The point is that $\langle a, b \rangle = \langle a', b' \rangle$ if and only if $a = a'$ and $b = b'$. To see this, we just note that $\{a'\} \in \langle a, b \rangle$ forces $\{a'\} = \{a\}$ or $\{a'\} = \{a, b\}$, from which we can read $a = a'$. Then $b = b'$ follows similarly. ∎

Exercise 6.30. We encode the integers as sets.

Proof. The idea is to use the pair $\langle n, m \rangle$ of natural numbers to encode $n - m$. This does not quite work, however, because $\langle 2, 4 \rangle$ and $\langle 4, 6 \rangle$ should be the same negative number. To fix this, we define the equivalence relation

$$\langle n, m \rangle \sim \langle n', m' \rangle \iff n + m' = n' + m.$$

This condition is meant to encode $n - m = n' - m'$ without ever writing down subtraction. As such, we define $\mathbb{Z}$ to be the equivalence classes of $\sim$; concretely, the equivalence class for $\langle n, m\rangle$ as $[\langle n, m\rangle] \in \mathbb{Z}$. For example,

$$-2 \coloneqq [\langle n, m\rangle] = \{\langle 0, 2\rangle, \langle 1, 3\rangle, \langle 2, 4\rangle, \dots\}.$$

Then we can define addition and multiplication by

$$\begin{aligned}
[\langle n, m\rangle] + [\langle n', m'\rangle] &= [\langle n + n', m + m'\rangle], \\
[\langle n, m\rangle] \times [\langle n', m'\rangle] &= [\langle (n \times n') + (m \times m'), (n \times m') + (m \times n')\rangle].
\end{aligned}$$

Notably, these operations make sense because they are encoding the intuitive differences as

$$\begin{aligned}
(n - m) + (n' - m') &\coloneqq (n + n') - (m + m'), \\
(n - m) \times (n' - m') &\coloneqq (n \times n') - (m \times n') - (n \times m') + (m \times m').
\end{aligned}$$

One needs to verify that these operations behave that we want them to while also being well-defined operations (i.e., do not depend on the "representative chosen" in the equivalence class). ■

Remark 6.31. In the future, to be able to embed $\mathbb{N}$ into $\mathbb{Z}$, we will think about our natural numbers as embedded by the image of the map $n \mapsto [\langle n, 0\rangle]$. This might be a little dangerous because we have multiple notions of 2 (namely, in $\mathbb{N}$ and in $\mathbb{Z}$), but in practice we might as well just forget that we had a 2 in $\mathbb{N}$ to begin with and work with its image in $\mathbb{Z}$ instead.

And now we keep going. Next up is rationals.

Exercise 6.32. We encode rational numbers as sets.

Proof. The idea is to represent the rational number $\frac{a}{b}$ by some ordered pair $\langle a, b\rangle$ where a and b are integers and $b \neq [\langle 0, 0\rangle]$. Again, we need to identify some ordered pairs because (for example) $\frac{1}{2} = \frac{2}{4}$. As such, we define the equivalence relation $\sim$ by

$$\langle a, b\rangle \sim \langle a', b'\rangle \iff a \times b' = a' \times b.$$

As before, this condition $a \times b' = a' \times b$ is intended to encode $\frac{a}{b} = \frac{a'}{b'}$ while never writing down division.

Thus, we define $\mathbb{Q}$ to be the set of equivalence classes $[\langle a, b\rangle]$ of $\sim$ above, as $\langle a, b\rangle$ varies with $a \in \mathbb{Z}$ and $b \in \mathbb{Z} \setminus \{[\langle 0, 0\rangle]\}$. We can now define addition and multiplication by

$$\begin{aligned}
[\langle a, b\rangle] + [\langle a', b'\rangle] &\coloneqq [\langle (a \times b') + (b \times a'), b \times b'\rangle] \\
[\langle a, b\rangle] \times [\langle a', b'\rangle] &\coloneqq [\langle a \times a', b \times b'\rangle].
\end{aligned}$$

Again, the point of these definitions are intended to codify what we want:

$$\begin{aligned}
\frac{a}{b} + \frac{a'}{b'} &= \frac{ab' + ba'}{bb'}, \\
\frac{a}{b} \cdot \frac{a'}{b'} &= \frac{aa'}{bb'}.
\end{aligned}$$

We can check that these are well-defined and so on. ■

Remark 6.33. Lastly, we will identify our old integers $\mathbb{Z}$ canonically with their image in $\mathbb{Q}$ by taking a to $[\langle a, 1\rangle]$.

If we wanted to do real numbers, there are a few ways to do this. One way is to define a real number is an integer $\lfloor x \rfloor$ and then a function $f : \mathbb{N} \to \{0, \dots, 9\}$ to give the decimal expansion. To make sure that our real numbers are all distinct, we need to avoid

$$0.999\ldots = 1.000\ldots,$$

so we need to forbid functions f which are eventually all 9s. We won't be much more rigorous than this because defining addition and multiplication is really annoying (namely, we have to carry).

