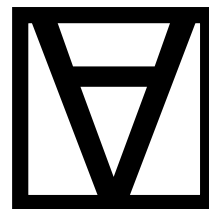


Logic II

Edited by Richard Zach



Logic II by Richard Zach is licensed under a Creative Commons Attribution 4.0 International License. It is based on *The Open Logic Text* by the Open Logic Project, used under a Creative Commons Attribution 4.0 International License.



Contents

I	Sets, Relations, Functions	1
1	Sets	3
1.1	Basics	3
1.2	Some Important Sets	4
1.3	Subsets	4
1.4	Unions and Intersections	5
1.5	Proofs about Sets	6
1.6	Pairs, Tuples, Cartesian Products	7
2	Relations	9
2.1	Relations as Sets	9
2.2	Special Properties of Relations	10
2.3	Orders	11
2.4	Operations on Relations	12
3	Functions	15
3.1	Basics	15
3.2	Kinds of Functions	16
3.3	Operations on Functions	17
3.4	Isomorphism	17
3.5	Partial Functions	18
4	The Size of Sets	19
4.1	Introduction	19
4.2	Countable Sets	19
4.3	Uncountable Sets	22
4.4	Reduction	23
4.5	Equinumerous Sets	24
4.6	Comparing Sizes of Sets	25

II	First-order Logic	27
5	Syntax and Semantics	29
5.1	First-Order Languages	29
5.2	Terms and Formulas	31
5.3	Unique Readability	33
5.4	Main operator of a Formula	35
5.5	Subformulas	36
5.6	Free Variables and Sentences	37
5.7	Substitution	38
5.8	Structures for First-order Languages	39
5.9	Satisfaction of a Formula in a Structure	42
5.10	Extensionality	45
5.11	Semantic Notions	46
6	Theories and Their Models	49
6.1	Introduction	49
6.2	Expressing Properties of Structures	51
6.3	Examples of First-Order Theories	51
6.4	Expressing Relations in a Structure	54
6.5	The Theory of Sets	55
6.6	Expressing the Size of Structures	57
7	Natural Deduction	59
7.1	Rules and Derivations	59
7.2	Examples of Derivations	61
7.3	Proof-Theoretic Notions	65
7.4	Properties of Derivability	66
7.5	Soundness	68
7.6	Derivations with Identity predicate	71
8	The Completeness Theorem	73
8.1	Introduction	73
8.2	Outline of the Proof	73
8.3	Maximally Consistent Sets of Sentences	75
8.4	Henkin Expansion	77
8.5	Lindenbaum's Lemma	78
8.6	Construction of a Model	78
8.7	Identity	80
8.8	The Completeness Theorem	82
8.9	The Compactness Theorem	82
8.10	The Löwenheim-Skolem Theorem	83
9	Beyond First-order Logic	85

9.1	Overview	85
9.2	Many-Sorted Logic	86
9.3	Second-Order logic	87
9.4	Higher-Order logic	91
9.5	Intuitionistic logic	93
9.6	Modal Logics	97
9.7	Other Logics	98
III Turing Machines		101
10	Turing Machine Computations	103
10.1	Introduction	103
10.2	Turing Machines	103
10.3	Configurations and Computations	104
10.4	Unary Representation of Numbers	105
11	Undecidability	107
11.1	Decision Problems	107
11.2	Representing Turing Machines	108
11.3	Verifying the Representation	110
Problems		113

Part I

Sets, Relations, Functions

Chapter 1

Sets

1.1 Basics

Sets are the most fundamental building blocks of mathematical objects. In fact, almost every mathematical object can be seen as a set of some kind. In logic, as in other parts of mathematics, sets and set theoretical talk is ubiquitous. So it will be important to discuss what sets are, and introduce the notations necessary to talk about sets and operations on sets in a standard way.

Definition 1.1. A *set* is a collection of objects, considered independently of the way it is specified, of the order of its elements, or of their multiplicity. The objects making up the set are called *elements* or *members* of the set. If a is an element of a set X , we write $a \in X$ (otherwise, $a \notin X$). The set which has no elements is called the empty set and denoted \emptyset .

Example 1.2. Whenever you have a bunch of objects, you can collect them together in a set. The set of Richard's siblings, for instance, is a set that contains one person, and we could write it as $S = \{\text{Ruth}\}$. In general, when we have some objects a_1, \dots, a_n , then the set consisting of exactly those objects is written $\{a_1, \dots, a_n\}$. Frequently we'll specify a set by some property that its elements share—as we just did, for instance, by specifying S as the set of Richard's siblings. We'll use the following shorthand notation for that: $\{x : \dots x \dots\}$, where the $\dots x \dots$ stands for the property that x has to have in order to be counted among the elements of the set. In our example, we could have specified S also as $S = \{x : x \text{ is a sibling of Richard}\}$.

When we say that sets are independent of the way they are specified, we mean that the elements of a set are all that matters. For instance, it so happens that $\{\text{Nicole, Jacob}\}$, $\{x : \text{is a niece or nephew of Richard}\}$ and $\{x : \text{is a child of Ruth}\}$ are three ways of specifying one and the same set.

Saying that sets are considered independently of the order of their elements and their multiplicity is a fancy way of saying that $\{\text{Nicole, Jacob}\}$ and

$\{\text{Jacob, Nicole}\}$ are two ways of specifying the same set; and that $\{\text{Nicole, Jacob}\}$ and $\{\text{Jacob, Nicole, Nicole}\}$ are two ways of specifying the same set.

1.2 Some Important Sets

Example 1.3. Mostly we'll be dealing with sets that have mathematical objects as members. You will remember the various sets of numbers: \mathbb{N} is the set of *natural* numbers $\{0, 1, 2, 3, \dots\}$; \mathbb{Z} the set of *integers* $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$; \mathbb{Q} the set of *rationals* ($\mathbb{Q} = \{z/n : z \in \mathbb{Z}, n \in \mathbb{N}, n \neq 0\}$); and \mathbb{R} the set of *real* numbers. These are all *infinite* sets, that is, they each have infinitely many elements. As it turns out, $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ have the same number of elements, while \mathbb{R} has a whole bunch more— $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ are “countable and infinite” whereas \mathbb{R} is “uncountable”.

We'll sometimes also use the set of positive integers $\mathbb{Z}^+ = \{1, 2, 3, \dots\}$ and the set containing just the first two natural numbers $\mathbb{B} = \{0, 1\}$.

Example 1.4 (Strings). Another interesting example is the set A^* of *finite strings* over an alphabet A : any finite sequence of elements of A is a string over A . We include the *empty string* Λ among the strings over A , for every alphabet A . For instance,

$$\mathbb{B}^* = \{\Lambda, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, 101, 110, 111, 0000, \dots\}.$$

If $x = x_1 \dots x_n \in A^*$ is a string consisting of n “letters” from A , then we say *length* of the string is n and write $\text{len}(x) = n$.

Example 1.5 (Infinite sequences). For any set A we may also consider the set A^ω of infinite sequences of elements of A . An infinite sequence $a_1 a_2 a_3 a_4 \dots$ consists of a one-way infinite list of objects, each one of which is an element of A .

1.3 Subsets

Sets are made up of their elements, and every element of a set is a part of that set. But there is also a sense that some of the elements of a set *taken together* are a “part of” that set. For instance, the number 2 is part of the set of integers, but the set of even numbers is also a part of the set of integers. It's important to keep those two senses of being part of a set separate.

Definition 1.6. If every element of a set X is also an element of Y , then we say that X is a *subset* of Y , and write $X \subseteq Y$.

Example 1.7. First of all, every set is a subset of itself, and \emptyset is a subset of every set. The set of even numbers is a subset of the set of natural numbers. Also, $\{a, b\} \subseteq \{a, b, c\}$.

But $\{a, b, e\}$ is not a subset of $\{a, b, c\}$.

Note that a set may contain other sets! In particular, a set may happen to *both* be an element and a subset of another, e.g., $\{0\} \in \{0, \{0\}\}$ and also $\{0\} \subseteq \{0, \{0\}\}$.

Definition 1.8. The set consisting of all subsets of a set X is called the *power set* of X , written $\wp(X)$.

$$\wp(X) = \{x : x \subseteq X\}$$

Example 1.9. What are all the possible subsets of $\{a, b, c\}$? They are: \emptyset , $\{a\}$, $\{b\}$, $\{c\}$, $\{a, b\}$, $\{a, c\}$, $\{b, c\}$, $\{a, b, c\}$. The set of all these subsets is $\wp(\{a, b, c\})$:

$$\wp(\{a, b, c\}) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$$

1.4 Unions and Intersections

Definition 1.10. The *union* of two sets X and Y , written $X \cup Y$, is the set of all things which are members of X , Y , or both.

$$X \cup Y = \{x : x \in X \vee x \in Y\}$$

Example 1.11. Since the multiplicity of elements doesn't matter, the union of two sets which have an element in common contains that element only once, e.g., $\{a, b, c\} \cup \{a, 0, 1\} = \{a, b, c, 0, 1\}$.

The union of a set and one of its subsets is just the bigger set: $\{a, b, c\} \cup \{a\} = \{a, b, c\}$.

The union of a set with the empty set is identical to the set: $\{a, b, c\} \cup \emptyset = \{a, b, c\}$.

Definition 1.12. The *intersection* of two sets X and Y , written $X \cap Y$, is the set of all things which are elements of both X and Y .

$$X \cap Y = \{x : x \in X \wedge x \in Y\}$$

Two sets are called *disjoint* if their intersection is empty. This means they have no elements in common.

Example 1.13. If two sets have no elements in common, their intersection is empty: $\{a, b, c\} \cap \{0, 1\} = \emptyset$.

If two sets do have elements in common, their intersection is the set of all those: $\{a, b, c\} \cap \{a, b, d\} = \{a, b\}$.

The intersection of a set with one of its subsets is just the smaller set: $\{a, b, c\} \cap \{a, b\} = \{a, b\}$.

The intersection of any set with the empty set is empty: $\{a, b, c\} \cap \emptyset = \emptyset$.

1. SETS

We can also form the union or intersection of more than two sets. An elegant way of dealing with this in general is the following: suppose you collect all the sets you want to form the union (or intersection) of into a single set. Then we can define the union of all our original sets as the set of all objects which belong to at least one element of the set, and the intersection as the set of all objects which belong to every element of the set.

Definition 1.14. If C is a set of sets, then $\bigcup C$ is the set of elements of elements of C :

$$\begin{aligned}\bigcup C &= \{x : x \text{ belongs to an element of } C\}, \text{ i.e.,} \\ \bigcup C &= \{x : \text{there is a } y \in C \text{ so that } x \in y\}\end{aligned}$$

Definition 1.15. If C is a set of sets, then $\bigcap C$ is the set of objects which all elements of C have in common:

$$\bigcap C = \{x : x \text{ belongs to every element of } C\}, \text{ i.e., } \bigcap C = \{x : \text{for all } y \in C, x \in y\}$$

Example 1.16. Suppose $C = \{\{a, b\}, \{a, d, e\}, \{a, d\}\}$. Then $\bigcup C = \{a, b, d, e\}$ and $\bigcap C = \{a\}$.

We could also do the same for a sequence of sets A_1, A_2, \dots

$$\begin{aligned}\bigcup_i A_i &= \{x : x \text{ belongs to one of the } A_i\} \\ \bigcap_i A_i &= \{x : x \text{ belongs to every } A_i\}.\end{aligned}$$

Definition 1.17. The *difference* $X \setminus Y$ is the set of all elements of X which are not also elements of Y , i.e.,

$$X \setminus Y = \{x : x \in X \text{ and } x \notin Y\}.$$

1.5 Proofs about Sets

Sets and the notations we've introduced so far provide us with convenient shorthands for specifying sets and expressing relationships between them. Often it will also be necessary to prove claims about such relationships. If you're not familiar with mathematical proofs, this may be new to you. So we'll walk through a simple example. We'll prove that for any sets X and Y , it's always the case that $X \cap (X \cup Y) = X$. How do you prove an identity between sets like this? Recall that sets are determined solely by their elements, i.e., sets are identical iff they have the same elements. So in this case we have to prove that (a) every element of $X \cap (X \cup Y)$ is also an element of X and, conversely, that (b) every element of X is also an element of $X \cap (X \cup Y)$. In other words, we show that both (a) $X \cap (X \cup Y) \subseteq X$ and (b) $X \subseteq X \cap (X \cup Y)$.

A proof of a general claim like “every element z of $X \cap (X \cup Y)$ is also an element of X ” is proved by first assuming that an arbitrary $z \in X \cap (X \cup Y)$ is given, and proving from this assumption that $z \in X$. You may know this pattern as “general conditional proof.” In this proof we’ll also have to make use of the definitions involved in the assumption and conclusion, e.g., in this case of “ \cap ” and “ \cup .” So case (a) would be argued as follows:

(a) We first want to show that $X \cap (X \cup Y) \subseteq X$, i.e., by definition of \subseteq , that if $z \in X \cap (X \cup Y)$ then $z \in X$, for any z . So assume that $z \in X \cap (X \cup Y)$. Since z is an element of the intersection of two sets iff it is an element of both sets, we can conclude that $z \in X$ and also $z \in X \cup Y$. In particular, $z \in X$. But this is what we wanted to show.

This completes the first half of the proof. Note that in the last step we used the fact that if a conjunction ($z \in X$ and $z \in X \cup Y$) follows from an assumption, each conjunct follows from that same assumption. You may know this rule as “conjunction elimination,” or \wedge Elim. Now let’s prove (b):

(b) We now prove that $X \subseteq X \cap (X \cup Y)$, i.e., by definition of \subseteq , that if $z \in X$ then also $z \in X \cap (X \cup Y)$, for any z . Assume $z \in X$. To show that $z \in X \cap (X \cup Y)$, we have to show (by definition of “ \cap ”) that (i) $z \in X$ and also (ii) $z \in X \cup Y$. Here (i) is just our assumption, so there is nothing further to prove. For (ii), recall that z is an element of a union of sets iff it is an element of at least one of those sets. Since $z \in X$, and $X \cup Y$ is the union of X and Y , this is the case here. So $z \in X \cup Y$. We’ve shown both (i) $z \in X$ and (ii) $z \in X \cup Y$, hence, by definition of “ \cap ,” $z \in X \cap (X \cup Y)$.

This was somewhat long-winded, but it illustrates how we reason about sets and their relationships. We usually aren’t this explicit; in particular, we might not repeat all the definitions. A “textbook” proof of our result would look something like this.

Proposition 1.18 (Absorption). *For all sets X, Y ,*

$$X \cap (X \cup Y) = X$$

Proof. (a) Suppose $z \in X \cap (X \cup Y)$. Then $z \in X$, so $X \cap (X \cup Y) \subseteq X$.

(b) Now suppose $z \in X$. Then also $z \in X \cup Y$, and therefore also $z \in X \cap (X \cup Y)$. Thus, $X \subseteq X \cap (X \cup Y)$. \square

1.6 Pairs, Tuples, Cartesian Products

Sets have no order to their elements. We just think of them as an unordered collection. So if we want to represent order, we use *ordered pairs* $\langle x, y \rangle$, or more generally, *ordered n -tuples* $\langle x_1, \dots, x_n \rangle$.

1. SETS

Definition 1.19. Given sets X and Y , their *Cartesian product* $X \times Y$ is $\{\langle x, y \rangle : x \in X \text{ and } y \in Y\}$.

Example 1.20. If $X = \{0, 1\}$, and $Y = \{1, a, b\}$, then their product is

$$X \times Y = \{\langle 0, 1 \rangle, \langle 0, a \rangle, \langle 0, b \rangle, \langle 1, 1 \rangle, \langle 1, a \rangle, \langle 1, b \rangle\}.$$

Example 1.21. If X is a set, the product of X with itself, $X \times X$, is also written X^2 . It is the set of *all* pairs $\langle x, y \rangle$ with $x, y \in X$.

Chapter 2

Relations

2.1 Relations as Sets

You will no doubt remember some interesting relations between objects of some of the sets we've mentioned. For instance, numbers come with an *order relation* $<$ and from the theory of whole numbers the relation of *divisibility without remainder* (usually written $n \mid m$) may be familiar. There is also the relation *is identical with* that every object bears to itself and to no other thing. But there are many more interesting relations that we'll encounter, and even more possible relations. Before we review them, we'll just point out that we can look at relations as a special sort of set. For this, first recall what a *pair* is: if a and b are two objects, we can combine them into the *ordered pair* $\langle a, b \rangle$. Note that for ordered pairs the order *does* matter, e.g. $\langle a, b \rangle \neq \langle b, a \rangle$, in contrast to unordered pairs, i.e., 2-element sets, where $\{a, b\} = \{b, a\}$.

If X and Y are sets, then the *Cartesian product* $X \times Y$ of X and Y is the set of all pairs $\langle a, b \rangle$ with $a \in X$ and $b \in Y$. In particular, $X^2 = X \times X$ is the set of all pairs from X .

Now consider a relation on a set, e.g., the $<$ -relation on the set \mathbb{N} of natural numbers, and consider the set of all pairs of numbers $\langle n, m \rangle$ where $n < m$, i.e.,

$$R = \{\langle n, m \rangle : n, m \in \mathbb{N} \text{ and } n < m\}.$$

Then there is a close connection between the number n being less than a number m and the corresponding pair $\langle n, m \rangle$ being a member of R , namely, $n < m$ if and only if $\langle n, m \rangle \in R$. In a sense we can consider the set R to be the $<$ -relation on the set \mathbb{N} . In the same way we can construct a subset of \mathbb{N}^2 for any relation between numbers. Conversely, given any set of pairs of numbers $S \subseteq \mathbb{N}^2$, there is a corresponding relation between numbers, namely, the relationship n bears to m if and only if $\langle n, m \rangle \in S$. This justifies the following definition:

2. RELATIONS

Definition 2.1. A *binary relation* on a set X is a subset of X^2 . If $R \subseteq X^2$ is a binary relation on X and $x, y \in X$, we write Rxy (or xRy) for $\langle x, y \rangle \in R$.

Example 2.2. The set \mathbb{N}^2 of pairs of natural numbers can be listed in a 2-dimensional matrix like this:

$$\begin{array}{cccccc} \langle \mathbf{0}, \mathbf{0} \rangle & \langle 0, 1 \rangle & \langle 0, 2 \rangle & \langle 0, 3 \rangle & \dots & \\ \langle 1, 0 \rangle & \langle \mathbf{1}, \mathbf{1} \rangle & \langle 1, 2 \rangle & \langle 1, 3 \rangle & \dots & \\ \langle 2, 0 \rangle & \langle 2, 1 \rangle & \langle \mathbf{2}, \mathbf{2} \rangle & \langle 2, 3 \rangle & \dots & \\ \langle 3, 0 \rangle & \langle 3, 1 \rangle & \langle 3, 2 \rangle & \langle \mathbf{3}, \mathbf{3} \rangle & \dots & \\ \vdots & \vdots & \vdots & \vdots & \ddots & \end{array}$$

The subset consisting of the pairs lying on the diagonal, $\{\langle 0, 0 \rangle, \langle 1, 1 \rangle, \langle 2, 2 \rangle, \dots\}$, is the *identity relation* on \mathbb{N} . (Since the identity relation is popular, let's define $\text{Id}_X = \{\langle x, x \rangle : x \in X\}$ for any set X .) The subset of all pairs lying above the diagonal, $L = \{\langle 0, 1 \rangle, \langle 0, 2 \rangle, \dots, \langle 1, 2 \rangle, \langle 1, 3 \rangle, \dots, \langle 2, 3 \rangle, \langle 2, 4 \rangle, \dots\}$ is the *less than* relation, i.e., Lnm iff $n < m$. The subset of pairs below the diagonal, $G = \{\langle 1, 0 \rangle, \langle 2, 0 \rangle, \langle 2, 1 \rangle, \langle 3, 0 \rangle, \langle 3, 1 \rangle, \langle 3, 2 \rangle, \dots\}$ is the *greater than* relation, i.e., Gnm iff $n > m$. The union of L with I , $K = L \cup I$, is the *less than or equal to* relation: Knm iff $n \leq m$. Similarly, $H = G \cup I$ is the *greater than or equal to* relation. L , G , K , and H are special kinds of relations called *orders*. L and G have the property that no number bears L or G to itself (i.e., for all n , neither Lnn nor Gnn). Relations with this property are called *antireflexive*, and, if they also happen to be orders, they are called *strict orders*.

Although orders and identity are important and natural relations, it should be emphasized that according to our definition *any* subset of X^2 is a relation on X , regardless of how unnatural or contrived it seems. In particular, \emptyset is a relation on any set (the *empty relation*, which no pair of elements bears), and X^2 itself is a relation on X as well (one which every pair bears). But also something like $E = \{\langle n, m \rangle : n > 5 \text{ or } m \times n \geq 34\}$ counts as a relation.

2.2 Special Properties of Relations

Some kinds of relations turn out to be so common that they have been given special names. For instance, \leq and \subseteq both relate their respective domains (say, \mathbb{N} in the case of \leq and $\wp(X)$ in the case of \subseteq) in similar ways. To get at exactly how these relations are similar, and how they differ, we categorize them according to some special properties that relations can have. It turns out that (combinations of) some of these special properties are especially important: orders and equivalence relations.

Definition 2.3. A relation $R \subseteq X^2$ is *reflexive* iff, for every $x \in X$, Rxx .

Definition 2.4. A relation $R \subseteq X^2$ is *transitive* iff, whenever Rxy and Ryz , then also Rxz .

Definition 2.5. A relation $R \subseteq X^2$ is *symmetric* iff, whenever Rxy , then also Ryx .

Definition 2.6. A relation $R \subseteq X^2$ is *anti-symmetric* iff, whenever both Rxy and Ryx , then $x = y$ (or, in other words: if $x \neq y$ then either $\neg Rxy$ or $\neg Ryx$).

In a symmetric relation, Rxy and Ryx always hold together, or neither holds. In an anti-symmetric relation, the only way for Rxy and Ryx to hold together is if $x = y$. Note that this does not *require* that Rxy and Ryx holds when $x = y$, only that it isn't ruled out. So an anti-symmetric relation can be reflexive, but it is not the case that every anti-symmetric relation is reflexive. Also note that being anti-symmetric and merely not being symmetric are different conditions. In fact, a relation can be both symmetric and anti-symmetric at the same time (e.g., the identity relation is).

Definition 2.7. A relation $R \subseteq X^2$ is *connected* if for all $x, y \in X$, if $x \neq y$, then either Rxy or Ryx .

Definition 2.8. A relation $R \subseteq X^2$ that is reflexive, transitive, and anti-symmetric is called a *partial order*. A partial order that is also connected is called a *linear order*.