To codify this properly, we do need to encode functions.

Exercise 6.34. We encode functions as sets.

Proof. The point is that functions $f : X \to Y$ need to give a unique output for given input. A function $f : X \to Y$ is subset of ordered pairs $X \times Y$ such that each $x \in X$ has exactly one $y \in Y$ such that $\langle x, y \rangle \in f$. ■

So we've put most math that we care about into sets. It remains to talk about how to reason with sets.

6.6.2 Set Theory

We close class by making the language of set theory just the language of predicate logic with a single binary relation $\dot\in$. Namely, $x \dot\in y$ is intended to mean that x is an element of y. There are some other relations that we can build from here.

- We write $x \dot\subseteq y$ to mean $\forall z(z \dot\in x \to z \dot\in y)$.
- If we want to talk about ordered pairs, we can write

  $$\forall x \forall y \exists s \forall z (z \doteq s \leftrightarrow (z \doteq x \vee z \doteq y)).$$

  Namely, we are saying that there is a set s containing precisely x and y.
- We can also give unions $x \cup y$ by writing

  $$\forall x \forall y \exists s \forall z (z \dot\in s \leftrightarrow (z \dot\in x \vee z \dot\in y)).$$

- Lastly, we can write $\mathcal{P}(x)$ by writing

  $$\forall x \exists s \forall z (z \dot\in s \leftrightarrow z \dot\subseteq s).$$

That these statements are true are axioms of set theory.

6.7 April 29

Welcome to the last day of class.

6.7.1 Zermelo–Fraenkel Speed-Run

We're going to try to cover a lot of ground, praying for a pedagogical miracle. Last class we brought in our language of set theory; this class we are going to take it further. The only object above propositional logic in our language of set theory is the relation $\dot\in$, which we think of as membership.

To reason about out sets, we are going to require some axioms for our set theory. So let's go through some axioms.

Axiom 6.35 (Existentionality)**.** Two sets are the same if and only if they have the same elements:

$$\forall x \forall y (x \doteq y \leftrightarrow \forall w (w \dot\in x \leftrightarrow w \dot\in y)).$$

Example 6.36. We will think that $\{\varnothing, \{\varnothing\}\} = \{\{\varnothing\}, \varnothing\} = \{\{\varnothing\}, \varnothing, \varnothing\}$.

So we can talk about equality of sets. Next we need sets to talk about. One way we might want to do this is by asserting any property φ can dictate a set by filtering through all elements and check φ.

Example 6.37. We can build the set of all prime numbers by forcing φ to assert that our elements are prime.

Thus, we might want formulae $\varphi(z)$ with a free variable z to dictate sets by writing

$$\exists s \forall z (z \mathrel{\dot\in} s \leftrightarrow \varphi(z)). \tag{$*$}$$

Namely, there is a set s made up of the z which have $\varphi(z)$.

Example 6.38. Let's make ordered pairs. We have

$$\exists s \forall z (z \mathrel{\dot\in} s \leftrightarrow (z \doteq a \vee z \doteq b)).$$

This makes the ordered pair $\{a, b\}$.

Similarly, we can build power sets and so on.

However, ($*$) is a problem because it derives contradiction. The idea is that it is impossible to have a town where a barber shaves exactly the people who do not shave themselves: the barber must either shave themself or not shave themself. So we have the following.

Proposition 6.39. The statement ($*$) derives contradiction.

Proof. Fix $\varphi(z) \coloneqq \neg(z \mathrel{\dot\in} z)$, so comprehension gives

$$\exists s \forall z (z \mathrel{\dot\in} s \leftrightarrow \neg(z \mathrel{\dot\in} z)).$$

But then $s \mathrel{\dot\in} s$ is equivalent to $s \mathrel{\dot\notin} s$, so we have our contradiction. Formally, we have the following.

1	$\exists s \forall z (z \mathrel{\dot\in} s \leftrightarrow \neg(z \mathrel{\dot\in} z))$	
2	$\forall z (z \mathrel{\dot\in} r \leftrightarrow \neg(z \mathrel{\dot\in} z))$ $\boxed{r}$	
3	$r \mathrel{\dot\in} r \leftrightarrow \neg(r \mathrel{\dot\in} r)$	$\forall$Elim, 2
4	$r \mathrel{\dot\in} r$	
5	$r \mathrel{\dot\in} r \leftrightarrow \neg(r \mathrel{\dot\in} r)$	Reit, 3
6	$\neg(r \mathrel{\dot\in} r)$	$\leftrightarrow$Elim, 4, 5
7	$(r \mathrel{\dot\in} r) \wedge \neg(r \mathrel{\dot\in} r)$	$\wedge$Intro, 6
8	$\neg(r \mathrel{\dot\in} r)$	$\neg$Intro, 4–7
9	$r \mathrel{\dot\in} r$	$\leftrightarrow$Elim, 3, 8
10	$(r \mathrel{\dot\in} r) \wedge \neg(r \mathrel{\dot\notin} r)$	$\wedge$Intro, 8, 9
11	$\neg\exists s \forall z (z \mathrel{\dot\in} s \leftrightarrow \neg(z \mathrel{\dot\in} z))$	

Thus, we have derived the negation of ($*$) in this case. ∎

The point is that we need to be more careful about how we make our sets: not all properties are okay. Namely, we want to restrict our φ in $(*)$. So here are some sets we'll make.