Definition 2.9. A relation $R \subseteq X^2$ that is reflexive, symmetric, and transitive is called an *equivalence relation*.

2.3 Orders

Definition 2.10. A relation which is both reflexive and transitive is called a *preorder*. A preorder which is also anti-symmetric is called a *partial order*. A partial order which is also connected is called a *total order* or *linear order*. (If we want to emphasize that the order is reflexive, we add the adjective “weak”—see below).

Example 2.11. Every linear order is also a partial order, and every partial order is also a preorder, but the converses don't hold. For instance, the identity relation and the full relation on X are preorders, but they are not partial orders, because they are not anti-symmetric (if X has more than one element). For a somewhat less silly example, consider the *no longer than* relation \preceq on \mathbb{B}^* : $x \preceq y$ iff $\text{len}(x) \leq \text{len}(y)$. This is a preorder, even a linear preorder, but not a partial order.

The relation of *divisibility without remainder* gives us an example of a partial order which isn't a linear order: for integers n, m , we say n (evenly) divides m , in symbols: $n \mid m$, if there is some k so that $m = kn$. On \mathbb{N} , this is a partial order, but not a linear order: for instance, $2 \nmid 3$ and also $3 \nmid 2$. Considered as a relation on \mathbb{Z} , divisibility is only a preorder since anti-symmetry fails: $1 \mid -1$ and $-1 \mid 1$ but $1 \neq -1$. Another important partial order is the relation \subseteq on a set of sets.

2. RELATIONS

Notice that the examples L and G from [Example 2.2](#), although we said there that they were called “strict orders” are not linear orders even though they are connected (they are not reflexive). But there is a close connection, as we will see momentarily.

Definition 2.12. A relation R on X is called *irreflexive* if, for all $x \in X$, $\neg Rxx$. R is called *asymmetric* if for no pair $x, y \in X$ we have Rxy and Ryx . A *strict partial order* is a relation which is irreflexive, asymmetric, and transitive. A strict partial order which is also connected is called a *strict linear order*.

A strict partial order R on X can be turned into a weak partial order R' by adding the identity relation on X : $R' = R \cup \text{Id}_X$. Conversely, starting from a weak partial order, one can get a strict partial order by removing Id_X , i.e., $R' = R \setminus \text{Id}_X$.

Proposition 2.13. R is a strict partial (linear) order on X iff R' is a weak partial (linear) order. Moreover, Rxy iff $R'xy$ for all $x \neq y$.

Example 2.14. \leq is the weak linear order corresponding to the strict linear order $<$. \subseteq is the weak partial order corresponding to the strict partial order \subsetneq .

2.4 Operations on Relations

It is often useful to modify or combine relations. We’ve already used the union of relations above (which is just the union of two relations considered as sets of pairs). Here are some other ways:

Definition 2.15. Let $R, S \subseteq X^2$ be relations and Y a set.

1. The *inverse* R^{-1} of R is $R^{-1} = \{\langle y, x \rangle : \langle x, y \rangle \in R\}$.
2. The *relative product* $R \mid S$ of R and S is

$$(R \mid S) = \{\langle x, z \rangle : \text{for some } y, Rxy \text{ and } Syz\}$$

3. The *restriction* $R \upharpoonright Y$ of R to Y is $R \cap Y^2$
4. The *application* $R[Y]$ of R to Y is

$$R[Y] = \{y : \text{for some } x \in X, Rxy\}$$

Example 2.16. Let $S \subseteq \mathbb{Z}^2$ be the successor relation on \mathbb{Z} , i.e., the set of pairs $\langle x, y \rangle$ where $x + 1 = y$, for $x, y \in \mathbb{Z}$. Sxy holds iff y is the successor of x .

1. The inverse S^{-1} of S is the predecessor relation, i.e., $S^{-1}xy$ iff $x - 1 = y$.
2. The relative product $S \mid S$ is the relation x bears to y if $x + 2 = y$.

3. The restriction of S to \mathbb{N} is the successor relation on \mathbb{N} .

4. The application of S to a set, e.g., $S[\{1, 2, 3\}]$ is $\{2, 3, 4\}$.

Definition 2.17. The *transitive closure* R^+ of a relation $R \subseteq X^2$ is $R^+ = \bigcup_{i=1}^{\infty} R^i$ where $R^1 = R$ and $R^{i+1} = R^i \mid R$.

The *reflexive transitive closure* of R is $R^* = R^+ \cup I_X$.

Example 2.18. Take the successor relation $S \subseteq \mathbb{Z}^2$. S^2xy iff $x + 2 = y$, S^3xy iff $x + 3 = y$, etc. So R^*xy iff for some $i \geq 1$, $x + i = y$. In other words, S^+xy iff $x < y$ (and R^*xy iff $x \leq y$).

Chapter 3

Functions

3.1 Basics

A function is a relation in which each object is related to a unique partner. Many functions are familiar to us from basic arithmetic. For instance, addition and multiplication are functions. They take in two numbers and return a third. A function, more generally, is something that takes one or more things as input and returns some kind of output. A function is a *black box*: what matters is only what output is paired with what input, not the method for calculating the output.

Definition 3.1. A function $f: X \rightarrow Y$ is a mapping of each element of X to an element of Y . We call X the *domain* of f and Y the *codomain* of f . The *range* of f is the subset of the codomain that is actually output by f for some input.

Example 3.2. Multiplication goes from $\mathbb{N} \times \mathbb{N}$ (the domain) to \mathbb{N} (the codomain). As it turns out, the range is also \mathbb{N} , since every $n \in \mathbb{N}$ is $n \times 1$.

Multiplication is a function because it pairs each input—each pair of natural numbers—with a single output: $\times: \mathbb{N}^2 \rightarrow \mathbb{N}$. In contrast, the square root operation applied to the domain \mathbb{N} is not functional, since each positive integer n has two square roots: \sqrt{n} and $-\sqrt{n}$. We can make it functional by only returning the positive square root: $\sqrt{\cdot}: \mathbb{N} \rightarrow \mathbb{R}$. The relation that pairs each student in a class with their final grade is a function—no student can get two different final grades in the same class. The relation that pairs each student in a class with their parents is not a function—generally each student will have at least two parents.

Example 3.3. Let $f: \mathbb{N} \rightarrow \mathbb{N}$ be defined such that $f(x) = x + 1$. This tells us that f is a function which takes in natural numbers and outputs natural numbers. It then tells us that, given a natural number, f will output its successor.

3. FUNCTIONS

In this case, the codomain \mathbb{N} is not the range of f , since the natural number 0 is not the successor of any natural number. The range of f is the set of all positive integers, \mathbb{Z}^+ .

Example 3.4. Let $g: \mathbb{N} \rightarrow \mathbb{N}$ be defined such that $g(x) = x - 1 + 2$. This tells us that g is a function which takes in natural numbers and outputs natural numbers. It then tells us that, given a natural number, g will output the successor of the successor of its predecessor. Despite their different definitions, g and f are the same function.

Functions f and g defined above are the same because for any natural number x , $x - 1 + 2 = x + 1$. f and g pair each natural number with the same output. Functions, just like relations more generally, can be treated as just sets of pairs. The definitions for f and g specify the same set by means of different equations, and as we know, sets are independent of how they are specified.

Example 3.5. We can also define functions by cases. For instance, we could define $f: \mathbb{N} \rightarrow \mathbb{N}$ by

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ \frac{x+1}{2} & \text{if } x \text{ is odd.} \end{cases}$$

This is fine, since every natural number is either even or odd, and the output of this function will always be a natural number. Just remember that if you define a function by cases, every possible input must fall into exactly one case.

3.2 Kinds of Functions

Definition 3.6. A function $f: X \rightarrow Y$ is *surjective* iff Y is also the range of f , i.e., for every $y \in Y$ there is at least one $x \in X$ such that $f(x) = y$.

If you want to show that a function is surjective, then you need to show that every object in the codomain is the output of the function given some input or other.

Definition 3.7. A function $f: X \rightarrow Y$ is *injective* iff for each $y \in Y$ there is at most one $x \in X$ such that $f(x) = y$.

Any function pairs each input with a unique output. An injective function has a unique input for each possible output. If you want to show that a function f is injective, you need to show that for any element y of the codomain, if $f(x) = y$ and $f(w) = y$, then $x = w$.

A function which is neither injective, nor surjective, is the constant function $f: \mathbb{N} \rightarrow \mathbb{N}$ where $f(x) = 1$.

A function which is both injective and surjective is the identity function $f: \mathbb{N} \rightarrow \mathbb{N}$ where $f(x) = x$.

The successor function $f: \mathbb{N} \rightarrow \mathbb{N}$ where $f(x) = x + 1$ is injective, but not surjective.

The function

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } x \text{ is even} \\ \frac{x+1}{2} & \text{if } x \text{ is odd.} \end{cases}$$

is surjective, but not injective.

Definition 3.8. A function $f: X \rightarrow Y$ is *bijective* iff it is both surjective and injective. We call such a function a *bijection* from X to Y (or between X and Y).

3.3 Operations on Functions

We've already seen that the inverse f^{-1} of a bijective function f is itself a function. It is also possible to compose functions.

Definition 3.9. Let $f: X \rightarrow Y$ and $g: Y \rightarrow Z$. Then we *compose* g and f to form the function $(g \circ f): X \rightarrow Z$, where $(g \circ f)(x) = g(f(x))$.

The function $(g \circ f): X \rightarrow Z$ pairs each member of X with a member of Z . We specify which member of Z a member of X is paired with as follows—given an input $x \in X$, first apply the function f to x , which will output some $y \in Y$. Then apply the function g to y , which will output some $z \in Z$.

Example 3.10. Consider the functions $f(x) = x + 1$, and $g(x) = 2x$. What function do you get when you compose these two? $(g \circ f)(x) = g(f(x))$. So that means for every natural number you give this function, you first add one, and then you multiply the result by two. So their composition is $(g \circ f)(x) = 2(x + 1)$.

3.4 Isomorphism

An *isomorphism* is a bijection that preserves the structure of the sets it relates, where structure is a matter of the relationships that obtain between the members of the sets. Consider the following two sets $X = \{1, 2, 3\}$ and $Y = \{4, 5, 6\}$. These sets are both structured by the relations successor, less than and greater than. An isomorphism between the two sets is a bijection that preserves those structures. So a function $f: X \rightarrow Y$ is an isomorphism if, among other things, $i < j$ iff $f(i) < f(j)$, and j is the successor of i iff $f(j)$ is the successor of $f(i)$.

Definition 3.11. Let U be the pair $\langle X, R \rangle$ and V be the pair $\langle Y, S \rangle$ such that X and Y are sets and R and S are relations on X and Y respectively. A bijection

f from X to Y is an *isomorphism* from U to V iff it preserves the relational structure, that is, for any x_1 and x_2 in X , $\langle x_1, x_2 \rangle \in R$ iff $\langle f(x_1), f(x_2) \rangle \in S$.

Example 3.12. Consider the following two sets $X = \{1, 2, 3\}$ and $Y = \{4, 5, 6\}$, and the relations successor, less than, and greater than. The function $f: X \rightarrow Y$ where $f(x) = x + 3$ is an isomorphism between X and Y .

3.5 Partial Functions

It is sometimes useful to relax the definition of function so that it is not required that the output of the function is defined for all possible inputs. Such mappings are called *partial functions*.

Definition 3.13. A *partial function* $f: X \rightarrow Y$ is a mapping which assigns to every element of X at most one element of Y . If f assigns an element of Y to $x \in X$, we say $f(x)$ is *defined*, and otherwise undefined. If $f(x)$ is defined, we write $f(x) \downarrow$, otherwise $f(x) \uparrow$. The *domain* of a partial function f is the subset of X where it is defined, i.e., $\text{dom}(f) = \{x : f(x) \downarrow\}$.

Example 3.14. Every function $f: X \rightarrow Y$ is also a partial function. Partial functions that are defined everywhere on X are called *total*.

Example 3.15. The partial function $f: \mathbb{R} \rightarrow \mathbb{R}$ given by $f(x) = 1/x$ is undefined for $x = 0$, and defined everywhere else.

Chapter 4

The Size of Sets

4.1 Introduction

When Georg Cantor developed set theory in the 1870s, his interest was in part to make palatable the idea of an infinite collection—an actual infinity, as the medievals would say. Key to this rehabilitation of the notion of the infinite was a way to assign sizes—“cardinalities”—to sets. The cardinality of a finite set is just a natural number, e.g., \emptyset has cardinality 0, and a set containing five things has cardinality 5. But what about infinite sets? Do they all have the same cardinality, ∞ ? It turns out, they do not.

The first important idea here is that of an enumeration. We can list every finite set by listing all its elements. For some infinite sets, we can also list all their elements if we allow the list itself to be infinite. Such sets are called countable. Cantor’s surprising result was that some infinite sets are not countable.

4.2 Countable Sets

Definition 4.1. Informally, an *enumeration* of a set X is a list (possibly infinite) such that every element of X appears some finite number of places into the list. If X has an enumeration, then X is said to be *countable*. If X is countable and infinite, we say X is countably infinite.

A couple of points about enumerations:

1. The order of elements of X in the enumeration does not matter, as long as every element appears: 4, 1, 25, 16, 9 enumerates the (set of the) first five square numbers just as well as 1, 4, 9, 16, 25 does.
2. Redundant enumerations are still enumerations: 1, 1, 2, 2, 3, 3, ... enumerates the same set as 1, 2, 3, ... does.

4. THE SIZE OF SETS

3. Order and redundancy *do* matter when we specify an enumeration: we can enumerate the natural numbers beginning with $1, 2, 3, 1, \dots$, but the pattern is easier to see when enumerated in the standard way as $1, 2, 3, 4, \dots$
4. Enumerations must have a beginning: $\dots, 3, 2, 1$ is not an enumeration of the natural numbers because it has no first element. To see how this follows from the informal definition, ask yourself, “at what place in the list does the number 76 appear?”
5. The following is not an enumeration of the natural numbers: $1, 3, 5, \dots, 2, 4, 6, \dots$. The problem is that the even numbers occur at places $\infty + 1, \infty + 2, \infty + 3$, rather than at finite positions.
6. Lists may be gappy: $2, -, 4, -, 6, -, \dots$ enumerates the even natural numbers.
7. The empty set is enumerable: it is enumerated by the empty list!

The following provides a more formal definition of an enumeration:

Definition 4.2. An *enumeration* of a set X is any surjective function $f : \mathbb{N} \rightarrow X$.

Let’s convince ourselves that the formal definition and the informal definition using a possibly gappy, possibly infinite list are equivalent. A surjective function (partial or total) from \mathbb{N} to a set X enumerates X . Such a function determines an enumeration as defined informally above. Then an enumeration for X is the list $f(0), f(1), f(2), \dots$. Since f is surjective, every element of X is guaranteed to be the value of $f(n)$ for some $n \in \mathbb{N}$. Hence, every element of X appears at some finite place in the list. Since the function may be partial or not injective, the list may be gappy or redundant, but that is acceptable (as noted above). On the other hand, given a list that enumerates all elements of X , we can define a surjective function $f : \mathbb{N} \rightarrow X$ by letting $f(n)$ be the $(n + 1)$ st member of the list, or undefined if the list has a gap in the $(n + 1)$ st spot.

Example 4.3. A function enumerating the natural numbers (\mathbb{N}) is simply the identity function given by $f(n) = n$.

Example 4.4. The functions $f : \mathbb{N} \rightarrow \mathbb{N}$ and $g : \mathbb{N} \rightarrow \mathbb{N}$ given by

$$f(n) = 2n \text{ and} \tag{4.1}$$

$$g(n) = 2n + 1 \tag{4.2}$$

enumerate the even natural numbers and the odd natural numbers, respectively. However, neither function is an enumeration of \mathbb{N} , since neither is surjective.

Example 4.5. The function $f(n) = \lceil \frac{(-1)^n n}{2} \rceil$ (where $\lceil x \rceil$ denotes the *ceiling* function, which rounds x up to the nearest integer) enumerates the set of integers \mathbb{Z} . Notice how f generates the values of \mathbb{Z} by “hopping” back and forth between positive and negative integers:

$$\begin{array}{cccccc} f(1) & f(2) & f(3) & f(4) & f(5) & f(6) & \dots \\ \lceil -\frac{1}{2} \rceil & \lceil \frac{2}{2} \rceil & \lceil -\frac{3}{2} \rceil & \lceil \frac{4}{2} \rceil & \lceil -\frac{5}{2} \rceil & \lceil \frac{6}{2} \rceil & \dots \\ 0 & 1 & -1 & 2 & -2 & 3 & \dots \end{array}$$

That is fine for “easy” sets. What about the set of, say, pairs of natural numbers?

$$\mathbb{N}^2 = \mathbb{N} \times \mathbb{N} = \{ \langle n, m \rangle : n, m \in \mathbb{N} \}$$

Another method we can use to enumerate sets is to organize them in an *array*, such as the following:

	1	2	3	4	...
1	$\langle 1, 1 \rangle$	$\langle 1, 2 \rangle$	$\langle 1, 3 \rangle$	$\langle 1, 4 \rangle$...
2	$\langle 2, 1 \rangle$	$\langle 2, 2 \rangle$	$\langle 2, 3 \rangle$	$\langle 2, 4 \rangle$...
3	$\langle 3, 1 \rangle$	$\langle 3, 2 \rangle$	$\langle 3, 3 \rangle$	$\langle 3, 4 \rangle$...
4	$\langle 4, 1 \rangle$	$\langle 4, 2 \rangle$	$\langle 4, 3 \rangle$	$\langle 4, 4 \rangle$...
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots

Clearly, every ordered pair in \mathbb{N}^2 will appear at least once in the array. In particular, $\langle n, m \rangle$ will appear in the n th column and m th row. But how do we organize the elements of an array into a list? The pattern in the array below demonstrates one way to do this:

	1	2	4	7	...
	3	5	8
	6	9
	10
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots

This pattern is called *Cantor’s zig-zag method*. Other patterns are perfectly permissible, as long as they “zig-zag” through every cell of the array. By Cantor’s zig-zag method, the enumeration for \mathbb{N}^2 according to this scheme would be:

$$\langle 1, 1 \rangle, \langle 1, 2 \rangle, \langle 2, 1 \rangle, \langle 1, 3 \rangle, \langle 2, 2 \rangle, \langle 3, 1 \rangle, \langle 1, 4 \rangle, \langle 2, 3 \rangle, \langle 3, 2 \rangle, \langle 4, 1 \rangle, \dots$$

What ought we do about enumerating, say, the set of ordered triples of natural numbers?

$$\mathbb{N}^3 = \mathbb{N} \times \mathbb{N} \times \mathbb{N} = \{ \langle n, m, k \rangle : n, m, k \in \mathbb{N} \}$$

4. THE SIZE OF SETS

We can think of \mathbb{N}^3 as the Cartesian product of \mathbb{N}^2 and \mathbb{N} , that is,

$$\mathbb{N}^3 = \mathbb{N}^2 \times \mathbb{N} = \{(\vec{a}, k) : \vec{a} \in \mathbb{N}^2, k \in \mathbb{N}\}$$

and thus we can enumerate \mathbb{N}^3 with an array by labelling one axis with the enumeration of \mathbb{N} , and the other axis with the enumeration of \mathbb{N}^2 :

	1	2	3	4	...
\langle 1, 1 \rangle	\langle 1, 1, 1 \rangle	\langle 1, 1, 2 \rangle	\langle 1, 1, 3 \rangle	\langle 1, 1, 4 \rangle	...
\langle 1, 2 \rangle	\langle 1, 2, 1 \rangle	\langle 1, 2, 2 \rangle	\langle 1, 2, 3 \rangle	\langle 1, 2, 4 \rangle	...
\langle 2, 1 \rangle	\langle 2, 1, 1 \rangle	\langle 2, 1, 2 \rangle	\langle 2, 1, 3 \rangle	\langle 2, 1, 4 \rangle	...
\langle 1, 3 \rangle	\langle 1, 3, 1 \rangle	\langle 1, 3, 2 \rangle	\langle 1, 3, 3 \rangle	\langle 1, 3, 4 \rangle	...
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots

Thus, by using a method like Cantor's zig-zag method, we may similarly obtain an enumeration of \mathbb{N}^3 .

4.3 Uncountable Sets

Some sets, such as the set \mathbb{N} of natural numbers, are infinite. So far we've seen examples of infinite sets which were all countable. However, there are also infinite sets which do not have this property. Such sets are called *uncountable*.

Cantor's method of diagonalization shows a set to be uncountable via a reductio proof. We start with the assumption that the set is countable, and show that a contradiction results from this assumption. Our first example is the set \mathbb{B}^ω of all infinite, non-gappy sequences of 0's and 1's.

Theorem 4.6. \mathbb{B}^ω is uncountable.

Proof. Suppose, for reductio, that \mathbb{B}^ω is countable, so that there is a list $s_1, s_2, s_3, s_4, \dots$ of all the elements of \mathbb{B}^ω . We may arrange this list, and the elements of each sequence s_i in it vertically in an array with the positive integers on the horizontal axis, as so:

	1	2	3	4	...
s_1	$s_1(1)$	$s_1(2)$	$s_1(3)$	$s_1(4)$...
s_2	$s_2(1)$	$s_2(2)$	$s_2(3)$	$s_2(4)$...
s_3	$s_3(1)$	$s_3(2)$	$s_3(3)$	$s_3(4)$...
s_4	$s_4(1)$	$s_4(2)$	$s_4(3)$	$s_4(4)$...
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots

Here $s_1(1)$ is a name for whatever number, a 0 or a 1, is the first member in the sequence s_1 , and so on.

Now define \bar{s} as follows: The n th member $\bar{s}(n)$ of the sequence \bar{s} is set to

$$\bar{s}(n) = \begin{cases} 1 & \text{if } s_n(n) = 0 \\ 0 & \text{if } s_n(n) = 1. \end{cases}$$

In other words, $\bar{s}(n)$ has the opposite value to $s_n(n)$.

Clearly \bar{s} is a non-gappy infinite sequence of 0s and 1s, since it is just the mirror sequence to the sequence of 0s and 1s that appear on the diagonal of our array. So \bar{s} is an element of \mathbb{B}^ω . Since it is an element of \mathbb{B}^ω , it must appear somewhere in the enumeration of \mathbb{B}^ω , that is, $\bar{s} = s_n$ for some n .

If $\bar{s} = s_n$, then for any m , $\bar{s}(m) = s_n(m)$. (This is just the criterion of identity for sequences—sequences are identical when they agree at every place.)

So in particular, $\bar{s}(n) = s_n(n)$. $\bar{s}(n)$ must be either a 0 or a 1. If it is a 0 then (given the definition of \bar{s}) $s_n(n)$ must be a 1. But if it is a 1 then $s_n(n)$ must be a 0. In either case $\bar{s}(n) \neq s_n(n)$. \square

Diagonalization need not involve the presence of an array, though the array method is where it takes its name.

Theorem 4.7. $\wp(\mathbb{Z}^+)$ is not enumerable.

Proof. Suppose, for reductio, that $\wp(\mathbb{Z}^+)$ is countable, and so it has an enumeration, i.e., a list of all subsets of \mathbb{Z}^+ :

$$Z_1, Z_2, Z_3, \dots$$

We now define a set \bar{Z} such that for any positive integer i , $i \in \bar{Z}$ iff $i \notin Z_i$:

$$\bar{Z} = \{i \in \mathbb{Z}^+ : i \notin Z_i\}$$

\bar{Z} is clearly a set of positive integers, and thus $\bar{Z} \in \wp(\mathbb{Z}^+)$. So \bar{Z} must be $= Z_k$ for some $k \in \mathbb{Z}^+$. And if that is the case, i.e., $\bar{Z} = Z_k$, then $i \in \bar{Z}$ iff $i \in Z_k$ for all $i \in \mathbb{Z}^+$.

In particular, $k \in \bar{Z}$ iff $k \in Z_k$.

Now either $k \in Z_k$ or $k \notin Z_k$. In the first case, by the previous line, $k \in \bar{Z}$. But we've defined \bar{Z} so that it contains exactly those $i \in \mathbb{Z}^+$ which are *not* elements of Z_i . So by that definition, we would have to also have $k \notin Z_k$. In the second case, $k \notin Z_k$. But now k satisfies the condition by which we have defined \bar{Z} , and that means that $k \in \bar{Z}$. And as $\bar{Z} = Z_k$, we get that $k \in Z_k$ after all. Either case leads to a contradiction. \square

4.4 Reduction

We showed $\wp(\mathbb{Z}^+)$ to be uncountable by a diagonalization argument. However, with the proof of the non-enumerability of \mathbb{B}^ω , the set of all infinite sequences of 0s and 1s, in place, we could have instead showed $\wp(\mathbb{Z}^+)$ to be

4. THE SIZE OF SETS

uncountable by showing that if $\wp(\mathbb{Z}^+)$ is enumerable then \mathbb{B}^ω is also enumerable. This called *reducing* one problem to another.

Proof of Theorem 4.7 by reduction. Suppose, for reductio, $\wp(\mathbb{Z}^+)$ is enumerable, and thus that there is an enumeration of it Z_1, Z_2, Z_3, \dots

Define the function $f: \wp(\mathbb{Z}^+) \rightarrow \mathbb{B}^\omega$ by letting $f(Z)$ be the sequence s_k such that $s_k(j) = 1$ iff $j \in Z$.

Every sequence of 0s and 1s corresponds to some set of positive integers, namely the one which has as its members those integers corresponding to the places where the sequence has 1s. In other words, this is a surjective function.

Now consider the list

$$f(Z_1), f(Z_2), f(Z_3), \dots$$

Since f is surjective, every member of \mathbb{B}^ω must appear as a value of f for some argument, and so must appear on the list. So this list must enumerate \mathbb{B}^ω .

So if $\wp(\mathbb{Z}^+)$ were countable, \mathbb{B}^ω would be enumerable. But \mathbb{B}^ω is uncountable (Theorem 4.6). \square

4.5 Equinumerous Sets

We have an intuitive notion of “size” of sets, which works fine for finite sets. But what about infinite sets? If we want to come up with a formal way of comparing the sizes of two sets of *any* size, it is a good idea to start with defining when sets are the same size. Let’s say sets of the same size are *equinumerous*. We want the formal notion of equinumerosity to correspond with our intuitive notion of “same size,” hence the formal notion ought to satisfy the following properties:

Reflexivity: Every set is equinumerous with itself.

Symmetry: For any sets X and Y , if X is equinumerous with Y , then Y is equinumerous with X .

Transitivity: For any sets X, Y , and Z , if X is equinumerous with Y and Y is equinumerous with Z , then X is equinumerous with Z .

In other words, we want equinumerosity to be an *equivalence relation*.

Definition 4.8. A set X is *equinumerous* with a set Y if and only if there is a total bijection f from X to Y (that is, $f: X \rightarrow Y$).

Proposition 4.9. *Equinumerosity defines an equivalence relation.*

Proof. Let X, Y , and Z be sets.

Reflexivity: Using the identity map $1_X: X \rightarrow X$, where $1_X(x) = x$ for all $x \in X$, we see that X is equinumerous with itself (clearly, 1_X is bijective).

Symmetry: Suppose that X is equinumerous with Y . Then there is a bijection $f: X \rightarrow Y$. Since f is bijective, its inverse f^{-1} is also a bijection. Since f is surjective, f^{-1} is total. Hence, $f^{-1}: Y \rightarrow X$ is a total bijection from Y to X , so Y is also equinumerous with X .

Transitivity: Suppose that X is equinumerous with Y via the total bijection f and that Y is equinumerous with Z via the total bijection g . Then the composition of $g \circ f: X \rightarrow Z$ is a total bijection, and X is thus equinumerous with Z .

Therefore, equinumerosity is an equivalence relation by the given definition. \square

Theorem 4.10. *Suppose X and Y are equinumerous. Then X is countable if and only if Y is.*

Proof. Let X and Y be equinumerous. Suppose that X is countable. Then there is a possibly partial, surjective function $f: \mathbb{N} \rightarrow X$. Since X and Y are equinumerous, there is a total bijection $g: X \rightarrow Y$. Claim: $g \circ f: \mathbb{N} \rightarrow Y$ is surjective. Clearly, $g \circ f$ is a function (since functions are closed under composition). To see $g \circ f$ is surjective, let $y \in Y$. Since g is surjective, there is an $x \in X$ such that $g(x) = y$. Since f is surjective, there is an $n \in \mathbb{N}$ such that $f(n) = x$. Hence,

$$(g \circ f)(n) = g(f(n)) = g(x) = y$$

and thus $g \circ f$ is surjective. Since $g \circ f: \mathbb{N} \rightarrow Y$ is surjective, it is an enumeration of Y , and so Y is countable. \square

4.6 Comparing Sizes of Sets

Just like we were able to make precise when two sets have the same size in a way that also accounts for the size of infinite sets, we can also compare the sizes of sets in a precise way. Our definition of “is smaller than (or equinumerous)” will require, instead of a bijection between the sets, a total injective function from the first set to the second. If such a function exists, the size of the first set is less than or equal to the size of the second. Intuitively, an injective function from one set to another guarantees that the range of the function has at least as many elements as the domain, since no two elements of the domain map to the same element of the range.

Definition 4.11. $|X| \leq |Y|$ if and only if there is an injective function $f: X \rightarrow Y$.

4. THE SIZE OF SETS

Theorem 4.12 (Schröder-Bernstein). *Let X and Y be sets. If $|X| \leq |Y|$ and $|Y| \leq |X|$, then $|X| = |Y|$.*

In other words, if there is a total injective function from X to Y , and if there is a total injective function from Y back to X , then there is a total bijection from X to Y . Sometimes, it can be difficult to think of a bijection between two equinumerous sets, so the Schröder-Bernstein theorem allows us to break the comparison down into cases so we only have to think of an injection from the first to the second, and vice-versa. The Schröder-Bernstein theorem, apart from being convenient, justifies the act of discussing the “sizes” of sets, for it tells us that set cardinalities have the familiar anti-symmetric property that numbers have.

Definition 4.13. $|X| < |Y|$ if and only if there is an injective function $f: X \rightarrow Y$ but no bijective $g: X \rightarrow Y$.

Theorem 4.14 (Cantor). *For all X , $|X| < |\wp(X)|$.*

Proof. The function $f: X \rightarrow \wp(X)$ that maps any $x \in X$ to its singleton $\{x\}$ is injective, since if $x \neq y$ then also $f(x) = \{x\} \neq \{y\} = f(y)$.

There cannot be a surjective function $g: X \rightarrow \wp(X)$, let alone a bijective one. For assume that a surjective $g: X \rightarrow \wp(X)$ exists. Then let $Y = \{x \in X : x \notin g(x)\}$. If $g(x)$ is defined for all $x \in X$, then Y is clearly a well-defined subset of X . If g is surjective, Y must be the value of g for some $x_0 \in X$, i.e., $Y = g(x_0)$. Now consider x_0 : it cannot be an element of Y , since if $x_0 \in Y$ then $x_0 \in g(x_0)$, and the definition of Y then would have $x_0 \notin Y$. On the other hand, it must be an element of Y , since if it were not, then $x_0 \notin Y = g(x_0)$. But then x_0 satisfies the defining condition of Y , and so $x_0 \in Y$. In either case, we have a contradiction. \square

Part II

First-order Logic

Chapter 5

Syntax and Semantics

5.1 First-Order Languages

Expressions of first-order logic are built up from a basic vocabulary containing *variables*, *constant symbols*, *predicate symbols* and sometimes *function symbols*. From them, together with logical connectives, quantifiers, and punctuation symbols such as parentheses and commas, *terms* and *formulas* are formed.

Informally, predicate symbols are names for properties and relations, constant symbols are names for individual objects, and function symbols are names for mappings. These, except for the identity predicate $=$, are the *non-logical symbols* and together make up a language. Any first-order language \mathcal{L} is determined by its non-logical symbols. In the most general case, \mathcal{L} contains infinitely many symbols of each kind.

In the general case, we make use of the following symbols in first-order logic:

1. Logical symbols
 - a) Logical connectives: \neg (negation), \wedge (conjunction), \vee (disjunction), \rightarrow (conditional), \forall (universal quantifier), \exists (existential quantifier).
 - b) The propositional constant for falsity \perp .
 - c) The two-place identity predicate $=$.
 - d) A countably infinite set of variables: v_0, v_1, v_2, \dots
2. Non-logical symbols, making up the *standard language* of first-order logic
 - a) A countably infinite set of n -place predicate symbols for each $n > 0$: $A_0^n, A_1^n, A_2^n, \dots$
 - b) A countably infinite set of constant symbols: c_0, c_1, c_2, \dots
 - c) A countably infinite set of n -place function symbols for each $n > 0$: $f_0^n, f_1^n, f_2^n, \dots$

3. Punctuation marks: (,), and the comma.

Most of our definitions and results will be formulated for the full standard language of first-order logic. However, depending on the application, we may also restrict the language to only a few predicate symbols, constant symbols, and function symbols.

Example 5.1. The language \mathcal{L}_A of arithmetic contains a single two-place predicate symbol $<$, a single constant symbol 0 , one one-place function symbol $'$, and two two-place function symbols $+$ and \times .

Example 5.2. The language of set theory \mathcal{L}_Z contains only the single two-place predicate symbol \in .

Example 5.3. The language of orders \mathcal{L}_{\leq} contains only the two-place predicate symbol \leq .

Again, these are conventions: officially, these are just aliases, e.g., $<$, \in , and \leq are aliases for A_0^2 , 0 for c_0 , $'$ for f_0^1 , $+$ for f_0^2 , \times for f_1^2 .

In addition to the primitive connectives and quantifiers introduced above, we also use the following *defined* symbols: \leftrightarrow (biconditional), truth \top

A defined symbol is not officially part of the language, but is introduced as an informal abbreviation: it allows us to abbreviate formulas which would, if we only used primitive symbols, get quite long. This is obviously an advantage. The bigger advantage, however, is that proofs become shorter. If a symbol is primitive, it has to be treated separately in proofs. The more primitive symbols, therefore, the longer our proofs.

You may be familiar with different terminology and symbols than the ones we use above. Logic texts (and teachers) commonly use either \sim , \neg , and $!$ for “negation”, \wedge , \cdot , and $\&$ for “conjunction”. Commonly used symbols for the “conditional” or “implication” are \rightarrow , \Rightarrow , and \supset . Symbols for “biconditional,” “bi-implication,” or “(material) equivalence” are \leftrightarrow , \Leftrightarrow , and \equiv . The \perp symbol is variously called “falsity,” “falsum,” “absurdity,” or “bottom.” The \top symbol is variously called “truth,” “verum,” or “top.”

It is conventional to use lower case letters (e.g., a , b , c) from the beginning of the Latin alphabet for constant symbols (sometimes called names), and lower case letters from the end (e.g., x , y , z) for variables. Quantifiers combine with variables, e.g., x ; notational variations include $\forall x$, $(\forall x)$, (x) , Πx , \bigwedge_x for the universal quantifier and $\exists x$, $(\exists x)$, (Ex) , Σx , \bigvee_x for the existential quantifier.

We might treat all the propositional operators and both quantifiers as primitive symbols of the language. We might instead choose a smaller stock of primitive symbols and treat the other logical operators as defined. “Truth functionally complete” sets of Boolean operators include $\{\neg, \vee\}$, $\{\neg, \wedge\}$, and $\{\neg, \rightarrow\}$ —these can be combined with either quantifier for an expressively complete first-order language.

You may be familiar with two other logical operators: the Sheffer stroke $|$ (named after Henry Sheffer), and Peirce's arrow \downarrow , also known as Quine's dagger. When given their usual readings of "nand" and "nor" (respectively), these operators are truth functionally complete by themselves.

5.2 Terms and Formulas

Once a first-order language \mathcal{L} is given, we can define expressions built up from the basic vocabulary of \mathcal{L} . These include in particular *terms* and *formulas*.

Definition 5.4 (Terms). The set of *terms* $\text{Trm}(\mathcal{L})$ of \mathcal{L} is defined inductively by:

1. Every variable is a term.
2. Every constant symbol of \mathcal{L} is a term.
3. If f is an n -place function symbol and t_1, \dots, t_n are terms, then $f(t_1, \dots, t_n)$ is a term.
4. Nothing else is a term.

A term containing no variables is a *closed term*.

The constant symbols appear in our specification of the language and the terms as a separate category of symbols, but they could instead have been included as zero-place function symbols. We could then do without the second clause in the definition of terms. We just have to understand $f(t_1, \dots, t_n)$ as just f by itself if $n = 0$.

Definition 5.5 (Formula). The set of *formulas* $\text{Frm}(\mathcal{L})$ of the language \mathcal{L} is defined inductively as follows:

1. \perp is an atomic formula.
2. If R is an n -place predicate symbol of \mathcal{L} and t_1, \dots, t_n are terms of \mathcal{L} , then $R(t_1, \dots, t_n)$ is an atomic formula.
3. If t_1 and t_2 are terms of \mathcal{L} , then $=(t_1, t_2)$ is an atomic formula.
4. If A is a formula, then $\neg A$ is formula.
5. If A and B are formulas, then $(A \wedge B)$ is a formula.
6. If A and B are formulas, then $(A \vee B)$ is a formula.
7. If A and B are formulas, then $(A \rightarrow B)$ is a formula.
8. If A is a formula and x is a variable, then $\forall x A$ is a formula.

9. If A is a formula and x is a variable, then $\exists x A$ is a formula.

10. Nothing else is a formula.

The definitions of the set of terms and that of formulas are *inductive definitions*. Essentially, we construct the set of formulas in infinitely many stages. In the initial stage, we pronounce all atomic formulas to be formulas; this corresponds to the first few cases of the definition, i.e., the cases for \perp , $R(t_1, \dots, t_n)$ and $=(t_1, t_2)$. “Atomic formula” thus means any formula of this form.

The other cases of the definition give rules for constructing new formulas out of formulas already constructed. At the second stage, we can use them to construct formulas out of atomic formulas. At the third stage, we construct new formulas from the atomic formulas and those obtained in the second stage, and so on. A formula is anything that is eventually constructed at such a stage, and nothing else.

By convention, we write $=$ between its arguments and leave out the parentheses: $t_1 = t_2$ is an abbreviation for $=(t_1, t_2)$. Moreover, $\neg=(t_1, t_2)$ is abbreviated as $t_1 \neq t_2$. When writing a formula $(B * C)$ constructed from B , C using a two-place connective $*$, we will often leave out the outermost pair of parentheses and write simply $B * C$.

Some logic texts require that the variable x must occur in A in order for $\exists x A$ and $\forall x A$ to count as formulas. Nothing bad happens if you don't require this, and it makes things easier.

Definition 5.6. Formulas constructed using the defined operators are to be understood as follows:

1. \top abbreviates $\neg\perp$.
2. $A \leftrightarrow B$ abbreviates $(A \rightarrow B) \wedge (B \rightarrow A)$.

If we work in a language for a specific application, we will often write two-place predicate symbols and function symbols between the respective terms, e.g., $t_1 < t_2$ and $(t_1 + t_2)$ in the language of arithmetic and $t_1 \in t_2$ in the language of set theory. The successor function in the language of arithmetic is even written conventionally *after* its argument: t' . Officially, however, these are just conventional abbreviations for $A_0^2(t_1, t_2)$, $f_0^2(t_1, t_2)$, $A_0^2(t_1, t_2)$ and $f_0^1(t)$, respectively.

Definition 5.7. The symbol \equiv expresses syntactic identity between strings of symbols, i.e., $A \equiv B$ iff A and B are strings of symbols of the same length and which contain the same symbol in each place.

The \equiv symbol may be flanked by strings obtained by concatenation, e.g., $A \equiv (B \vee C)$ means: the string of symbols A is the same string as the one obtained by concatenating an opening parenthesis, the string B , the \vee symbol,

the string C , and a closing parenthesis, in this order. If this is the case, then we know that the first symbol of A is an opening parenthesis, A contains B as a substring (starting at the second symbol), that substring is followed by \vee , etc.

5.3 Unique Readability

The way we defined formulas guarantees that every formula has a *unique reading*, i.e., there is essentially only one way of constructing it according to our formation rules for formulas and only one way of “interpreting” it. If this were not so, we would have ambiguous formulas, i.e., formulas that have more than one reading or interpretation—and that is clearly something we want to avoid. But more importantly, without this property, most of the definitions and proofs we are going to give will not go through.

Perhaps the best way to make this clear is to see what would happen if we had given bad rules for forming formulas that would not guarantee unique readability. For instance, we could have forgotten the parentheses in the formation rules for connectives, e.g., we might have allowed this:

If A and B are formulas, then so is $A \rightarrow B$.

Starting from an atomic formula D , this would allow us to form $D \rightarrow D$, and from this, together with D , we would get $D \rightarrow D \rightarrow D$. But there are two ways to do this: one where we take D to be A and $D \rightarrow D$ to be B , and the other where A is $D \rightarrow D$ and B is D . Correspondingly, there are two ways to “read” the formula $D \rightarrow D \rightarrow D$. It is of the form $B \rightarrow C$ where B is D and C is $D \rightarrow D$, but *it is also* of the form $B \rightarrow C$ with B being $D \rightarrow D$ and C being D .

If this happens, our definitions will not always work. For instance, when we define the main operator of a formula, we say: in a formula of the form $B \rightarrow C$, the main operator is the indicated occurrence of \rightarrow . But if we can match the formula $D \rightarrow D \rightarrow D$ with $B \rightarrow C$ in the two different ways mentioned above, then in one case we get the first occurrence of \rightarrow as the main operator, and in the second case the second occurrence. But we intend the main operator to be a *function* of the formula, i.e., every formula must have exactly one main operator occurrence.

Lemma 5.8. *The number of left and right parentheses in a formula A are equal.*

Proof. We prove this by induction on the way A is constructed. This requires two things: (a) We have to prove first that all atomic formulas have the property in question (the induction basis). (b) Then we have to prove that when we construct new formulas out of given formulas, the new formulas have the property provided the old ones do.

Let $l(A)$ be the number of left parentheses, and $r(A)$ the number of right parentheses in A , and $l(t)$ and $r(t)$ similarly the number of left and right

parentheses in a term t . We leave the proof that for any term t , $l(t) = r(t)$ as an exercise.

1. $A \equiv \perp$: A has 0 left and 0 right parentheses.
2. $A \equiv R(t_1, \dots, t_n)$: $l(A) = 1 + l(t_1) + \dots + l(t_n) = 1 + r(t_1) + \dots + r(t_n) = r(A)$. Here we make use of the fact, left as an exercise, that $l(t) = r(t)$ for any term t .
3. $A \equiv t_1 = t_2$: $l(A) = l(t_1) + l(t_2) = r(t_1) + r(t_2) = r(A)$.
4. $A \equiv \neg B$: By induction hypothesis, $l(B) = r(B)$. Thus $l(A) = l(B) = r(B) = r(A)$.
5. $A \equiv (B * C)$: By induction hypothesis, $l(B) = r(B)$ and $l(C) = r(C)$. Thus $l(A) = 1 + l(B) + l(C) = 1 + r(B) + r(C) = r(A)$.
6. $A \equiv \forall x B$: By induction hypothesis, $l(B) = r(B)$. Thus, $l(A) = l(B) = r(B) = r(A)$.
7. $A \equiv \exists x B$: Similarly.

□

Definition 5.9. A string of symbols B is a proper prefix of a string of symbols A if concatenating B and a non-empty string of symbols yields A .

Lemma 5.10. If A is a formula, and B is a proper prefix of A , then B is not a formula.

Proof. Exercise. □

Proposition 5.11. If A is an atomic formula, then it satisfies one, and only one of the following conditions.

1. $A \equiv \perp$.
2. $A \equiv R(t_1, \dots, t_n)$ where R is an n -place predicate symbol, t_1, \dots, t_n are terms, and each of R, t_1, \dots, t_n is uniquely determined.
3. $A \equiv t_1 = t_2$ where t_1 and t_2 are uniquely determined terms.

Proof. Exercise. □

Proposition 5.12 (Unique Readability). Every formula satisfies one, and only one of the following conditions.

1. A is atomic.
2. A is of the form $\neg B$.
3. A is of the form $(B \wedge C)$.

4. A is of the form $(B \vee C)$.
5. A is of the form $(B \rightarrow C)$.
6. A is of the form $\forall x B$.
7. A is of the form $\exists x B$.

Moreover, in each case B , or B and C , are uniquely determined. This means that, e.g., there are no different pairs B, C and B', C' so that A is both of the form $(B \rightarrow C)$ and $(B' \rightarrow C')$.

Proof. The formation rules require that if a formula is not atomic, it must start with an opening parenthesis $($, \neg , or with a quantifier. On the other hand, every formula that start with one of the following symbols must be atomic: a predicate symbol, a function symbol, a constant symbol, \perp .

So we really only have to show that if A is of the form $(B * C)$ and also of the form $(B' *' C')$, then $B \equiv B'$, $C \equiv C'$, and $* = *'$.