Axiom 6.40 (Empty set)**.** There is an empty set: $\exists s \forall x (\neg (x \mathrel{\dot\in} s))$.

Axiom 6.41 (Unordered pair)**.** There is an ordered pair: $\forall x \forall y \exists s \forall z (z \mathrel{\dot\in} s \leftrightarrow (z \doteq x \wedge z \doteq y))$.

Axiom 6.42 (Union)**.** We can take the union of some set of sets:

$$\forall x \exists s \forall z (z \mathrel{\dot\in} s \leftrightarrow \exists y (y \mathrel{\dot\in} x \wedge z \mathrel{\dot\in} y)).$$

Here y is to be thought of as "looping" through the various sets within x.

Axiom 6.43 (Power set)**.** We can build power sets by writing

$$\forall x \exists s \forall z (z \mathrel{\dot\in} s \leftrightarrow z \mathrel{\dot\subseteq} s).$$

Namely, s contains all the various subsets z of x.

Axiom 6.44 (Separation)**.** Given a set x, we can filter out a subset from a property: given a formula φ with a free variable z (but s is not free in z),

$$\forall x \exists s \forall z (z \mathrel{\dot\in} s \leftrightarrow (z \mathrel{\dot\in} s \wedge \varphi(z))).$$

We will also want to make an infinite set.

Axiom 6.45 (Infinity)**.** There is an infinite set, closed under successor: $\exists s (\varnothing \mathrel{\dot\in} s \wedge \forall x (x \mathrel{\dot\in} s \to x \cup \{x\} \mathrel{\dot\in} s))$.

These are most of the axioms in set theory, with a few others. They make up Zermelo–Fraenkel set theory, and they can create a foundation for all current mathematics.

6.7.2 Sizes of Infinity

Zermelo–Fraenkel set theory is going to permit infinities of different sizes in a non-paradoxical way. The key point is that the definition of "size" is fairly technical.

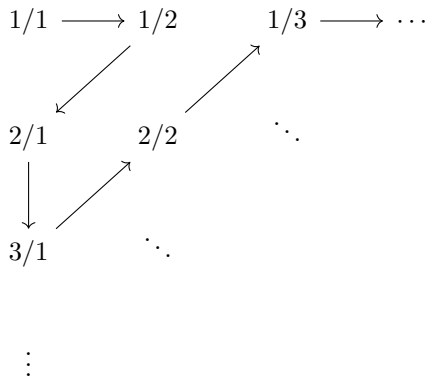
Definition 6.46 (Bijection)**.** Given two sets X and Y, a bijection $f \colon X \to Y$ is an onto, one-to-one function between X and Y.

The point is that bijections should assert that sets have the same size without having to enumerate all the elements.

Definition 6.47 (Equinumerous)**.** Two sets X and Y are *equinumerous* if and only if there is a bijection between them. We might also say that they have the "same size."

Remark 6.48. If $X \subsetneq Y$ and X and Y are finite, then X and Y have different sizes. However, $\mathbb{N}$ and $2\mathbb{N}$ have the same size because we can biject $\mathbb{N} \to 2\mathbb{N}$ by sending $n \mapsto 2n$.

One can even put $\mathbb{Q}$ and $\mathbb{N}$ into bijection, by moving diagonally and across through the following table.



However, it turns out that there are more real numbers than natural numbers.

Proposition 6.49. There are strictly more real numbers than natural numbers.

Proof. We show that any injective function $f\colon \mathbb{N} \to \mathbb{R}$ is not surjective, thus implying there is no bijection. Now, write out the decimal expansion of any $d(n)$ for $n \in \mathbb{N}$ by

$$f(n) \coloneqq a_n.d_{n,0}d_{n,1}d_{n,2}\dots.$$

As such, we construct a real number as having decimal expansion given by

$$d_n \coloneqq \begin{cases} 4 & d_{n,n} = 5, \\ 5 & d_{n,n} \neq 4. \end{cases}$$

Then the number $\alpha \coloneqq 0.d_0d_1d_2\dots$ is not in the image of f because α and $f(n)$ differ in the nth decimal digit, implying $f(n) \neq \alpha$ for each $n \in \mathbb{N}$. This finishes. ∎

Remark 6.50. From here, one can build larger and larger infinities. However, we cannot tell the difference between these larger infinities. For example, the Continuum hypothesis (which asks if there is a set X of size strictly between $\mathbb{N}$ and $\mathbb{R}$) cannot be settled by Zermelo–Fraenkel set theory. Some questions must be left unanswered.

List of Definitions