So suppose both $A \equiv (B * C)$ and $A \equiv (B' *' C')$. Then either $B \equiv B'$ or not. If it is, clearly $* = *'$ and $C \equiv C'$, since they then are substrings of A that begin in the same place and are of the same length. The other case is $C \not\equiv C'$. Since C and C' are both substrings of A that begin at the same place, one must be a prefix of the other. But this is impossible by [Lemma 5.10](#). \square

5.4 Main operator of a Formula

It is often useful to talk about the last operator used in constructing a formula A . This operator is called the *main operator* of A . Intuitively, it is the “outermost” operator of A . For example, the main operator of $\neg A$ is \neg , the main operator of $(A \vee B)$ is \vee , etc.

Definition 5.13 (Main operator). The *main operator* of a formula A is defined as follows:

1. A is atomic: A has no main operator.
2. $A \equiv \neg B$: the main operator of A is \neg .
3. $A \equiv (B \wedge C)$: the main operator of A is \wedge .
4. $A \equiv (B \vee C)$: the main operator of A is \vee .
5. $A \equiv (B \rightarrow C)$: the main operator of A is \rightarrow .
6. $A \equiv \forall x B$: the main operator of A is \forall .
7. $A \equiv \exists x B$: the main operator of A is \exists .

In each case, we intend the specific indicated *occurrence* of the main operator in the formula. For instance, since the formula $((D \rightarrow E) \rightarrow (E \rightarrow D))$ is of the form $(B \rightarrow C)$ where B is $(D \rightarrow E)$ and C is $(E \rightarrow D)$, the second occurrence of \rightarrow is the main operator.

This is a *recursive* definition of a function which maps all non-atomic formulas to their main operator occurrence. Because of the way formulas are defined inductively, every formula A satisfies one of the cases in [Definition 5.13](#). This guarantees that for each non-atomic formula A a main operator exists. Because each formula satisfies only one of these conditions, and because the smaller formulas from which A is constructed are uniquely determined in each case, the main operator occurrence of A is unique, and so we have defined a function.

We call formulas by the following names depending on which symbol their main operator is:

Main operator	Type of formula	Example
none	atomic (formula)	$\perp, R(t_1, \dots, t_n), t_1 = t_2$
\neg	negation	$\neg A$
\wedge	conjunction	$(A \wedge B)$
\vee	disjunction	$(A \vee B)$
\rightarrow	conditional	$(A \rightarrow B)$
\forall	universal (formula)	$\forall x A$
\exists	existential (formula)	$\exists x A$

5.5 Subformulas

It is often useful to talk about the formulas that “make up” a given formula. We call these its *subformulas*. Any formula counts as a subformula of itself; a subformula of A other than A itself is a *proper subformula*.

Definition 5.14 (Immediate Subformula). If A is a formula, the *immediate subformulas* of A are defined inductively as follows:

1. Atomic formulas have no immediate subformulas.
2. $A \equiv \neg B$: The only immediate subformula of A is B .
3. $A \equiv (B * C)$: The immediate subformulas of A are B and C ($*$ is any one of the two-place connectives).
4. $A \equiv \forall x B$: The only immediate subformula of A is B .
5. $A \equiv \exists x B$: The only immediate subformula of A is B .

Definition 5.15 (Proper Subformula). If A is a formula, the *proper subformulas* of A are recursively as follows:

1. Atomic formulas have no proper subformulas.
2. $A \equiv \neg B$: The proper subformulas of A are B together with all proper subformulas of B .
3. $A \equiv (B * C)$: The proper subformulas of A are B, C , together with all proper subformulas of B and those of C .
4. $A \equiv \forall x B$: The proper subformulas of A are B together with all proper subformulas of B .
5. $A \equiv \exists x B$: The proper subformulas of A are B together with all proper subformulas of B .

Definition 5.16 (Subformula). The subformulas of A are A itself together with all its proper subformulas.

Note the subtle difference in how we have defined immediate subformulas and proper subformulas. In the first case, we have directly defined the immediate subformulas of a formula A for each possible form of A . It is an explicit definition by cases, and the cases mirror the inductive definition of the set of formulas. In the second case, we have also mirrored the way the set of all formulas is defined, but in each case we have also included the proper subformulas of the smaller formulas B, C in addition to these formulas themselves. This makes the definition *recursive*. In general, a definition of a function on an inductively defined set (in our case, formulas) is recursive if the cases in the definition of the function make use of the function itself. To be well defined, we must make sure, however, that we only ever use the values of the function for arguments that come “before” the one we are defining—in our case, when defining “proper subformula” for $(B * C)$ we only use the proper subformulas of the “earlier” formulas B and C .

5.6 Free Variables and Sentences

Definition 5.17 (Free occurrences of a variable). The *free* occurrences of a variable in a formula are defined inductively as follows:

1. A is atomic: all variable occurrences in A are free.
2. $A \equiv \neg B$: the free variable occurrences of A are exactly those of B .
3. $A \equiv (B * C)$: the free variable occurrences of A are those in B together with those in C .
4. $A \equiv \forall x B$: the free variable occurrences in A are all of those in B except for occurrences of x .

5. $A \equiv \exists x B$: the free variable occurrences in A are all of those in B except for occurrences of x .

Definition 5.18 (Bound Variables). An occurrence of a variable in a formula A is *bound* if it is not free.

Definition 5.19 (Scope). If $\forall x B$ is an occurrence of a subformula in a formula A , then the corresponding occurrence of B in A is called the *scope* of the corresponding occurrence of $\forall x$. Similarly for $\exists x$.

If B is the scope of a quantifier occurrence $\forall x$ or $\exists x$ in A , then all occurrences of x which are free in B are said to be *bound by* the mentioned quantifier occurrence.

Example 5.20. Here is a somewhat complicated formula A :

$$\forall x_0 \underbrace{(A_0^1(x_0) \rightarrow A_0^2(x_0, x_1))}_B \rightarrow \exists x_1 \underbrace{(A_1^2(x_0, x_1) \vee \forall x_0 \overbrace{\neg A_1^1(x_0)}^D)}_C$$

B is the scope of the first $\forall x_0$, C is the scope of $\exists x_1$, and D is the scope of the second $\forall x_0$. The first $\forall x_0$ binds the occurrences of x_0 in B , $\exists x_1$ the occurrence of x_1 in C , and the second $\forall x_0$ binds the occurrence of x_0 in D . The first occurrence of x_1 and the fourth occurrence of x_0 are free in A . The last occurrence of x_0 is free in D , but bound in C and A .

Definition 5.21 (Sentence). A formula A is a *sentence* iff it contains no free occurrences of variables.

5.7 Substitution

Definition 5.22 (Substitution in a term). We define $s[t/x]$, the result of *substituting* t for every occurrence of x in s , recursively:

1. $s \equiv c$: $s[t/x]$ is just s .
2. $s \equiv y$: $s[t/x]$ is also just s , provided y is a variable other than x .
3. $s \equiv x$: $s[t/x]$ is t .
4. $s \equiv f(t_1, \dots, t_n)$: $s[t/x]$ is $f(t_1[t/x], \dots, t_n[t/x])$.

Definition 5.23. A term t is *free for* x in A if none of the free occurrences of x in A occur in the scope of a quantifier that binds a variable in t .

Definition 5.24 (Substitution in a formula). If A is a formula, x is a variable, and t is a term free for x in A , then $A[t/x]$ is the result of substituting t for all free occurrences of x in A .

1. $A \equiv P(t_1, \dots, t_n)$: $A[t/x]$ is $P(t_1[t/x], \dots, t_n[t/x])$.
2. $A \equiv t_1 = t_2$: $A[t/x]$ is $t_1[t/x] = t_2[t/x]$.
3. $A \equiv \neg B$: $A[t/x]$ is $\neg B[t/x]$.
4. $A \equiv (B \wedge C)$: $A[t/x]$ is $(B[t/x] \wedge C[t/x])$.
5. $A \equiv (B \vee C)$: $A[t/x]$ is $(B[t/x] \vee C[t/x])$.
6. $A \equiv (B \rightarrow C)$: $A[t/x]$ is $(B[t/x] \rightarrow C[t/x])$.
7. $A \equiv \forall y B$: $A[t/x]$ is $\forall y B[t/x]$, provided y is a variable other than x ; otherwise $A[t/x]$ is just A .
8. $A \equiv \exists y B$: $A[t/x]$ is $\exists y B[t/x]$, provided y is a variable other than x ; otherwise $A[t/x]$ is just A .

Note that substitution may be vacuous: If x does not occur in A at all, then $A[t/x]$ is just A .

The restriction that t must be free for x in A is necessary to exclude cases like the following. If $A \equiv \exists y x < y$ and $t \equiv y$, then $A[t/y]$ would be $\exists y y < y$. In this case the free variable y is “captured” by the quantifier $\exists y$ upon substitution, and that is undesirable. For instance, we would like it to be the case that whenever $\forall x B$ holds, so does $B[t/x]$. But consider $\forall x \exists y x < y$ (here B is $\exists y x < y$). It is sentence that is true about, e.g., the natural numbers: for every number x there is a number y greater than it. If we allowed y as a possible substitution for x , we would end up with $B[y/x] \equiv \exists y y < y$, which is false. We prevent this by requiring that none of the free variables in t would end up being bound by a quantifier in A .

We often use the following convention to avoid cumbersome notation: If A is a formula with a free variable x , we write $A(x)$ to indicate this. When it is clear which A and x we have in mind, and t is a term (assumed to be free for x in $A(x)$), then we write $A(t)$ as short for $A(x)[t/x]$.

5.8 Structures for First-order Languages

First-order languages are, by themselves, *uninterpreted*: the constant symbols, function symbols, and predicate symbols have no specific meaning attached to them. Meanings are given by specifying a *structure*. It specifies the *domain*, i.e., the objects which the constant symbols pick out, the function symbols operate on, and the quantifiers range over. In addition, it specifies which constant symbols pick out which objects, how a function symbol maps objects to objects, and which objects the predicate symbols apply to. Structures are the basis for *semantic* notions in logic, e.g., the notion of consequence, validity, satisfiability. They are variously called “structures,” “interpretations,” or “models” in the literature.

Definition 5.25 (Structures). A *structure* \mathbf{M} , for a language \mathcal{L} of first-order logic consists of the following elements:

1. *Domain*: a non-empty set, $|\mathbf{M}|$
2. *Interpretation of constant symbols*: for each constant symbol c of \mathcal{L} , an element $c^{\mathbf{M}} \in |\mathbf{M}|$
3. *Interpretation of predicate symbols*: for each n -place predicate symbol R of \mathcal{L} (other than $=$), an n -ary relation $R^{\mathbf{M}} \subseteq |\mathbf{M}|^n$
4. *Interpretation of function symbols*: for each n -place function symbol f of \mathcal{L} , an n -ary function $f^{\mathbf{M}}: |\mathbf{M}|^n \rightarrow |\mathbf{M}|$

Example 5.26. A structure \mathbf{M} for the language of arithmetic consists of a set, an element of $|\mathbf{M}|$, $o^{\mathbf{M}}$, as interpretation of the constant symbol o , a one-place function $\iota^{\mathbf{M}}: |\mathbf{M}| \rightarrow |\mathbf{M}|$, two two-place functions $+^{\mathbf{M}}$ and $\times^{\mathbf{M}}$, both $|\mathbf{M}|^2 \rightarrow |\mathbf{M}|$, and a two-place relation $<^{\mathbf{M}} \subseteq |\mathbf{M}|^2$.

An obvious example of such a structure is the following:

1. $|\mathbf{N}| = \mathbb{N}$
2. $o^{\mathbf{N}} = 0$
3. $\iota^{\mathbf{N}}(n) = n + 1$ for all $n \in \mathbb{N}$
4. $+^{\mathbf{N}}(n, m) = n + m$ for all $n, m \in \mathbb{N}$
5. $\times^{\mathbf{N}}(n, m) = n \cdot m$ for all $n, m \in \mathbb{N}$
6. $<^{\mathbf{N}} = \{ \langle n, m \rangle : n \in \mathbb{N}, m \in \mathbb{N}, n < m \}$

The structure \mathbf{N} for \mathcal{L}_A so defined is called the *standard model of arithmetic*, because it interprets the non-logical constants of \mathcal{L}_A exactly how you would expect.

However, there are many other possible structures for \mathcal{L}_A . For instance, we might take as the domain the set \mathbb{Z} of integers instead of \mathbb{N} , and define the interpretations of $o, \iota, +, \times, <$ accordingly. But we can also define structures for \mathcal{L}_A which have nothing even remotely to do with numbers.

Example 5.27. A structure \mathbf{M} for the language \mathcal{L}_Z of set theory requires just a set and a single-two place relation. So technically, e.g., the set of people plus the relation “ x is older than y ” could be used as a structure for \mathcal{L}_Z , as well as \mathbb{N} together with $n \geq m$ for $n, m \in \mathbb{N}$.

A particularly interesting structure for \mathcal{L}_Z in which the elements of the domain are actually sets, and the interpretation of \in actually is the relation “ x is an element of y ” is the structure **HF** of *hereditarily finite sets*:

1. $|\mathbf{HF}| = \emptyset \cup \wp(\emptyset) \cup \wp(\wp(\emptyset)) \cup \wp(\wp(\wp(\emptyset))) \cup \dots;$

$$2. \in^{\mathbf{HF}} = \{\langle x, y \rangle : x, y \in |\mathbf{HF}|, x \in y\}.$$

Recall that a term is *closed* if it contains no variables.

Definition 5.28 (Value of closed terms). If t is a closed term of the language \mathcal{L} and M is a structure for \mathcal{L} , the *value* $\text{Val}^M(t)$ is defined as follows:

1. If t is just the constant symbol c , then $\text{Val}^M(c) = c^M$.
2. If t is of the form $f(t_1, \dots, t_n)$, then $\text{Val}^M(t)$ is $f^M(\text{Val}^M(t_1), \dots, \text{Val}^M(t_n))$.

Definition 5.29 (Covered structure). A structure is *covered* if every element of the domain is the value of some closed term.

Example 5.30. Let \mathcal{L} be the language with constant symbols $zero, one, two, \dots$, the binary predicate symbols $=$ and $<$, and the binary function symbols $+$ and \times . Then a structure M for \mathcal{L} is the one with domain $|M| = \{0, 1, 2, \dots\}$ and name assignment $zero^M = 0, one^M = 1, two^M = 2$, and so forth. For the binary relation symbol $<$, the set $<^M$ is the set of all pairs $\langle c_1, c_2 \rangle \in |M|^2$ such that the integer c_1 is less than the integer c_2 : for example, $\langle 1, 3 \rangle \in <^M$ but $\langle 2, 2 \rangle \notin <^M$. For the binary function symbol $+$, define $+^M$ in the usual way—for example, $+^M(2, 3)$ maps to 5, and similarly for the binary function symbol \times . Hence, the value of $four$ is just 4, and the value of $\times(two, +(three, zero))$ (or in infix notation, $two \times (three + zero)$) is

$$\begin{aligned} \text{Val}^M(\times(two, +(three, zero))) &= \times^M(\text{Val}^M(two), \text{Val}^M(two, +(three, zero))) \\ &= \times^M(\text{Val}^M(two), +^M(\text{Val}^M(three), \text{Val}^M(zero))) \\ &= \times^M(two^M, +^M(three^M, zero^M)) \\ &= \times^M(2, +^M(3, 0)) \\ &= \times^M(2, 3) \\ &= 6 \end{aligned}$$

The stipulations we make as to what counts as a structure impact our logic. For example, the choice to prevent empty domains ensures, given the usual account of satisfaction (or truth) for quantified sentences, that $\exists x (A(x) \vee \neg A(x))$ is valid—that is, a logical truth. And the stipulation that all constant symbols must refer to an object in the domain ensures that the existential generalization is a sound pattern of inference: $A(a)$, therefore $\exists x A(x)$. If we allowed names to refer outside the domain, or to not refer, then we would be on our way to a *free logic*, in which existential generalization requires an additional premise: $A(a)$ and $\exists x x = a$, therefore $\exists x A(x)$.

5.9 Satisfaction of a Formula in a Structure

The basic notion that relates expressions such as terms and formulas, on the one hand, and structures on the other, are those of *value* of a term and *satisfaction* of a formula. Informally, the value of a term is an element of a structure—if the term is just a constant, its value is the object assigned to the constant by the structure, and if it is built up using function symbols, the value is computed from the values of constants and the functions assigned to the functions in the term. A formula is *satisfied* in a structure if the interpretation given to the predicates makes the formula true in the domain of the structure. This notion of satisfaction is specified inductively: the specification of the structure directly states when atomic formulas are satisfied, and we define when a complex formula is satisfied depending on the main connective or quantifier and whether or not the immediate subformulas are satisfied. The case of the quantifiers here is a bit tricky, as the immediate subformula of a quantified formula has a free variable, and structures don't specify the values of variables. In order to deal with this difficulty, we also introduce *variable assignments* and define satisfaction not with respect to a structure alone, but with respect to a structure plus a variable assignment.

Definition 5.31 (Variable Assignment). A *variable assignment* s for a structure M is a function which maps each variable to an element of $|M|$, i.e., $s: \text{Var} \rightarrow |M|$.

A structure assigns a value to each constant symbol, and a variable assignment to each variable. But we want to use terms built up from them to also name elements of the domain. For this we define the value of terms inductively. For constant symbols and variables the value is just as the structure or the variable assignment specifies it; for more complex terms it is computed recursively using the functions the structure assigns to the function symbols.

Definition 5.32 (Value of Terms). If t is a term of the language \mathcal{L} , M is a structure for \mathcal{L} , and s is a variable assignment for M , the *value* $\text{Val}_s^M(t)$ is defined as follows:

1. $t \equiv c$: $\text{Val}_s^M(t) = c^M$.
2. $t \equiv x$: $\text{Val}_s^M(t) = s(x)$.
3. $t \equiv f(t_1, \dots, t_n)$:

$$\text{Val}_s^M(t) = f^M(\text{Val}_s^M(t_1), \dots, \text{Val}_s^M(t_n)).$$

Definition 5.33 (x -Variant). If s is a variable assignment for a structure M , then any variable assignment s' for M which differs from s at most in what it assigns to x is called an *x -variant* of s . If s' is an x -variant of s we write $s \sim_x s'$.

Note that an x -variant of an assignment s does not *have* to assign something different to x . In fact, every assignment counts as an x -variant of itself.

Definition 5.34 (Satisfaction). Satisfaction of a formula A in a structure M relative to a variable assignment s , in symbols: $M, s \models A$, is defined recursively as follows. (We write $M, s \not\models A$ to mean “not $M, s \models A$.”)

1. $A \equiv \perp$: not $M, s \models A$.
2. $A \equiv R(t_1, \dots, t_n)$: $M, s \models A$ iff $\langle \text{Val}_s^M(t_1), \dots, \text{Val}_s^M(t_n) \rangle \in R^M$.
3. $A \equiv t_1 = t_2$: $M, s \models A$ iff $\text{Val}_s^M(t_1) = \text{Val}_s^M(t_2)$.
4. $A \equiv \neg B$: $M, s \models A$ iff $M, s \not\models B$.
5. $A \equiv (B \wedge C)$: $M, s \models A$ iff $M, s \models B$ and $M, s \models C$.
6. $A \equiv (B \vee C)$: $M, s \models A$ iff $M, s \models B$ or $M, s \models C$ (or both).
7. $A \equiv (B \rightarrow C)$: $M, s \models A$ iff $M, s \not\models B$ or $M, s \models C$ (or both).
8. $A \equiv \forall x B$: $M, s \models A$ iff for every x -variant s' of s , $M, s' \models B$.
9. $A \equiv \exists x B$: $M, s \models A$ iff there is an x -variant s' of s so that $M, s' \models B$.

The variable assignments are important in the last two clauses. We cannot define satisfaction of $\forall x B(x)$ by “for all $a \in |M|$, $M \models B(a)$.” We cannot define satisfaction of $\exists x B(x)$ by “for at least one $a \in |M|$, $M \models B(a)$.” The reason is that a is not symbol of the language, and so $B(a)$ is not a formula (that is, $B[a/x]$ is undefined). We also cannot assume that we have constant symbols or terms available that name every element of M , since there is nothing in the definition of structures that requires it. Even in the standard language the set of constant symbols is countably infinite, so if $|M|$ is not countable there aren’t even enough constant symbols to name every object.

A variable assignment s provides a value for *every* variable in the language. This is of course not necessary: whether or not a formula A is satisfied in a structure with respect to s only depends on the assignments s makes to the free variables that actually occur in A . This is the content of the next theorem. We require variable assignments to assign values to all variables simply because it makes things a lot easier.

Proposition 5.35. *If x_1, \dots, x_n are the only free variables in A and $s(x_i) = s'(x_i)$ for $i = 1, \dots, n$, then $M, s \models A$ iff $M, s' \models A$.*

Proof. We use induction on the complexity of A . For the base case, where A is atomic, A can be: \perp , $R(t_1, \dots, t_k)$ for a k -place predicate R and terms t_1, \dots, t_k , or $t_1 = t_2$ for terms t_1 and t_2 .

1. $A \equiv \perp$: both $M, s \not\models A$ and $M, s' \not\models A$.

5. SYNTAX AND SEMANTICS

2. $A \equiv R(t_1, \dots, t_k)$: let $M, s \models A$. Then $\langle \text{Val}_s^M(t_1), \dots, \text{Val}_s^M(t_k) \rangle \in R^M$. For $i = 1, \dots, k$, if t_i is a constant, then $\text{Val}_s^M(t_i) = \text{Val}^M(t_i) = \text{Val}_{s'}^M(t_i)$. If t_i is a free variable, then since the mappings s and s' agree on all free variables, $\text{Val}_s^M(t_i) = s(t_i) = s'(t_i) = \text{Val}_{s'}^M(t_i)$. Similarly, if t_i is of the form $f(t'_1, \dots, t'_j)$, we will also get $\text{Val}_s^M(t_i) = \text{Val}_{s'}^M(t_i)$. Hence, $\text{Val}_s^M(t_i) = \text{Val}_{s'}^M(t_i)$ for any term t_i for $i = 1, \dots, k$, so we also have $\langle \text{Val}_{s'}^M(t_1), \dots, \text{Val}_{s'}^M(t_k) \rangle \in R^M$.
3. $A \equiv t_1 = t_2$: if $M, s \models A$, $\text{Val}_{s'}^M(t_1) = \text{Val}_s^M(t_1) = \text{Val}_s^M(t_2) = \text{Val}_{s'}^M(t_2)$, so $M, s' \models t_1 = t_2$.

Now assume $M, s \models B$ iff $M, s' \models B$ for all formulas B less complex than A . The induction step proceeds by cases determined by the main operator of A . In each case, we only demonstrate the forward direction of the biconditional; the proof of the reverse direction is symmetrical.

1. $A \equiv \neg B$: if $M, s \models A$, then $M, s \not\models B$, so by the induction hypothesis, $M, s' \not\models B$, hence $M, s' \models A$.
2. $A \equiv B \wedge C$: exercise.
3. $A \equiv B \vee C$: if $M, s \models A$, then $M, s \models B$ or $M, s \models C$. By induction hypothesis, $M, s' \models B$ or $M, s' \models C$, so $M, s' \models A$.
4. $A \equiv B \rightarrow C$: exercise.
5. $A \equiv \exists x B$: if $M, s \models A$, there is an x -variant \bar{s} of s so that $M, \bar{s} \models B$. Let \bar{s}' denote the x -variant of s' that assigns the same thing to x as does \bar{s} : then by the induction hypothesis, $M, \bar{s}' \models B$. Hence, there is an x -variant of s' that satisfies B , so $M, s' \models A$.
6. $A \equiv \forall x B$: exercise.

By induction, we get that $M, s \models A$ iff $M, s' \models A$ whenever x_1, \dots, x_n are the only free variables in A and $s(x_i) = s'(x_i)$ for $i = 1, \dots, n$. \square

Definition 5.36. If A is a sentence, we say that a structure M satisfies A , $M \models A$, iff $M, s \models A$ for all variable assignments s .

If $M \models A$, we also say that A is true in M .

Proposition 5.37. Suppose $A(x)$ only contains x free, and M is a structure. Then:

1. $M \models \exists x A(x)$ iff $M, s \models A(x)$ for at least one variable assignment s .
2. $M \models \forall x A(x)$ iff $M, s \models A(x)$ for all variable assignments s .

Proof. Exercise. \square

5.10 Extensionality

Extensionality, sometimes called relevance, can be expressed informally as follows: the only thing that bears upon the satisfaction of formula A in a structure M relative to a variable assignment s , are the assignments made by M and s to the elements of the language that actually appear in A .

One immediate consequence of extensionality is that where two structures M and M' agree on all the elements of the language appearing in a sentence A and have the same domain, M and M' must also agree on A itself.

Proposition 5.38 (Extensionality). *Let A be a sentence, and M and M' be structures. If $c^M = c^{M'}$, $R^M = R^{M'}$, and $f^M = f^{M'}$ for every constant symbol c , relation symbol R , and function symbol f occurring in A , then $M \models A$ iff $M' \models A$.*

Moreover, the value of a term, and whether or not a structure satisfies a formula, only depends on the values of its subterms.

Proposition 5.39. *Let M be a structure, t and t' terms, and s a variable assignment. Let $s' \sim_x s$ be the x -variant of s given by $s'(x) = \text{Val}_s^M(t')$. Then $\text{Val}_s^M(t[t'/x]) = \text{Val}_{s'}^M(t)$.*

Proof. By induction on t .

1. If t is a constant, say, $t \equiv c$, then $t[t'/x] = c$, and $\text{Val}_s^M(c) = c^M = \text{Val}_{s'}^M(c)$.
2. If t is a variable other than x , say, $t \equiv y$, then $t[t'/x] = y$, and $\text{Val}_s^M(y) = \text{Val}_{s'}^M(y)$ since $s' \sim_x s$.
3. If $t \equiv x$, then $t[t'/x] = t'$. But $\text{Val}_s^M(x) = \text{Val}_s^M(t')$ by definition of s' .
4. If $t \equiv f(t_1, \dots, t_n)$ then we have:

$$\begin{aligned}
 \text{Val}_s^M(t[t'/x]) &= \text{Val}_s^M(f(t_1[t'/x], \dots, t_n[t'/x])) \text{ by definition of } t[t'/x] \\
 &= f^M(\text{Val}_s^M(t_1[t'/x]), \dots, \text{Val}_s^M(t_n[t'/x])) \text{ by definition of } \text{Val}_s^M(f(\dots)) \\
 &= f^M(\text{Val}_{s'}^M(t_1), \dots, \text{Val}_{s'}^M(t_n)) \text{ by induction hypothesis} \\
 &= \text{Val}_{s'}^M(t) \text{ by definition of } \text{Val}_{s'}^M(f(\dots))
 \end{aligned}$$

□

Proposition 5.40. *Let M be a structure, A a formula, t a term, and s a variable assignment. Let $s' \sim_x s$ be the x -variant of s given by $s'(x) = \text{Val}_s^M(t)$. Then $M, s \models A[t/x]$ iff $M, s' \models A$.*

Proof. Exercise. □

5.11 Semantic Notions

Give the definition of structures for first-order languages, we can define some basic semantic properties of and relationships between sentences. The simplest of these is the notion of *validity* of a sentence. A sentence is valid if it is satisfied in every structure. Valid sentences are those that are satisfied regardless of how the non-logical symbols in it are interpreted. Valid sentences are therefore also called *logical truths*—they are true, i.e., satisfied, in any structure and hence their truth depends only on the logical symbols occurring in them and their syntactic structure, but not on the non-logical symbols or their interpretation.

Definition 5.41 (Validity). A sentence A is *valid*, $\vDash A$, iff $M \models A$ for every structure M .

Definition 5.42 (Entailment). A set of sentences Γ *entails* a sentence A , $\Gamma \vDash A$, iff for every structure M with $M \models \Gamma$, $M \models A$.

Definition 5.43 (Satisfiability). A set of sentences Γ is *satisfiable* if $M \models \Gamma$ for some structure M . If Γ is not satisfiable it is called *unsatisfiable*.

Proposition 5.44. A sentence A is valid iff $\Gamma \vDash A$ for every set of sentences Γ .

Proof. For the forward direction, let A be valid, and let Γ be a set of sentences. Let M be a structure so that $M \models \Gamma$. Since A is valid, $M \models A$, hence $\Gamma \vDash A$.

For the contrapositive of the reverse direction, let A be invalid, so there is a structure M with $M \not\models A$. When $\Gamma = \{\top\}$, since \top is valid, $M \models \Gamma$. Hence, there is a structure M so that $M \models \Gamma$ but $M \not\models A$, hence Γ does not entail A . \square

Proposition 5.45. $\Gamma \vDash A$ iff $\Gamma \cup \{\neg A\}$ is unsatisfiable.

Proof. For the forward direction, suppose $\Gamma \vDash A$ and suppose to the contrary that there is a structure M so that $M \models \Gamma \cup \{\neg A\}$. Since $M \models \Gamma$ and $\Gamma \vDash A$, $M \models A$. Also, since $M \models \Gamma \cup \{\neg A\}$, $M \models \neg A$, so we have both $M \models A$ and $M \models \neg A$, a contradiction. Hence, there can be no such structure M , so $\Gamma \cup \{A\}$ is unsatisfiable.

For the reverse direction, suppose $\Gamma \cup \{\neg A\}$ is unsatisfiable. So for every structure M , either $M \not\models \Gamma$ or $M \models A$. Hence, for every structure M with $M \models \Gamma$, $M \models A$, so $\Gamma \vDash A$. \square

Proposition 5.46. If $\Gamma \subseteq \Gamma'$ and $\Gamma \vDash A$, then $\Gamma' \vDash A$.

Proof. Suppose that $\Gamma \subseteq \Gamma'$ and $\Gamma \vDash A$. Let M be such that $M \models \Gamma'$; then $M \models \Gamma$, and since $\Gamma \vDash A$, we get that $M \models A$. Hence, whenever $M \models \Gamma'$, $M \models A$, so $\Gamma' \vDash A$. \square

Theorem 5.47 (Semantic Deduction Theorem). $\Gamma \cup \{A\} \models B$ iff $\Gamma \models A \rightarrow B$.

Proof. For the forward direction, let $\Gamma \cup \{A\} \models B$ and let \mathbf{M} be a structure so that $\mathbf{M} \models \Gamma$. If $\mathbf{M} \models A$, then $\mathbf{M} \models \Gamma \cup \{A\}$, so since $\Gamma \cup \{A\}$ entails B , we get $\mathbf{M} \models B$. Therefore, $\mathbf{M} \models A \rightarrow B$, so $\Gamma \models A \rightarrow B$.

For the reverse direction, let $\Gamma \models A \rightarrow B$ and \mathbf{M} be a structure so that $\mathbf{M} \models \Gamma \cup \{A\}$. Then $\mathbf{M} \models \Gamma$, so $\mathbf{M} \models A \rightarrow B$, and since $\mathbf{M} \models A$, $\mathbf{M} \models B$. Hence, whenever $\mathbf{M} \models \Gamma \cup \{A\}$, $\mathbf{M} \models B$, so $\Gamma \cup \{A\} \models B$. \square

Chapter 6

Theories and Their Models

6.1 Introduction

The development of the axiomatic method is a significant achievement in the history of science, and is of special importance in the history of mathematics. An axiomatic development of a field involves the clarification of many questions: What is the field about? What are the most fundamental concepts? How are they related? Can all the concepts of the field be defined in terms of these fundamental concepts? What laws do, and must, these concepts obey?

The axiomatic method and logic were made for each other. Formal logic provides the tools for formulating axiomatic theories, for proving theorems from the axioms of the theory in a precisely specified way, for studying the properties of all systems satisfying the axioms in a systematic way.

Definition 6.1. A set of sentences Γ is *closed* iff, whenever $\Gamma \models A$ then $A \in \Gamma$. The *closure* of a set of sentences Γ is $\{A : \Gamma \models A\}$.

We say that Γ is *axiomatized* by a set of sentences Δ if Γ is the closure of Δ

We can think of an axiomatic theory as the set of sentences that is axiomatized by its set of axioms Δ . In other words, when we have a first-order language which contains non-logical symbols for the primitives of the axiomatically developed science we wish to study, together with a set of sentences that express the fundamental laws of the science, we can think of the theory as represented by all the sentences in this language that are entailed by the axioms. This ranges from simple examples with only a single primitive and simple axioms, such as the theory of partial orders, to complex theories such as Newtonian mechanics.

The important logical facts that make this formal approach to the axiomatic method so important are the following. Suppose Γ is an axiom system for a theory, i.e., a set of sentences.

6. THEORIES AND THEIR MODELS

1. We can state precisely when an axiom system captures an intended class of structures. That is, if we are interested in a certain class of structures, we will successfully capture that class by an axiom system Γ iff the structures are exactly those M such that $M \models \Gamma$.
2. We may fail in this respect because there are M such that $M \models \Gamma$, but M is not one of the structures we intend. This may lead us to add axioms which are not true in M .
3. If we are successful at least in the respect that Γ is true in all the intended structures, then a sentence A is true in all intended structures whenever $\Gamma \models A$. Thus we can use logical tools (such as proof methods) to show that sentences are true in all intended structures simply by showing that they are entailed by the axioms.
4. Sometimes we don't have intended structures in mind, but instead start from the axioms themselves: we begin with some primitives that we want to satisfy certain laws which we codify in an axiom system. One thing that we would like to verify right away is that the axioms do not contradict each other: if they do, there can be no concepts that obey these laws, and we have tried to set up an incoherent theory. We can verify that this doesn't happen by finding a model of Γ . And if there are models of our theory, we can use logical methods to investigate them, and we can also use logical methods to construct models.
5. The independence of the axioms is likewise an important question. It may happen that one of the axioms is actually a consequence of the others, and so is redundant. We can prove that an axiom A in Γ is redundant by proving $\Gamma \setminus \{A\} \models A$. We can also prove that an axiom is not redundant by showing that $(\Gamma \setminus \{A\}) \cup \{\neg A\}$ is satisfiable. For instance, this is how it was shown that the parallel postulate is independent of the other axioms of geometry.
6. Another important question is that of definability of concepts in a theory: The choice of the language determines what the models of a theory consists of. But not every aspect of a theory must be represented separately in its models. For instance, every ordering \leq determines a corresponding strict ordering $<$ —given one, we can define the other. So it is not necessary that a model of a theory involving such an order must *also* contain the corresponding strict ordering. When is it the case, in general, that one relation can be defined in terms of others? When is it impossible to define a relation in terms of other (and hence must add it to the primitives of the language)?

6.2 Expressing Properties of Structures

It is often useful and important to express conditions on functions and relations, or more generally, that the functions and relations in a structure satisfy these conditions. For instance, we would like to have ways of distinguishing those structures for a language which “capture” what we want the predicate symbols to “mean” from those that do not. Of course we’re completely free to specify which structures we “intend,” e.g., we can specify that the interpretation of the predicate symbol \leq must be an ordering, or that we are only interested in interpretations of \mathcal{L} in which the domain consists of sets and \in is interpreted by the “is an element of” relation. But can we do this with sentences of the language? In other words, which conditions on a structure M can we express by a sentence (or perhaps a set of sentences) in the language of M ? There are some conditions that we will not be able to express. For instance, there is no sentence of \mathcal{L}_A which is only true in a structure M if $|M| = \mathbb{N}$. We cannot express “the domain contains only natural numbers.” But there are “structural properties” of structures that we perhaps can express. Which properties of structures can we express by sentences? Or, to put it another way, which collections of structures can we describe as those making a sentence (or set of sentences) true?

Definition 6.2. Let Γ be a set of sentences in a language \mathcal{L} . We say that a structure M is a model of Γ if $M \models A$ for all $A \in \Gamma$.

Example 6.3. The sentence $\forall x x \leq x$ is true in M iff \leq^M is a reflexive relation. The sentence $\forall x \forall y ((x \leq y \wedge y \leq x) \rightarrow x = y)$ is true in M iff \leq^M is anti-symmetric. The sentence $\forall x \forall y \forall z ((x \leq y \wedge y \leq z) \rightarrow x \leq z)$ is true in M iff \leq^M is transitive. Thus, the models of

$$\left\{ \begin{array}{l} \forall x x \leq x, \\ \forall x \forall y ((x \leq y \wedge y \leq x) \rightarrow x = y), \\ \forall x \forall y \forall z ((x \leq y \wedge y \leq z) \rightarrow x \leq z) \end{array} \right\}$$

are exactly those structures in which \leq^M is reflexive, anti-symmetric, and transitive, i.e., a partial order. Hence, we can take them as axioms for the *first-order theory of partial orders*.

6.3 Examples of First-Order Theories

Example 6.4. The theory of strict linear orders in the language $\mathcal{L}_<$ is axiomatized by the set

$$\left\{ \begin{array}{l} \forall x \neg x < x, \\ \forall x \forall y ((x < y \vee y < x) \vee x = y), \\ \forall x \forall y \forall z ((x < y \wedge y < z) \rightarrow x < z) \end{array} \right\}$$

It completely captures the intended structures: every strict linear order is a model of this axiom system, and vice versa, if R is a linear order on a set X , then the structure \mathbf{M} with $|\mathbf{M}| = X$ and $<^{\mathbf{M}} = R$ is a model of this theory.

Example 6.5. The theory of groups in the language \mathcal{L}_1 (constant symbol), \cdot (two-place function symbol) is axiomatized by

$$\begin{aligned}\forall x (x \cdot 1) &= x \\ \forall x \forall y \forall z (x \cdot (y \cdot z)) &= ((x \cdot y) \cdot z) \\ \forall x \exists y (x \cdot y) &= 1\end{aligned}$$

Example 6.6. The theory of Peano arithmetic is axiomatized by the following sentences in the language of arithmetic \mathcal{L}_A .

$$\begin{aligned}\neg \exists x x' &= 0 \\ \forall x \forall y (x' = y' \rightarrow x = y) \\ \forall x \forall y (x < y \leftrightarrow \exists z (x + z' = y)) \\ \forall x (x + 0) &= x \\ \forall x \forall y (x + y') &= (x + y)' \\ \forall x (x \times 0) &= 0 \\ \forall x \forall y (x \times y') &= ((x \times y) + x)\end{aligned}$$

plus all sentences of the form

$$(A(0) \wedge \forall x (A(x) \rightarrow A(x'))) \rightarrow \forall x A(x)$$

Since there are infinitely many sentences of the latter form, this axiom system is infinite. The latter form is called the *induction schema*. (Actually, the induction schema is a bit more complicated than we let on here.)

The third axiom is an *explicit definition* of $<$.

Example 6.7. The theory of pure sets plays an important role in the foundations (and in the philosophy) of mathematics. A set is pure if all its elements are also pure sets. The empty set counts therefore as pure, but a set that has something as an element that is not a set would not be pure. So the pure sets are those that are formed just from the empty set and no “urelements,” i.e., objects that are not themselves sets.

The following might be considered as an axiom system for a theory of pure sets:

$$\begin{aligned}\exists x \neg \exists y y \in x \\ \forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y) \\ \forall x \forall y \exists z \forall u (u \in z \leftrightarrow (u = x \vee u = y)) \\ \forall x \exists y \forall z (z \in y \leftrightarrow \exists u (z \in u \vee u \in x))\end{aligned}$$

plus all sentences of the form

$$\exists x \forall y (y \in x \leftrightarrow A(y))$$

The first axiom says that there is a set with no elements (i.e., \emptyset exists); the second says that sets are extensional; the third that for any sets X and Y , the set $\{X, Y\}$ exists; the fourth that for any sets X and Y , the set $X \cup Y$ exists.

The sentences mentioned last are collectively called the *naive comprehension scheme*. It essentially says that for every $A(x)$, the set $\{x : A(x)\}$ exists—so at first glance a true, useful, and perhaps even necessary axiom. It is called “naive” because, as it turns out, it makes this theory unsatisfiable: if you take $A(y)$ to be $\neg y \in y$, you get the sentence

$$\exists x \forall y (y \in x \leftrightarrow \neg y \in y)$$

and this sentence is not satisfied in any structure.

Example 6.8. In the area of *mereology*, the relation of *parthood* is a fundamental relation. Just like theories of sets, there are theories of parthood that axiomatize various conceptions (sometimes conflicting) of this relation.

The language of mereology contains a single two-place predicate symbol P , and $P(x, y)$ “means” that x is a part of y . When we have this interpretation in mind, a structure for this language is called a *parthood structure*. Of course, not every structure for a single two-place predicate will really deserve this name. To have a chance of capturing “parthood,” P^M must satisfy some conditions, which we can lay down as axioms for a theory of parthood. For instance, parthood is a partial order on objects: every object is a part (albeit an *improper* part) of itself; no two different objects can be parts of each other; a part of a part of an object is itself part of that object. Note that in this sense “is a part of” resembles “is a subset of,” but does not resemble “is an element of” which is neither reflexive nor transitive.

$$\begin{aligned} &\forall x P(x, x), \\ &\forall x \forall y ((P(x, y) \wedge P(y, x)) \rightarrow x = y), \\ &\forall x \forall y \forall z ((P(x, y) \wedge P(y, z)) \rightarrow P(x, z)), \end{aligned}$$

Moreover, any two objects have a fusion (an object that has only these two objects and all their parts as parts).

$$\forall x \forall y \exists z \forall u (P(u, z) \leftrightarrow (P(u, x) \wedge P(u, y)))$$

These are only some of the basic principles of parthood considered by metaphysicians. Further principles, however, quickly become hard to formulate or write down without first introducing some defined relations. For instance, most metaphysicians interested in mereology also view the following as a valid principle: whenever an object x has a proper part y , it also has a part z that has no parts in common with y , and so that the fusion of y and z is x .

6.4 Expressing Relations in a Structure

One main use formulas can be put to is to express properties and relations in a structure M in terms of the primitives of the language \mathcal{L} of M . By this we mean the following: the domain of M is a set of objects. The constant symbols, function symbols, and predicate symbols are interpreted in M by some objects in $|M|$, functions on $|M|$, and relations on $|M|$. For instance, if A_0^2 is in \mathcal{L} , then M assigns to it a relation $R = A_0^{2M}$. Then the formula $A_0^2(x_1, x_2)$ expresses that very relation, in the following sense: if a variable assignment s maps x_1 to $a \in |M|$ and x_2 to $b \in |M|$, then

$$Rab \text{ iff } M, s \models A_0^2(x_1, x_2).$$

Note that we have to involve variable assignments here: we can't just say " Rab iff $M \models A_0^2(a, b)$ " because a and b are not symbols of our language: they are elements of $|M|$.

Since we don't just have atomic formulas, but can combine them using the logical connectives and the quantifiers, more complex formulas can define other relations which aren't directly built into M . We're interested in how to do that, and specifically, which relations we can define in a structure.

Definition 6.9. Let $A(x_1, \dots, x_n)$ be a formula of \mathcal{L} in which only x_1, \dots, x_n occur free, and let M be a structure for \mathcal{L} . $A(x_1, \dots, x_n)$ expresses the relation $R \subseteq |M|^n$ iff

$$Ra_1 \dots a_n \text{ iff } M, s \models A(x_1, \dots, x_n)$$

for any variable assignment s with $s(x_i) = a_i$ ($i = 1, \dots, n$).

Example 6.10. In the standard model of arithmetic N , the formula $x_1 < x_2 \vee x_1 = x_2$ expresses the \leq relation on \mathbb{N} . The formula $x_2 = x_1'$ expresses the successor relation, i.e., the relation $R \subseteq \mathbb{N}^2$ where Rnm holds if m is the successor of n . The formula $x_1 = x_2'$ expresses the predecessor relation. The formulas $\exists x_3 (x_3 \neq 0 \wedge x_2 = (x_1 + x_3))$ and $\exists x_3 (x_1 + x_3' = x_2)$ both express the $<$ relation. This means that the predicate symbol $<$ is actually superfluous in the language of arithmetic; it can be defined.

This idea is not just interesting in specific structures, but generally whenever we use a language to describe an intended model or models, i.e., when we consider theories. These theories often only contain a few predicate symbols as basic symbols, but in the domain they are used to describe often many other relations play an important role. If these other relations can be systematically expressed by the relations that interpret the basic predicate symbols of the language, we say we can *define* them in the language.

6.5 The Theory of Sets

Almost all of mathematics can be developed in the theory of sets. Developing mathematics in this theory involves a number of things. First, it requires a set of axioms for the relation \in . A number of different axiom systems have been developed, sometimes with conflicting properties of \in . The axiom system known as **ZFC**, Zermelo-Fraenkel set theory with the axiom of choice stands out: it is by far the most widely used and studied, because it turns out that its axioms suffice to prove almost all the things mathematicians expect to be able to prove. But before that can be established, it first is necessary to make clear how we can even *express* all the things mathematicians would like to express. For starters, the language contains no constant symbols or function symbols, so it seems at first glance unclear that we can talk about particular sets (such as \emptyset or \mathbb{N}), can talk about operations on sets (such as $X \cup Y$ and $\wp(X)$), let alone other constructions which involve things other than sets, such as relations and functions.

To begin with, “is an element of” is not the only relation we are interested in: “is a subset of” seems almost as important. But we can *define* “is a subset of” in terms of “is an element of.” To do this, we have to find a formula $A(x, y)$ in the language of set theory which is satisfied by a pair of sets $\langle X, Y \rangle$ iff $X \subseteq Y$. But X is a subset of Y just in case all elements of X are also elements of Y . So we can define \subseteq by the formula

$$\forall z (z \in x \rightarrow z \in y)$$

Now, whenever we want to use the relation \subseteq in a formula, we could instead use that formula (with x and y suitably replaced, and the bound variable z renamed if necessary). For instance, extensionality of sets means that if any sets x and y are contained in each other, then x and y must be the same set. This can be expressed by $\forall x \forall y ((x \subseteq y \wedge y \subseteq x) \rightarrow x = y)$, or, if we replace \subseteq by the above definition, by

$$\forall x \forall y ((\forall z (z \in x \rightarrow z \in y) \wedge \forall z (z \in y \rightarrow z \in x)) \rightarrow x = y).$$

This is in fact one of the axioms of **ZFC**, the “axiom of extensionality.”

There is no constant symbol for \emptyset , but we can express “ x is empty” by $\neg \exists y y \in x$. Then “ \emptyset exists” becomes the sentence $\exists x \neg \exists y y \in x$. This is another axiom of **ZFC**. (Note that the axiom of extensionality implies that there is only one empty set.) Whenever we want to talk about \emptyset in the language of set theory, we would write this as “there is a set that’s empty and ...” As an example, to express the fact that \emptyset is a subset of every set, we could write

$$\exists x (\neg \exists y y \in x \wedge \forall z x \subseteq z)$$

where, of course, $x \subseteq z$ would in turn have to be replaced by its definition.

To talk about operations on sets, such as $X \cup Y$ and $\wp(X)$, we have to use a similar trick. There are no function symbols in the language of set theory, but we can express the functional relations $X \cup Y = Z$ and $\wp(X) = Y$ by

$$\begin{aligned} \forall u ((u \in x \vee u \in y) \leftrightarrow u \in z) \\ \forall u (u \subseteq x \leftrightarrow u \in y) \end{aligned}$$

since the elements of $X \cup Y$ are exactly the sets that are either elements of X or elements of Y , and the elements of $\wp(X)$ are exactly the subsets of X . However, this doesn't allow us to use $x \cup y$ or $\wp(x)$ as if they were terms: we can only use the entire formulas that define the relations $X \cup Y = Z$ and $\wp(X) = Y$. In fact, we do not know that these relations are ever satisfied, i.e., we do not know that unions and power sets always exist. For instance, the sentence $\forall x \exists y \wp(x) = y$ is another axiom of **ZFC** (the power set axiom).

Now what about talk of ordered pairs or functions? Here we have to explain how we can think of ordered pairs and functions as special kinds of sets. One way to define the ordered pair $\langle x, y \rangle$ is as the set $\{\{x\}, \{x, y\}\}$. But like before, we cannot introduce a function symbol that names this set; we can only define the relation $\langle x, y \rangle = z$, i.e., $\{\{x\}, \{x, y\}\} = z$:

$$\forall u (u \in z \leftrightarrow (\forall v (v \in u \leftrightarrow v = x) \vee \forall v (v \in u \leftrightarrow (v = x \vee v = y))))$$

This says that the elements u of z are exactly those sets which either have x as its only element or have x and y as its only elements (in other words, those sets that are either identical to $\{x\}$ or identical to $\{x, y\}$). Once we have this, we can say further things, e.g., that $X \times Y = Z$:

$$\forall z (z \in Z \leftrightarrow \exists x \exists y (x \in X \wedge y \in Y \wedge \langle x, y \rangle = z))$$

A function $f: X \rightarrow Y$ can be thought of as the relation $f(x) = y$, i.e., as the set of pairs $\{\langle x, y \rangle : f(x) = y\}$. We can then say that a set f is a function from X to Y if (a) it is a relation $\subseteq X \times Y$, (b) it is total, i.e., for all $x \in X$ there is some $y \in Y$ such that $\langle x, y \rangle \in f$ and (c) it is functional, i.e., whenever $\langle x, y \rangle, \langle x, y' \rangle \in f$, $y = y'$ (because values of functions must be unique). So " f is a function from X to Y " can be written as:

$$\begin{aligned} \forall u (u \in f \rightarrow \exists x \exists y (x \in X \wedge y \in Y \wedge \langle x, y \rangle = u)) \wedge \\ \forall x (x \in X \rightarrow (\exists y (y \in Y \wedge \text{maps}(f, x, y)) \wedge \\ (\forall y \forall y' ((\text{maps}(f, x, y) \wedge \text{maps}(f, x, y')) \rightarrow y = y')))) \end{aligned}$$

where $\text{maps}(f, x, y)$ abbreviates $\exists v (v \in f \wedge \langle x, y \rangle = v)$ (this formula expresses " $f(x) = y$ ").

It is now also not hard to express that $f: X \rightarrow Y$ is injective, for instance:

$$\begin{aligned} f: X \rightarrow Y \wedge \\ \forall x \forall x' ((x \in X \wedge x' \in X \wedge \exists y (\text{maps}(f, x, y) \wedge \text{maps}(f, x', y))) \rightarrow x = x') \end{aligned}$$

A function $f: X \rightarrow Y$ is injective iff, whenever f maps $x, x' \in X$ to a single y , $x = x'$. If we abbreviate this formula as $\text{inj}(f, X, Y)$, we're already in a position to state in the language of set theory something as non-trivial as Cantor's theorem: there is no injective function from $\wp(X)$ to X :

$$\forall X \forall Y (\wp(X) = Y \rightarrow \neg \exists f \text{inj}(f, Y, X))$$

6.6 Expressing the Size of Structures

There are some properties of structures we can express even without using the non-logical symbols of a language. For instance, there are sentences which are true in a structure iff the domain of the structure has at least, at most, or exactly a certain number n of elements.

Proposition 6.11. *The sentence*

$$\begin{aligned} !A_{\geq n} \equiv \exists x_1 \exists x_2 \dots \exists x_n \quad & (x_1 \neq x_2 \wedge x_1 \neq x_3 \wedge x_1 \neq x_4 \wedge \dots \wedge x_1 \neq x_n \wedge \\ & x_2 \neq x_3 \wedge x_2 \neq x_4 \wedge \dots \wedge x_2 \neq x_n \wedge \\ & \vdots \\ & x_{n-1} \neq x_n) \end{aligned}$$

is true in a structure \mathbf{M} iff $|\mathbf{M}|$ contains at least n elements. Consequently, $\mathbf{M} \models \neg A_{\geq n+1}$ iff $|\mathbf{M}|$ contains at most n elements.

Proposition 6.12. *The sentence*

$$\begin{aligned} !A_{=n} \equiv \exists x_1 \exists x_2 \dots \exists x_n \quad & (x_1 \neq x_2 \wedge x_1 \neq x_3 \wedge x_1 \neq x_4 \wedge \dots \wedge x_1 \neq x_n \wedge \\ & x_2 \neq x_3 \wedge x_2 \neq x_4 \wedge \dots \wedge x_2 \neq x_n \wedge \\ & \vdots \\ & x_{n-1} \neq x_n \wedge \\ & \forall y (y = x_1 \vee \dots \vee y = x_n) \dots) \end{aligned}$$

is true in a structure \mathbf{M} iff $|\mathbf{M}|$ contains exactly n elements.

Proposition 6.13. *A structure is infinite iff it is a model of*

$$\{A_{\geq 1}, A_{\geq 2}, A_{\geq 3}, \dots\}$$

There is no single purely logical sentence which is true in \mathbf{M} iff $|\mathbf{M}|$ is infinite. However, one can give sentences with non-logical predicate symbols which only have infinite models (although not every infinite structure is a model of them). The property of being a finite structure, and the property of being an uncountable structure cannot even be expressed with an infinite set of sentences. These facts follow from the compactness and Löwenheim-Skolem theorems.

Chapter 7

Natural Deduction

7.1 Rules and Derivations

Let \mathcal{L} be a first-order language with the usual constants, variables, logical symbols, and auxiliary symbols (parentheses and the comma).

Definition 7.1 (Inference). An *inference* is an expression of the form

$$\frac{A}{C} \quad \text{or} \quad \frac{A \quad B}{C}$$

where A, B , and C are formulas. A and B are called the *upper formulas* or *premises* and C the *lower formulas* or *conclusion* of the inference.

The rules for natural deduction are divided into two main types: *propositional* rules (quantifier-free) and *quantifier* rules. The rules come in pairs, an introduction and an elimination rule for each logical operator. They introduced a logical operator in the conclusion or remove a logical operator from a premise of the rule. Some of the rules allow an assumption of a certain type to be *discharged*. To indicate which assumption is discharged by which inference, we also assign labels to both the assumption and the inference. This is indicated by writing the assumption formula as “[A]”.

It is customary to consider rules for all logical operators, even for those (if any) that we consider as defined.

Propositional Rules

Rules for \perp

$$\frac{A \quad \neg A}{\perp} \perp\text{Intro} \quad \frac{\perp}{A} \perp\text{Elim}$$

Rules for \wedge

$$\frac{A \quad B}{A \wedge B} \wedge\text{Intro} \quad \frac{A \wedge B}{A} \wedge\text{Elim} \quad \frac{A \wedge B}{B} \wedge\text{Elim}$$

Rules for \vee

$$\frac{A}{A \vee B} \vee\text{Intro} \quad \frac{B}{A \vee B} \vee\text{Intro} \quad \begin{array}{c} [A]^n \\ \vdots \\ C \\ \hline [B]^n \\ \vdots \\ C \end{array} \vee\text{Elim}$$

Rules for \neg

$$\begin{array}{c} [A]^n \\ \vdots \\ \perp \\ \hline \neg A \end{array} \neg\text{Intro} \quad \frac{\neg\neg A}{A} \neg\text{Elim}$$

Rules for \rightarrow

$$\begin{array}{c} [A]^n \\ \vdots \\ B \\ \hline A \rightarrow B \end{array} \rightarrow\text{Intro} \quad \frac{A \quad A \rightarrow B}{B} \rightarrow\text{Elim}$$

Quantifier Rules

Rules for \forall

$$\frac{A(a)}{\forall x A(x)} \forall\text{Intro} \quad \frac{\forall x A(x)}{A(t)} \forall\text{Elim}$$

where t is a ground term, and a is a constant which does not occur in A , or in any assumption which is undischarged in the derivation ending with the premise A . We call a the *eigenvariable* of the $\forall\text{Intro}$ inference.

Rules for \exists

$$\frac{A(a)}{\exists x A(x)} \exists\text{Intro} \quad \begin{array}{c} [A(a)]^n \\ \vdots \\ C \\ \hline \exists x A(x) \end{array} \exists\text{Elim}_n$$

where t is a ground term, and a is a constant which does not occur in the premise $\exists x A(x)$, in C , or any assumption which is undischarged in the derivations ending with the two premises C (other than the assumptions $A(a)$). We call a the *eigenvariable* of the $\exists\text{Elim}$ inference.

The condition that an eigenvariable not occur in the upper sequent of the \forall intro or \exists elim inference is called the *eigenvariable condition*.

We use the term “eigenvariable” even though a in the above rules is a constant. This has historical reasons.

In \exists Intro and \forall Elim there are no restrictions, and the term t can be anything, so we do not have to worry about any conditions. However, because the t may appear elsewhere in the derivation, the values of t for which the formula is satisfied are constrained. On the other hand, in the \exists Elim and \forall Intro rules, the eigenvariable condition requires that a does not occur anywhere else in the formula. Thus, if the upper formula is valid, the truth values of the formulas other than $A(a)$ are independent of a .

Natural deduction systems are meant to closely parallel the informal reasoning used in mathematical proof (hence it is somewhat “natural”). Natural deduction proofs begin with assumptions. Inference rules are then applied. Assumptions are “discharged” by the \neg Intro, \rightarrow Intro, \forall Elim and \exists Elim inference rules, and the label of the discharged assumption is placed beside the inference for clarity.

Definition 7.2 (Initial Formula). An *initial formula* or *assumption* is any formula in the topmost position of any branch.

Definition 7.3 (Derivation). A *derivation* of a formula A from assumptions Γ is a tree of formulas satisfying the following conditions:

1. The topmost formulas of the tree are either in Γ or are discharged by an inference in the tree.
2. Every formula in the tree is an upper formula of an inference whose lower formula stands directly below that formula in the tree.

We then say that A is the *end-formula* of the derivation and that A is *derivable* from Γ .

Definition 7.4 (Theorem). A sentence A is a *theorem* if it is derivable from the empty set.

7.2 Examples of Derivations

Example 7.5. Let’s give a derivation of the formula $(A \wedge B) \rightarrow A$.

We begin by writing the desired end-formula at the bottom of the derivation.

$$\overline{(A \wedge B) \rightarrow A}$$

Next, we need to figure out what kind of inference could result in a formula of this form. The main operator of the end-formula is \rightarrow , so we’ll try to arrive at the end-formula using the \rightarrow Intro rule. It is best to write down the assumptions involved and label the inference rules as you progress, so it

is easy to see whether all assumptions have been discharged at the end of the proof.

$$\begin{array}{c} [A \wedge B]^1 \\ \vdots \\ \vdots \\ A \\ 1 \frac{\quad}{(A \wedge B) \rightarrow A} \rightarrow \text{Intro} \end{array}$$

We now need to fill in the steps from the assumption $A \wedge B$ to A . Since we only have one connective to deal with, \wedge , we must use the \wedge elim rule. This gives us the following proof:

$$\begin{array}{c} [A \wedge B]^1 \\ \frac{\quad}{A} \wedge \text{Elim} \\ 1 \frac{\quad}{(A \wedge B) \rightarrow A} \rightarrow \text{Intro} \end{array}$$

We now have a correct derivation of the formula $(A \wedge B) \rightarrow A$.

Example 7.6. Now let's give a derivation of the formula $(\neg A \vee B) \rightarrow (A \rightarrow B)$.

We begin by writing the desired end-formula at the bottom of the derivation.

$$\overline{(\neg A \vee B) \rightarrow (A \rightarrow B)}$$

To find a logical rule that could give us this end-formula, we look at the logical connectives in the end-formula: \neg , \vee , and \rightarrow . We only care at the moment about the first occurrence of \rightarrow because it is the main operator of the sentence in the end-sequent, while \neg , \vee and the second occurrence of \rightarrow are inside the scope of another connective, so we will take care of those later. We therefore start with the \rightarrow Intro rule. A correct application must look as follows:

$$\begin{array}{c} [\neg A \vee B]^1 \\ \vdots \\ \vdots \\ A \rightarrow B \\ 1 \frac{\quad}{(\neg A \vee B) \rightarrow (A \rightarrow B)} \rightarrow \text{Intro} \end{array}$$

This leaves us with two possibilities to continue. Either we can keep working from the bottom up and look for another application of the \rightarrow Intro rule, or we can work from the top down and apply a \vee Elim rule. Let us apply the latter. We will use the assumption $\neg A \vee B$ as the leftmost premise of \vee Elim. For a valid application of \vee Elim, the other two premises must be identical to the conclusion $A \rightarrow B$, but each may be derived in turn from another assumption, namely the two disjuncts of $\neg A \vee B$. So our derivation will look like

this:

$$\begin{array}{c}
 \begin{array}{ccc}
 & [\neg A]^2 & [B]^2 \\
 & \vdots & \vdots \\
 & A \rightarrow B & A \rightarrow B \\
 2 \frac{[\neg A \vee B]^1}{A \rightarrow B} & & \vee\text{Elim} \\
 1 \frac{A \rightarrow B}{(\neg A \vee B) \rightarrow (A \rightarrow B)} & \rightarrow \text{Intro} &
 \end{array}
 \end{array}$$

In each of the two branches on the right, we want to derive $A \rightarrow B$, which is best done using \rightarrow Intro.

$$\begin{array}{c}
 \begin{array}{ccc}
 & [\neg A]^2, [A]^3 & [B]^2, [A]^4 \\
 & \vdots & \vdots \\
 & B & B \\
 2 \frac{[\neg A \vee B]^1}{A \rightarrow B} & 3 \frac{B}{A \rightarrow B} \rightarrow \text{Intro} & 4 \frac{B}{A \rightarrow B} \rightarrow \text{Intro} \\
 1 \frac{A \rightarrow B}{(\neg A \vee B) \rightarrow (A \rightarrow B)} & \rightarrow \text{Intro} & \vee\text{Elim}
 \end{array}
 \end{array}$$

For the two missing parts of the derivation, we need derivations of B from $\neg A$ and A in the middle, and from A and B on the left. Let's take the former first. $\neg A$ and A are the two premises of \perp Intro:

$$\begin{array}{c}
 \frac{[\neg A]^2 \quad [A]^3}{\perp} \perp \text{Intro} \\
 \vdots \\
 B
 \end{array}$$

By using \perp Elim, we can obtain B as a conclusion and complete the branch.

$$\begin{array}{c}
 \begin{array}{ccc}
 & [\neg A]^2 \quad [A]^3 & [B]^2, [A]^4 \\
 & \frac{\perp}{\perp} \perp \text{Intro} & \vdots \\
 & \frac{\perp}{B} \perp \text{Elim} & \vdots \\
 & 3 \frac{B}{A \rightarrow B} \rightarrow \text{Intro} & 4 \frac{B}{A \rightarrow B} \rightarrow \text{Intro} \\
 2 \frac{[\neg A \vee B]^1}{A \rightarrow B} & & \vee\text{Elim} \\
 1 \frac{A \rightarrow B}{(\neg A \vee B) \rightarrow (A \rightarrow B)} & \rightarrow \text{Intro} &
 \end{array}
 \end{array}$$

Let's now look at the rightmost branch. Here it's important to realize that the definition of derivation *allows assumptions to be discharged* but *does not require* them to be. In other words, if we can derive B from one of the assumptions A and B without using the other, that's ok. And to derive B from B is trivial: B by itself is such a derivation, and no inferences are needed. So we

7. NATURAL DEDUCTION

can simply delete the assumption A .

$$\begin{array}{c}
 \frac{[\neg A]^2 \quad [A]^3}{\perp} \perp\text{Intro} \\
 \frac{\perp}{B} \perp\text{Elim} \\
 \frac{A \rightarrow B}{A \rightarrow B} \rightarrow\text{Intro} \\
 \frac{[B]^2}{A \rightarrow B} \rightarrow\text{Intro} \\
 \frac{[\neg A \vee B]^1 \quad \frac{A \rightarrow B}{A \rightarrow B} \rightarrow\text{Intro} \quad \frac{[B]^2}{A \rightarrow B} \rightarrow\text{Intro}}{A \rightarrow B} \vee\text{Elim} \\
 \frac{A \rightarrow B}{(\neg A \vee B) \rightarrow (A \rightarrow B)} \rightarrow\text{Intro}
 \end{array}$$

Note that in the finished derivation, the rightmost \rightarrow Intro inference does not actually discharge any assumptions.

Example 7.7. When dealing with quantifiers, we have to make sure not to violate the eigenvariable condition, and sometimes this requires us to play around with the order of carrying out certain inferences. In general, it helps to try and take care of rules subject to the eigenvariable condition first (they will be lower down in the finished proof).

Let's see how we'd give a derivation of the formula $\exists x \neg A(x) \rightarrow \neg \forall x A(x)$. Starting as usual, we write

$$\overline{\exists x \neg A(x) \rightarrow \neg \forall x A(x)}$$

We start by writing down what it would take to justify that last step using the \rightarrow Intro rule.

$$\begin{array}{c}
 [\exists x \neg A(x)]^1 \\
 \vdots \\
 \neg \forall x A(x) \\
 \hline
 \exists x \neg A(x) \rightarrow \neg \forall x A(x) \quad \rightarrow\text{Intro}
 \end{array}$$

Since there is no obvious rule to apply to $\neg \forall x A(x)$, we will proceed by setting up the derivation so we can use the \exists Elim rule. Here we must pay attention to the eigenvariable condition, and choose a constant that does not appear in $\exists x A(x)$ or any assumptions that it depends on. (Since no constant symbols appear, however, any choice will do fine.)

$$\begin{array}{c}
 [\neg A(a)]^2 \\
 \vdots \\
 \neg \forall x A(x) \\
 \frac{[\exists x \neg A(x)]^1 \quad \neg \forall x A(x)}{\neg \forall x A(x)} \exists\text{Elim} \\
 \hline
 \exists x \neg A(x) \rightarrow \neg \forall x A(x) \quad \rightarrow\text{Intro}
 \end{array}$$

In order to derive $\neg \forall x A(x)$, we will attempt to use the \neg Intro rule: this requires that we derive a contradiction, possibly using $\forall x A(x)$ as an additional assumption. Of course, this contradiction may involve the assumption $\neg A(a)$ which will be discharged by the \rightarrow Intro inference. We can set it

up as follows:

$$\begin{array}{c}
 [\neg A(a)]^2, [\forall x A(x)]^3 \\
 \vdots \\
 \perp \\
 \frac{3 \frac{\perp}{\neg \forall x A(x)} \text{ } \neg\text{-Intro}}{2 \frac{[\exists x \neg A(x)]^1}{\neg \forall x A(x)} \text{ } \exists\text{-Elim}}{\exists x \neg A(x) \rightarrow \neg \forall x A(x)} \rightarrow \text{Intro}
 \end{array}$$

It looks like we are close to getting a contradiction. The easiest rule to apply is the \forall Elim, which has no eigenvariable conditions. Since we can use any term we want to replace the universally quantified x , it makes the most sense to continue using a so we can reach a contradiction.

$$\begin{array}{c}
 \frac{[\neg A(a)]^2 \quad \frac{[\forall x A(x)]^3}{A(a)} \text{ } \forall\text{-Elim}}{\perp} \perp\text{-Intro} \\
 \frac{3 \frac{\perp}{\neg \forall x A(x)} \text{ } \neg\text{-Intro}}{2 \frac{[\exists x \neg A(x)]^1}{\neg \forall x A(x)} \text{ } \exists\text{-Elim}}{\exists x \neg A(x) \rightarrow \neg \forall x A(x)} \rightarrow \text{Intro}
 \end{array}$$

It is important, especially when dealing with quantifiers, to double check at this point that the eigenvariable condition has not been violated. Since the only rule we applied that is subject to the eigenvariable condition was \exists Elim, and the eigenvariable a does not occur in any assumptions it depends on, this is a correct derivation.

7.3 Proof-Theoretic Notions

Just as we've defined a number of important semantic notions (validity, entailment, satisfiability), we now define corresponding *proof-theoretic notions*. These are not defined by appeal to satisfaction of sentences in structures, but by appeal to the derivability or non-derivability of certain formulas. It was an important discovery, due to Gödel, that these notions coincide. That they do is the content of the *completeness theorem*.

Definition 7.8 (Derivability). A formula A is *derivable* from a set of formulas Γ , $\Gamma \vdash A$, if there is a derivation with end-formula A and in which every assumption is either discharged or is in Γ . If A is not derivable from Γ we write $\Gamma \not\vdash A$.

Definition 7.9 (Theorems). A formula A is a *theorem* if there is a derivation of A from the empty set. We write $\vdash A$ if A is a theorem and $\not\vdash A$ if it is not.

7. NATURAL DEDUCTION

Definition 7.10 (Consistency). A set of sentences Γ is *consistent* iff $\Gamma \not\vdash \perp$. If Γ is not consistent, i.e., if $\Gamma \vdash \perp$, we say it is *inconsistent*.

Proposition 7.11. $\Gamma \vdash A$ iff $\Gamma \cup \{\neg A\}$ is inconsistent.

Proof. Exercise. □

Proposition 7.12. Γ is inconsistent iff $\Gamma \vdash A$ for every sentence A .

Proof. Exercise. □

Proposition 7.13. If $\Gamma \vdash A$ iff for some finite $\Gamma_0 \subseteq \Gamma$, $\Gamma_0 \vdash A$.

Proof. Any derivation of A from Γ can only contain finitely many undischarged assumptions. If all these undischarged assumptions are in Γ , then the set of them is a finite subset of Γ . The other direction is trivial, since a derivation from a subset of Γ is also a derivation from Γ . □

7.4 Properties of Derivability

We will now establish a number of properties of the derivability relation. They are independently interesting, but each will play a role in the proof of the completeness theorem.

Proposition 7.14 (Monotony). If $\Gamma \subseteq \Delta$ and $\Gamma \vdash A$, then $\Delta \vdash A$.

Proof. Any derivation of A from Γ is also a derivation of A from Δ . □

Proposition 7.15. If $\Gamma \vdash A$ and $\Gamma \cup \{A\} \vdash \perp$, then Γ is inconsistent.

Proof. Let the derivation of A from Γ be δ_1 and the derivation of \perp from $\Gamma \cup \{A\}$ be δ_2 . We can then derive:

$$\begin{array}{c}
 [A]^1 \\
 \vdots \\
 \vdots \delta_2 \\
 \vdots \\
 \vdots \delta_1 \\
 \vdots \\
 A \quad 1 \frac{\perp}{\neg A} \neg\text{Intro} \\
 \hline
 \perp \quad \neg\text{Elim}
 \end{array}$$

In the new derivation, the assumption A is discharged, so it is a derivation from Γ . □

Proposition 7.16. If $\Gamma \cup \{A\} \vdash \perp$, then $\Gamma \vdash \neg A$.

Proof. Suppose that $\Gamma \cup \{A\} \vdash \perp$. Then there is a derivation of \perp from $\Gamma \cup \{A\}$. Let δ be the derivation of \perp , and consider

$$\frac{\begin{array}{c} [A]^1 \\ \vdots \\ \delta \\ \vdots \\ \perp \end{array}}{\neg A} \neg\text{Intro}$$

□

Proposition 7.17. *If $\Gamma \cup \{A\} \vdash \perp$ and $\Gamma \cup \{\neg A\} \vdash \perp$, then $\Gamma \vdash \perp$.*

Proof. There are derivations δ_1 and δ_2 of \perp from $\Gamma \cup \{A\}$ and \perp from $\Gamma \cup \{\neg A\}$, respectively. We can then derive

$$\frac{\frac{\begin{array}{c} [A]^1 \\ \vdots \\ \delta_1 \\ \vdots \\ \perp \end{array}}{\neg A} \neg\text{Intro} \quad \frac{\begin{array}{c} [\neg A]^2 \\ \vdots \\ \delta_2 \\ \vdots \\ \perp \end{array}}{\neg\neg A} \neg\text{Intro}}{\perp} \neg\text{Elim}$$

Since the assumptions A and $\neg A$ are discharged, this is a derivation from Γ alone. Hence $\Gamma \vdash \perp$. □

Proposition 7.18. *If $\Gamma \cup \{A\} \vdash \perp$ and $\Gamma \cup \{B\} \vdash \perp$, then $\Gamma \cup \{A \vee B\} \vdash \perp$.*

Proof. Exercise. □

Proposition 7.19. *If $\Gamma \vdash A$ or $\Gamma \vdash B$, then $\Gamma \vdash A \vee B$.*

Proof. Suppose $\Gamma \vdash A$. There is a derivation δ of A from Γ . We can derive

$$\frac{\begin{array}{c} \vdots \\ \delta \\ \vdots \\ A \end{array}}{A \vee B} \vee\text{Intro}$$

Therefore $\Gamma \vdash A \vee B$. The proof for when $\Gamma \vdash B$ is similar. □

Proposition 7.20. *If $\Gamma \vdash A \wedge B$ then $\Gamma \vdash A$ and $\Gamma \vdash B$.*

Proof. Exercise □

Proposition 7.21. *If $\Gamma \vdash A$ and $\Gamma \vdash B$, then $\Gamma \vdash A \wedge B$.*

Proof. Exercise. □

Proposition 7.22. *If $\Gamma \vdash A$ and $\Gamma \vdash A \rightarrow B$, then $\Gamma \vdash B$.*

Proof. Exercise. □

Proposition 7.23. *If $\Gamma \vdash \neg A$ or $\Gamma \vdash B$, then $\Gamma \vdash A \rightarrow B$.*

Proof. Exercise. □

Theorem 7.24. *If c is a constant not occurring in Γ or $A(x)$ and $\Gamma \vdash A(c)$, then $\Gamma \vdash \forall x A(x)$.*

Proof. Let δ be an derivation of $A(c)$ from Γ . By adding a \forall Intro inference, we obtain a proof of $\forall x A(x)$. Since c does not occur in Γ or $A(x)$, the eigenvariable condition is satisfied. □

Theorem 7.25. 1. *If $\Gamma \vdash A(t)$ then $\Gamma \vdash \exists x A(x)$.*

2. *If $\Gamma \vdash \forall x A(x)$ then $\Gamma \vdash A(t)$.*

Proof. 1. Suppose $\Gamma \vdash A(t)$. Then there is a derivation δ of $A(t)$ from Γ .
The derivation

$$\frac{\begin{array}{c} \vdots \\ \vdots \delta \\ \vdots \\ A(t) \end{array}}{\exists x A(x)} \exists\text{Intro}$$

shows that $\Gamma \vdash \exists x A(x)$.

2. Suppose $\Gamma \vdash \forall x A(x)$. Then there is a derivation δ of $\forall x A(x)$ from Γ .
The derivation

$$\frac{\begin{array}{c} \vdots \\ \vdots \delta \\ \vdots \\ \forall x A(x) \end{array}}{A(t)} \forall\text{Elim}$$

shows that $\Gamma \vdash A(t)$. □

7.5 Soundness

A derivation system, such as natural deduction, is *sound* if it cannot derive things that do not actually follow. Soundness is thus a kind of guaranteed safety property for derivation systems. Depending on which proof theoretic property is in question, we would like to know for instance, that

1. every derivable sentence is valid;

2. if a sentence is derivable from some others, it is also a consequence of them;
3. if a set of sentences is inconsistent, it is unsatisfiable.

These are important properties of a derivation system. If any of them do not hold, the derivation system is deficient—it would derive too much. Consequently, establishing the soundness of a derivation system is of the utmost importance.

Theorem 7.26 (Soundness). *If A is derivable from the undischarged assumptions Γ , then $\Gamma \models A$.*

Proof. Inductive Hypothesis: The premises of an inference rule follow from the undischarged assumptions of the subproofs ending in those premises.

Inductive Step: Show that A follows from the undischarged assumptions of the entire proof.

Let δ be a derivation of A . We proceed by induction on the number of inferences in δ .

If the number of inferences is 0, then δ consists only of an initial formula. Every initial formula A is an undischarged assumption, and as such, any structure \mathbf{M} that satisfies all of the undischarged assumptions of the proof also satisfies A .

If the number of inferences is greater than 0, we distinguish cases according to the type of the lowermost inference. By induction hypothesis, we can assume that the premises of that inference follow from the undischarged assumptions of the sub-derivations ending in those premises, respectively.

First, we consider the possible inferences with only one premise.

1. Suppose that the last inference is \neg -Intro: By inductive hypothesis, \perp follows from the undischarged assumptions $\Gamma \cup \{A\}$. Consider a structure \mathbf{M} . We need to show that, if $\mathbf{M} \models \Gamma$, then $\mathbf{M} \models \neg A$. Suppose for reductio that $\mathbf{M} \models \Gamma$, but $\mathbf{M} \not\models \neg A$, i.e., $\mathbf{M} \models A$. This would mean that $\mathbf{M} \models \Gamma \cup \{A\}$. This is contrary to our inductive hypothesis. So, $\mathbf{M} \models \neg A$.
2. The last inference is \neg -Elim: Exercise.
3. The last inference is \wedge -Elim: There are two variants: A or B may be inferred from the premise $A \wedge B$. Consider the first case. By inductive hypothesis, $A \wedge B$ follows from the undischarged assumptions Γ . Consider a structure \mathbf{M} . We need to show that, if $\mathbf{M} \models \Gamma$, then $\mathbf{M} \models A$. By our inductive hypothesis, we know that $\mathbf{M} \models A \wedge B$. So, $\mathbf{M} \models A$. The case where B is inferred from $A \wedge B$ is handled similarly.

4. The last inference is \vee Intro: There are two variants: $A \vee B$ may be inferred from the premise A or the premise B . Consider the first case. By inductive hypothesis, A follows from the undischarged assumptions Γ . Consider a structure M . We need to show that, if $M \models \Gamma$, then $M \models A \vee B$. Since $M \models \Gamma$, it must be the case that $M \models A$, by inductive hypothesis. So it must also be the case that $M \models A \vee B$. The case where $A \vee B$ is inferred from B is handled similarly.
5. The last inference is \rightarrow Intro: $A \rightarrow B$ is inferred from a subproof with assumption A and conclusion B . By inductive hypothesis, B follows from the undischarged assumptions Γ and A . Consider a structure M . We need to show that, if $\Gamma \models A \rightarrow B$. For reductio, suppose that for some structure M , $M \models \Gamma$ but $M \not\models A \rightarrow B$. So, $M \models A$ and $M \not\models B$. But by hypothesis, B is a consequence of $\Gamma \cup \{A\}$. So, $M \models A \rightarrow B$.
6. The last inference is \forall Intro: The premise $A(a)$ is a consequence of the undischarged assumptions Γ by induction hypothesis. Consider some structure, M , such that $M \models \Gamma$. Let M' be exactly like M except that $a^M \neq a^{M'}$. We must have $M' \models A(a)$.
We now show that $M \models \forall x A(x)$. Since $\forall x A(x)$ is a sentence, this means we have to show that for every variable assignment s , $M, s \models A(x)$. Since Γ consists entirely of sentences, $M, s \models B$ for all $B \in \Gamma$. Let M' be like M except that $a^{M'} = s(x)$. Then $M, s \models A(x)$ iff $M' \models A(a)$ (as $A(x)$ does not contain a). Since a also does not occur in Γ , $M' \models \Gamma$. Since $\Gamma \models A(a)$, $M' \models A(a)$. This means that $M, s \models A(x)$. Since s is an arbitrary variable assignment, $M \models \forall x A(x)$.
7. The last inference is \exists Intro: Exercise.
8. The last inference is \forall Elim: Exercise.

Now let's consider the possible inferences with several premises: \vee Elim, \wedge Intro, \rightarrow Elim, and \exists Elim.

1. The last inference is \wedge Intro. $A \wedge B$ is inferred from the premises A and B . By induction hypothesis, A follows from the undischarged assumptions Γ and B follows from the undischarged assumptions Δ . We have to show that $\Gamma \cup \Delta \models A \wedge B$. Consider a structure M with $M \models \Gamma \cup \Delta$. Since $M \models \Gamma$, it must be the case that $M \models A$, and since $M \models \Delta$, $M \models B$, by inductive hypothesis. Together, $M \models A \wedge B$.
2. The last inference is \vee Elim: Exercise.
3. The last inference is \rightarrow Elim. B is inferred from the premises $A \rightarrow B$ and A . By induction hypothesis, $A \rightarrow B$ follows from the undischarged assumptions Γ and A follows from the undischarged assumptions Δ .

Consider a structure M . We need to show that, if $M \models \Gamma \cup \Delta$, then $M \models B$. It must be the case that $M \models A \rightarrow B$, and $M \models A$, by inductive hypothesis. Thus it must be the case that $M \models B$.

4. The last inference is \exists Elim: Exercise. □

Corollary 7.27. *If $\vdash A$, then A is valid.*

Corollary 7.28. *If Γ is satisfiable, then it is consistent.*

Proof. We prove the contrapositive. Suppose that Γ is not consistent. Then $\Gamma \vdash \perp$, i.e., there is a derivation of \perp from undischarged assumptions in Γ . By [Theorem 7.26](#), any structure M that satisfies Γ must satisfy \perp . Since $M \not\models \perp$ for every structure M , no M can satisfy Γ , i.e., Γ is not satisfiable. □

7.6 Derivations with Identity predicate

Derivations with the identity predicate require additional inference rules.

Rules for =:

$$\frac{}{t = t} = \text{Intro} \quad \frac{t_1 = t_2 \quad A(t_1)}{A(t_2)} = \text{Elim} \quad \text{and} \quad \frac{t_1 = t_2 \quad A(t_2)}{A(t_1)} = \text{Elim}$$

where t_1 and t_2 are closed terms. The = Intro rule allows us to derive any identity statement of the form $t = t$ outright.

Example 7.29. If s and t are closed terms, then $A(s), s = t \vdash A(t)$:

$$\frac{A(s) \quad s = t}{A(t)} = \text{Elim}$$

This may be familiar as the “principle of substitutability of identicals,” or Leibniz’ Law.

Proposition 7.30. *Natural deduction with rules for identity is sound.*

Proof. Any formula of the form $t = t$ is valid, since for every structure M , $M \models t = t$. (Note that we assume the term t to be ground, i.e., it contains no variables, so variable assignments are irrelevant).

Suppose the last inference in a derivation is = Elim. Then the premises are $t_1 = t_2$ and $A(t_1)$; they are derived from undischarged assumptions Γ and Δ , respectively. We want to show that $A(s)$ follows from $\Gamma \cup \Delta$. Consider a structure M with $M \models \Gamma \cup \Delta$. By induction hypothesis, M satisfies the two premises by induction hypothesis. So, $M \models t_1 = t_2$. Therefore, $\text{Val}^M(t_1) = \text{Val}^M(t_2)$. Let s be any variable assignment, and s' be the x -variant given by

7. NATURAL DEDUCTION

$s'(x) = \text{Val}^M(t_1) = \text{Val}^M(t_2)$. By [Proposition 5.40](#), $M, s \models A(t_2)$ iff $M, s' \models A(x)$ iff $M, s \models A(t_1)$. Since $M \models A(t_1)$ therefore $M \models A(t_2)$. \square

Chapter 8

The Completeness Theorem

8.1 Introduction

The completeness theorem is one of the most fundamental results about logic. It comes in two formulations, the equivalence of which we'll prove. In its first formulation it says something fundamental about the relationship between semantic consequence and our proof system: if a sentence A follows from some sentences Γ , then there is also a derivation that establishes $\Gamma \vdash A$. Thus, the proof system is as strong as it can possibly be without proving things that don't actually follow. In its second formulation, it can be stated as a model existence result: every consistent set of sentences is satisfiable.

These aren't the only reasons the completeness theorem—or rather, its proof—is important. It has a number of important consequences, some of which we'll discuss separately. For instance, since any derivation that shows $\Gamma \vdash A$ is finite and so can only use finitely many of the sentences in Γ , it follows by the completeness theorem that if A is a consequence of Γ , it is already a consequence of a finite subset of Γ . This is called *compactness*. Equivalently, if every finite subset of Γ is consistent, then Γ itself must be consistent. It also follows from *the proof of* the completeness theorem that any satisfiable set of sentences has a finite or countably infinite model. This result is called the Löwenheim-Skolem theorem.

8.2 Outline of the Proof

The proof of the completeness theorem is a bit complex, and upon first reading it, it is easy to get lost. So let us outline the proof. The first step is a shift of perspective, that allows us to see a route to a proof. When completeness is thought of as “whenever $\Gamma \models A$ then $\Gamma \vdash A$,” it may be hard to even come up with an idea: for to show that $\Gamma \vdash A$ we have to find a derivation, and it does not look like the hypothesis that $\Gamma \models A$ helps us for this in any way. For some proof systems it is possible to directly construct a derivation, but we will take

a slightly different tack. The shift in perspective required is this: completeness can also be formulated as: “if Γ is consistent, it has a model.” Perhaps we can use the information in Γ together with the hypothesis that it is consistent to construct a model. After all, we know what kind of model we are looking for: one that is as Γ describes it!

If Γ contains only atomic sentences, it is easy to construct a model for it: for atomic sentences are all of the form $P(a_1, \dots, a_n)$ where the a_i are constant symbols. So all we have to do is come up with a domain $|M|$ and an interpretation for P so that $M \models P(a_1, \dots, a_n)$. But nothing’s easier than that: put $|M| = \mathbb{N}$, $c_i^M = i$, and for every $P(a_1, \dots, a_n) \in \Gamma$, put the tuple $\langle k_1, \dots, k_n \rangle$ into P^M , where k_i is the index of the constant symbol a_i (i.e., $a_i \equiv c_{k_i}$).

Now suppose Γ contains some sentence $\neg B$, with B atomic. We might worry that the construction of M interferes with the possibility of making $\neg B$ true. But here’s where the consistency of Γ comes in: if $\neg B \in \Gamma$, then $B \notin \Gamma$, or else Γ would be inconsistent. And if $B \notin \Gamma$, then according to our construction of M , $M \not\models B$, so $M \models \neg B$. So far so good.

Now what if Γ contains complex, non-atomic formulas? Say, it contains $A \wedge B$. Then we should proceed as if both A and B were in Γ . And if $A \vee B \in \Gamma$, then we will have to make at least one of them true, i.e., proceed as if one of them was in Γ .

This suggests the following idea: we add additional sentences to Γ so as to (a) keep the resulting set consistent and (b) make sure that for every possible atomic sentence A , either A is in the resulting set, or $\neg A$, and (c) such that, whenever $A \wedge B$ is in the set, so are both A and B , if $A \vee B$ is in the set, at least one of A or B is also, etc. We keep doing this (potentially forever). Call the set of all sentences so added Γ^* . Then our construction above would provide us with a structure for which we could prove, by induction, that all sentences in Γ^* are true in M , and hence also all sentence in Γ since $\Gamma \subseteq \Gamma^*$.

There is one wrinkle in this plan: if $\exists x A(x) \in \Gamma$ we would hope to be able to pick some constant symbol c and add $A(c)$ in this process. But how do we know we can always do that? Perhaps we only have a few constant symbols in our language, and for each one of them we have $\neg B(c) \in \Gamma$. We can’t also add $B(c)$, since this would make the set inconsistent, and we wouldn’t know whether M has to make $B(c)$ or $\neg B(c)$ true. Moreover, it might happen that Γ contains only sentences in a language that has no constant symbols at all (e.g., the language of set theory).

The solution to this problem is to simply add infinitely many constants at the beginning, plus sentences that connect them with the quantifiers in the right way. (Of course, we have to verify that this cannot introduce an inconsistency.)

Our original construction works well if we only have constant symbols in the atomic sentences. But the language might also contain function symbols. In that case, it might be tricky to find the right functions on \mathbb{N} to assign to

these function symbols to make everything work. So here's another trick: instead of using i to interpret c_i , just take the set of constant symbols itself as the domain. Then M can assign every constant symbol to itself: $c_i^M = c_i$. But why not go all the way: let $|M|$ be all *terms* of the language! If we do this, there is an obvious assignment of functions (that take terms as arguments and have terms as values) to function symbols: we assign to the function symbol f_i^n the function which, given n terms t_1, \dots, t_n as input, produces the term $f_i^n(t_1, \dots, t_n)$ as value.

The last piece of the puzzle is what to do with $=$. The predicate symbol $=$ has a fixed interpretation: $M \models t = t'$ iff $\text{Val}^M(t) = \text{Val}^M(t')$. Now if we set things up so that the value of a term t is t itself, then this structure will make *no* sentence of the form $t = t'$ true unless t and t' are one and the same term. And of course this is a problem, since basically every interesting theory in a language with function symbols will have as theorems sentences $t = t'$ where t and t' are not the same term (e.g., in theories of arithmetic: $(o + o) = o$). To solve this problem, we change the domain of M : instead of using terms as the objects in $|M|$, we use sets of terms, and each set is so that it contains all those terms which the sentences in Γ require to be equal. So, e.g., if Γ is a theory of arithmetic, one of these sets will contain: $o, (o + o), (o \times o)$, etc. This will be the set we assign to o , and it will turn out that this set is also the value of all the terms in it, e.g., also of $(o + o)$. Therefore, the sentence $(o + o) = o$ will be true in this revised structure.

8.3 Maximally Consistent Sets of Sentences

Definition 8.1. A set Γ of sentences is *maximally consistent* iff

1. Γ is consistent, and
2. if $\Gamma \subsetneq \Gamma'$, then Γ' is inconsistent.

An alternate definition equivalent to the above is: a set Γ of sentences is *maximally consistent* iff

1. Γ is consistent, and
2. If $\Gamma \cup \{A\}$ is consistent, then $A \in \Gamma$.

In other words, one cannot add sentences not already in Γ to a maximally consistent set Γ without making the resulting larger set inconsistent.

Maximally consistent sets are important in the completeness proof since we can guarantee that every consistent set of sentences Γ is contained in a maximally consistent set Γ^* , and a maximally consistent set contains, for each sentence A , either A or its negation $\neg A$. This is true in particular for atomic sentences, so from a maximally consistent set in a language suitably expanded by constant symbols, we can construct a structure where the interpretation of

8. THE COMPLETENESS THEOREM

predicate symbols is defined according to which atomic sentences are in Γ^* . This structure can then be shown to make all sentences in Γ^* (and hence also in Γ) true. The proof of this latter fact requires that $\neg A \in \Gamma^*$ iff $A \notin \Gamma^*$, $(A \vee B) \in \Gamma^*$ iff $A \in \Gamma^*$ or $B \in \Gamma^*$, etc.

Proposition 8.2. *Suppose Γ is maximally consistent. Then:*

1. *If $\Gamma \vdash A$, then $A \in \Gamma$.*
2. *For any A , either $A \in \Gamma$ or $\neg A \in \Gamma$.*
3. *$(A \wedge B) \in \Gamma$ iff both $A \in \Gamma$ and $B \in \Gamma$.*
4. *$(A \vee B) \in \Gamma$ iff either $A \in \Gamma$ or $B \in \Gamma$.*
5. *$(A \rightarrow B) \in \Gamma$ iff either $A \notin \Gamma$ or $B \in \Gamma$.*

Proof. Let us suppose for all of the following that Γ is maximally consistent.

1. If $\Gamma \vdash A$, then $A \in \Gamma$.

Suppose that $\Gamma \vdash A$. Suppose to the contrary that $A \notin \Gamma$: then since Γ is maximally consistent, $\Gamma \cup \{A\}$ is inconsistent, hence $\Gamma \cup \{A\} \vdash \perp$. By [Proposition 7.15](#), Γ is inconsistent. This contradicts the assumption that Γ is consistent. Hence, it cannot be the case that $A \notin \Gamma$, so $A \in \Gamma$.

2. For any A , either $A \in \Gamma$ or $\neg A \in \Gamma$.

Suppose to the contrary that for some A both $A \notin \Gamma$ and $\neg A \notin \Gamma$. Since Γ is maximally consistent, $\Gamma \cup \{A\}$ and $\Gamma \cup \{\neg A\}$ are both inconsistent, so $\Gamma \cup \{A\} \vdash \perp$ and $\Gamma \cup \{\neg A\} \vdash \perp$. By [Proposition 7.17](#), Γ is inconsistent, a contradiction. Hence there cannot be such a sentence A and, for every A , $A \in \Gamma$ or $\neg A \in \Gamma$.

3. Exercise.

4. $(A \vee B) \in \Gamma$ iff either $A \in \Gamma$ or $B \in \Gamma$.

For the contrapositive of the forward direction, suppose that $A \notin \Gamma$ and $B \notin \Gamma$. We want to show that $(A \vee B) \notin \Gamma$. Since Γ is maximally consistent, $\Gamma \cup \{A\} \vdash \perp$ and $\Gamma \cup \{B\} \vdash \perp$. By [Proposition 7.18](#), $\Gamma \cup \{(A \vee B)\}$ is inconsistent. Hence, $(A \vee B) \notin \Gamma$, as required.

For the reverse direction, suppose that $A \in \Gamma$ or $B \in \Gamma$. Then $\Gamma \vdash A$ or $\Gamma \vdash B$. By [Proposition 7.19](#), $\Gamma \vdash A \vee B$. By (1), $(A \vee B) \in \Gamma$, as required.

5. Exercise.

□

8.4 Henkin Expansion

Part of the challenge in proving the completeness theorem is that the model we construct from a maximally consistent set Γ must make all the quantified formulas in Γ true. In order to guarantee this, we use a trick due to Leon Henkin. In essence, the trick consists in expanding the language by infinitely many constants and adding, for each formula with one free variable $A(x)$ a formula of the form $\exists x A \rightarrow A(c)$, where c is one of the new constant symbols. When we construct the structure satisfying Γ , this will guarantee that each true existential sentence has a witness among the new constants.

Lemma 8.3. *If Γ is consistent in \mathcal{L} and \mathcal{L}' is obtained from \mathcal{L} by adding a countably infinite set of new constant symbols d_1, d_2, \dots , then Γ is consistent in \mathcal{L}' .*

Definition 8.4. A set Γ of formulas of a language \mathcal{L} is *saturated* if and only if for each formula $A \in \text{Frm}(\mathcal{L})$ and variable x there is a constant symbol c such that $\exists x A \rightarrow A(c) \in \Gamma$.

The following definition will be used in the proof of the next theorem.

Definition 8.5. Let \mathcal{L}' be as in Lemma 8.3. Fix an enumeration $\langle A_1, x_1 \rangle, \langle A_2, x_2 \rangle, \dots$ of all formula-variable pairs of \mathcal{L}' . We define the sentences D_n by recursion on n . Assuming that D_1, \dots, D_n have already been defined, let c_{n+1} be the first new constant symbol among the d_i that does not occur in D_1, \dots, D_n , and let D_{n+1} be the formula $\exists x_{n+1} A_{n+1}(x_{n+1}) \rightarrow A_{n+1}(c_{n+1})$. This includes the case where $n = 0$ and the list of previous D_i 's is empty, i.e., D_1 is $\exists x_1 A_1 \rightarrow A_1(c_1)$.

Theorem 8.6. *Every consistent set Γ can be extended to a saturated consistent set Γ' .*

Proof. Given a consistent set of sentences Γ in a language \mathcal{L} , expand the language by adding a countably infinite set of new constant symbols to form \mathcal{L}' . By the previous Lemma, Γ is still consistent in the richer language. Further, let D_i be as in the previous definition: then $\Gamma \cup \{D_1, D_2, \dots\}$ is saturated by construction. Let

$$\begin{aligned}\Gamma_0 &= \Gamma \\ \Gamma_{n+1} &= \Gamma_n \cup \{D_{n+1}\}\end{aligned}$$

i.e., $\Gamma_n = \Gamma \cup \{D_1, \dots, D_n\}$, and let $\Gamma' = \bigcup_n \Gamma_n$. To show that Γ' is consistent it suffices to show, by induction on n , that each set Γ_n is consistent.

The induction basis is simply the claim that $\Gamma_0 = \Gamma$ is consistent, which is the hypothesis of the theorem. For the induction step, suppose that Γ_{n-1} is consistent but $\Gamma_n = \Gamma_{n-1} \cup \{D_n\}$ is inconsistent. Recall that D_n is $\exists x_n A_n(x_n) \rightarrow A_n(c_n)$, where $A(x)$ is a formula of \mathcal{L}' with only the variable x_n free and not containing any constant symbols c_i where $i \geq n$.

8. THE COMPLETENESS THEOREM

If $\Gamma_{n-1} \cup \{D_n\}$ is inconsistent, then $\Gamma_{n-1} \vdash \neg D_n$, and hence both of the following hold:

$$\Gamma_{n-1} \vdash \exists x_n A_n(x_n) \quad \Gamma_{n-1} \vdash \neg A_n(c_n)$$

Here c_n does not occur in Γ_{n-1} or $A_n(x_n)$ (remember, it was added only with D_n). By [Theorem 7.24](#), from $\Gamma \vdash \neg A_n(c_n)$, we obtain $\Gamma \vdash \forall x_n \neg A_n(x_n)$. Thus we have that both $\Gamma_{n-1} \vdash \exists x_n A_n$ and $\Gamma_{n-1} \vdash \forall x_n \neg A_n(x_n)$, so Γ itself is inconsistent. (Note that $\forall x_n \neg A_n(x_n) \vdash \neg \exists x_n A_n(x_n)$.) Contradiction: Γ_{n-1} was supposed to be consistent. Hence $\Gamma_n \cup \{D_n\}$ is consistent. \square

8.5 Lindenbaum's Lemma

Lemma 8.7 (Lindenbaum's Lemma). *Every consistent set Γ can be extended to a maximally consistent saturated set Γ^* .*

Proof. Let Γ be consistent, and let Γ' be as in the proof of [Theorem 8.6](#): we proved there that $\Gamma \cup \Gamma'$ is a consistent saturated set in the richer language \mathcal{L}' (with the countably infinite set of new constants). Let A_0, A_1, \dots be an enumeration of all the formulas of \mathcal{L}' . Define $\Gamma_0 = \Gamma \cup \Gamma'$, and

$$\Gamma_{n+1} = \begin{cases} \Gamma_n \cup \{A_n\} & \text{if } \Gamma_n \cup \{A_n\} \text{ is consistent;} \\ \Gamma_n \cup \{\neg A_n\} & \text{otherwise.} \end{cases}$$

Let $\Gamma^* = \bigcup_{n \geq 0} \Gamma_n$. Since $\Gamma' \subseteq \Gamma^*$, for each formula A , Γ^* contains a formula of the form $\exists x A \rightarrow A(c)$ and thus is saturated.

Each Γ_n is consistent: Γ_0 is consistent by definition. If $\Gamma_{n+1} = \Gamma_n \cup \{A\}$, this is because the latter is consistent. If it isn't, $\Gamma_{n+1} = \Gamma_n \cup \{\neg A\}$, which must be consistent. If it weren't, i.e., both $\Gamma_n \cup \{A\}$ and $\Gamma_n \cup \{\neg A\}$ are inconsistent, then $\Gamma_n \vdash \neg A$ and $\Gamma_n \vdash A$, so Γ_n would be inconsistent contrary to induction hypothesis.

Every formula of $\text{Frm}(\mathcal{L}')$ appears on the list used to define Γ^* . If $A_n \notin \Gamma^*$, then that is because $\Gamma_n \cup \{A_n\}$ was inconsistent. But that means that Γ^* is maximally consistent. \square

8.6 Construction of a Model

We will begin by showing how to construct a structure which satisfies a maximally consistent, saturated set of sentences in a language \mathcal{L} without $=$.

Definition 8.8. Let Γ^* be a maximally consistent, saturated set of sentences in a language \mathcal{L} . The *term model* $M(\Gamma^*)$ of Γ^* is the structure defined as follows:

1. The domain $|M(\Gamma^*)|$ is the set of all closed terms of \mathcal{L} .
2. The interpretation of a constant symbol c is c itself: $c^{M(\Gamma^*)} = c$.

3. The function symbol f is assigned the function

$$f^{M(\Gamma^*)}(t_1, \dots, t_n) = f(\text{Val}^{M(\Gamma^*)}(t_1), \dots, \text{Val}^{M(\Gamma^*)}(t_n))$$

4. If R is an n -place predicate symbol, then $\langle t_1, \dots, t_n \rangle \in R^{M(\Gamma^*)}$ iff $R(t_1, \dots, t_n) \in \Gamma^*$.

Lemma 8.9 (Truth Lemma). *Suppose A does not contain $=$. Then $M(\Gamma^*) \models A$ iff $A \in \Gamma^*$.*

Proof. We prove both directions simultaneously, and by induction on A .

1. $A \equiv R(t_1, \dots, t_n)$: $M(\Gamma^*) \models R(t_1, \dots, t_n)$ iff $\langle t_1, \dots, t_n \rangle \in R^{M(\Gamma^*)}$ (by the definition of satisfaction) iff $R(t_1, \dots, t_n) \in \Gamma^*$ (the construction of $M(\Gamma^*)$).
2. $A \equiv \neg B$: $M(\Gamma^*) \models A$ iff $M(\Gamma^*) \not\models B$ (by definition of satisfaction). By induction hypothesis, $M(\Gamma^*) \not\models B$ iff $B \notin \Gamma^*$. By [Proposition 8.2\(2\)](#), $\neg B \in \Gamma^*$ if $B \notin \Gamma^*$; and $\neg B \notin \Gamma^*$ if $B \in \Gamma^*$ since Γ^* is consistent.
3. $A \equiv B \wedge C$: exercise.
4. $A \equiv B \vee C$: $M(\Gamma^*) \models A$ iff at $M(\Gamma^*) \models B$ or $M(\Gamma^*) \models C$ (by definition of satisfaction) iff $B \in \Gamma^*$ or $C \in \Gamma^*$ (by induction hypothesis). This is the case iff $(B \vee C) \in \Gamma^*$ (by [Proposition 8.2\(4\)](#)).
5. $A \equiv B \rightarrow C$: exercise.
6. $A \equiv \forall x B(x)$: exercise.
7. $A \equiv \exists x B(x)$: First suppose that $M(\Gamma^*) \models A$. By the definition of satisfaction, for some variable assignment s , $M(\Gamma^*), s \models B(x)$. The value $s(x)$ is some term $t \in |M(\Gamma^*)|$. Thus, $M(\Gamma^*) \models B(t)$, and by our induction hypothesis, $B(t) \in \Gamma^*$. By [Theorem 7.25](#) we have $\Gamma^* \vdash \exists x B(x)$. Then, by [Proposition 8.2\(1\)](#), we can conclude that $A \in \Gamma^*$.

Conversely, suppose that $\exists x B(x) \in \Gamma^*$. Because Γ^* is saturated, $(\exists x B(x) \rightarrow B(c)) \in \Gamma^*$. By [Proposition 7.22](#) together with [Proposition 8.2\(1\)](#), $B(c) \in \Gamma^*$. By inductive hypothesis, $M(\Gamma^*) \models B(c)$. Now consider the variable assignment with $s(x) = c^{M(\Gamma^*)}$. Then $M(\Gamma^*), s \models B(x)$. By definition of satisfaction, $M(\Gamma^*) \models \exists x B(x)$.

□

8.7 Identity

The construction of the term model given in the preceding section is enough to establish completeness for first-order logic for sets Γ that do not contain $=$. The term model satisfies every $A \in \Gamma^*$ which does not contain $=$ (and hence all $A \in \Gamma$). It does not work, however, if $=$ is present. The reason is that Γ^* then may contain a sentence $t = t'$, but in the term model the value of any term is that term itself. Hence, if t and t' are different terms, their values in the term model—i.e., t and t' , respectively—are different, and so $t = t'$ is false. We can fix this, however, using a construction known as “factoring.”

Definition 8.10. Let Γ^* be a maximally consistent set of sentences in \mathcal{L} . We define the relation \approx on the set of closed terms of \mathcal{L} by

$$t \approx t' \quad \text{iff} \quad t = t' \in \Gamma^*$$

Proposition 8.11. *The relation \approx has the following properties:*

1. \approx is reflexive.
2. \approx is symmetric.
3. \approx is transitive.
4. If $t \approx t'$, f is a function symbol, and $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n$ are terms, then

$$f(t_1, \dots, t_{i-1}, t, t_{i+1}, \dots, t_n) \approx f(t_1, \dots, t_{i-1}, t', t_{i+1}, \dots, t_n).$$

5. If $t \approx t'$, R is a function symbol, and $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n$ are terms, then

$$R(t_1, \dots, t_{i-1}, t, t_{i+1}, \dots, t_n) \in \Gamma^* \quad \text{iff} \quad R(t_1, \dots, t_{i-1}, t', t_{i+1}, \dots, t_n) \in \Gamma^*.$$

Proof. Since Γ^* is maximally consistent, $t = t' \in \Gamma^*$ iff $\Gamma^* \vdash t = t'$. Thus it is enough to show the following:

1. $\Gamma^* \vdash t = t$ for all terms t .
2. If $\Gamma^* \vdash t = t'$ then $\Gamma^* \vdash t' = t$.
3. If $\Gamma^* \vdash t = t'$ and $\Gamma^* \vdash t' = t''$, then $\Gamma^* \vdash t = t''$.
4. If $\Gamma^* \vdash t = t'$, then

$$\Gamma^* \vdash f(t_1, \dots, t_{i-1}, t, t_{i+1}, \dots, t_n) = f(t_1, \dots, t_{i-1}, t', t_{i+1}, \dots, t_n)$$

for every n -place function symbol f and terms $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n$.

5. If $\Gamma^* \vdash t = t'$ and $\Gamma^* \vdash R(t_1, \dots, t_{i-1}, t, t_{i+1}, \dots, t_n)$, then $\Gamma^* \vdash R(t_1, \dots, t_{i-1}, t', t_{i+1}, \dots, t_n)$ for every n -place predicate symbol R and terms $t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n$.

□

Definition 8.12. Suppose Γ^* is a maximally consistent set in a language \mathcal{L} , t is a term, and \approx as in the previous definition. Then:

$$[t]_{\approx} = \{t' : t' \in \text{Trm}(\mathcal{L}), t \approx t'\}$$

and $\text{Trm}(\mathcal{L})/\approx = \{[t]_{\approx} : t \in \text{Trm}(\mathcal{L})\}$.

Definition 8.13. Let $M = M(\Gamma^*)$ be the term model for Γ^* . Then M/\approx is the following structure:

1. $|M/\approx| = \text{Trm}(\mathcal{L})/\approx$.
2. $c^{M/\approx} = [c]_{\approx}$
3. $f^{M/\approx}([t_1]_{\approx}, \dots, [t_n]_{\approx}) = [f(t_1, \dots, t_n)]_{\approx}$
4. $\langle [t_1]_{\approx}, \dots, [t_n]_{\approx} \rangle \in R^{M/\approx}$ iff $M \models R(t_1, \dots, t_n)$.

Note that we have defined $f^{M/\approx}$ and $R^{M/\approx}$ for elements of $\text{Trm}(\mathcal{L})/\approx$ by referring to them as $[t]_{\approx}$, i.e., via *representatives* $t \in [t]_{\approx}$. We have to make sure that these definitions do not depend on the choice of these representatives, i.e., that for some other choices t' which determine the same equivalence classes ($[t]_{\approx} = [t']_{\approx}$), the definitions yield the same result. For instance, if R is a one-place predicate symbol, the last clause of the definition says that $[t]_{\approx} \in R^{M/\approx}$ iff $M \models R(t)$. If for some other term t' with $t \approx t'$, $M \not\models R(t)$, then the definition would require $[t']_{\approx} \notin R^{M/\approx}$. If $t \approx t'$, then $[t]_{\approx} = [t']_{\approx}$, but we can't have both $[t]_{\approx} \in R^{M/\approx}$ and $[t]_{\approx} \notin R^{M/\approx}$. However, [Proposition 8.11](#) guarantees that this cannot happen.

Proposition 8.14. M/\approx is well defined, i.e., if $t_1, \dots, t_n, t'_1, \dots, t'_n$ are terms, and $t_i \approx t'_i$ then

1. $[f(t_1, \dots, t_n)]_{\approx} = [f(t'_1, \dots, t'_n)]_{\approx}$, i.e., $f(t_1, \dots, t_n) \approx f(t'_1, \dots, t'_n)$ and
2. $M \models R(t_1, \dots, t_n)$ iff $M \models R(t'_1, \dots, t'_n)$, i.e., $R(t_1, \dots, t_n) \in \Gamma^*$ iff $R(t'_1, \dots, t'_n) \in \Gamma^*$.

Proof. Follows from [Proposition 8.11](#). □

Lemma 8.15. $M/\approx \models A$ iff $A \in \Gamma^*$ for all sentences A .

Proof. By induction on A , just as in the proof of [Lemma 8.9](#). The only case that needs additional attention is when $A \equiv t = t'$.

$$\begin{aligned} M/\approx \models t = t' &\text{ iff } [t]_{\approx} = [t']_{\approx} \text{ (by definition of } M/\approx) \\ &\text{ iff } t \approx t' \text{ (by definition of } [t]_{\approx}) \\ &\text{ iff } t = t' \in \Gamma^* \text{ (by definition of } \approx). \end{aligned}$$

□

Note that while $M(\Gamma^*)$ is always countable and infinite, M/\approx may be finite, since it may turn out that there are only finitely many classes $[t]_{\approx}$. This is to be expected, since Γ may contain sentences which require any structure in which they are true to be finite. For instance, $\forall x \forall y x = y$ is a consistent sentence, but is satisfied only in structures with a domain that contains exactly one element.

8.8 The Completeness Theorem

Theorem 8.16 (Completeness Theorem). *Let Γ be a set of sentences. If Γ is consistent, it is satisfiable.*

Proof. Suppose Γ is consistent. By [Lemma 8.7](#), there is a $\Gamma^* \supseteq \Gamma$ which is maximally consistent and saturated. If Γ does not contain $=$, then by [Lemma 8.9](#), $M(\Gamma^*) \models A$ iff $A \in \Gamma^*$. From this it follows in particular that for all $A \in \Gamma$, $M(\Gamma^*) \models A$, so Γ is satisfiable. If Γ does contain $=$, then by [Lemma 8.15](#), $M/\approx \models A$ iff $A \in \Gamma^*$ for all sentences A . In particular, $M/\approx \models A$ for all $A \in \Gamma$, so Γ is satisfiable. \square

Corollary 8.17 (Completeness Theorem, Second Version). *For all Γ and A sentences: if $\Gamma \models A$ then $\Gamma \vdash A$.*

Proof. Note that the Γ 's in [Corollary 8.17](#) and [Theorem 8.16](#) are universally quantified. To make sure we do not confuse ourselves, let us restate [Theorem 8.16](#) using a different variable: for any set of sentences Δ , if Δ is consistent, it is satisfiable. By contraposition, if Δ is not satisfiable, then Δ is inconsistent. We will use this to prove the corollary.

Suppose that $\Gamma \models A$. Then $\Gamma \cup \{\neg A\}$ is unsatisfiable by [Proposition 5.45](#). Taking $\Gamma \cup \{\neg A\}$ as our Δ , the previous version of [Theorem 8.16](#) gives us that $\Gamma \cup \{\neg A\}$ is inconsistent. By [Proposition 7.11](#), $\Gamma \vdash A$. \square

8.9 The Compactness Theorem

Definition 8.18. A set Γ of formulas is *finitely satisfiable* if and only if every finite $\Gamma_0 \subseteq \Gamma$ is satisfiable.

Theorem 8.19 (Compactness Theorem). *The following hold for any sentences Γ and A :*

1. $\Gamma \models A$ iff there is a finite $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \models A$.
2. Γ is satisfiable if and only if it is finitely satisfiable.

Proof. We prove (2). If Γ is satisfiable, then there is a structure M such that $M \models A$ for all $A \in \Gamma$. Of course, this M also satisfies every finite subset of Γ , so Γ is finitely satisfiable.

Now suppose that Γ is finitely satisfiable. Then every finite subset $\Gamma_0 \subseteq \Gamma$ is satisfiable. By soundness, every finite subset is consistent. Then Γ itself must be consistent. For assume it is not, i.e., $\Gamma \vdash \perp$. But derivations are finite, and so already some finite subset $\Gamma_0 \subseteq \Gamma$ must be inconsistent (cf. [Proposition 7.13](#)). But we just showed they are all consistent, a contradiction. Now by completeness, since Γ is consistent, it is satisfiable. \square

8.10 The Löwenheim-Skolem Theorem

Theorem 8.20. *If Γ is consistent then it has a countably infinite model, i.e., it is satisfiable in a structure whose domain is either finite or infinite but countable.*

Proof. If Γ is consistent, the structure \mathbf{M} delivered by the proof of the completeness theorem has a domain $|\mathbf{M}|$ whose cardinality is bounded by that of the set of the terms of the language \mathcal{L} . So \mathbf{M} is at most countably infinite. \square

Theorem 8.21. *If Γ is consistent set of sentences in the language of first-order logic without identity, then it has a countably infinite model, i.e., it is satisfiable in a structure whose domain is infinite and countable.*

Proof. If Γ is consistent and contains no sentences in which identity appears, then the structure \mathbf{M} delivered by the proof of the completeness theorem has a domain $|\mathbf{M}|$ whose cardinality is identical to that of the set of the terms of the language \mathcal{L} . So \mathbf{M} is denumerably infinite. \square

Chapter 9

Beyond First-order Logic

9.1 Overview

First-order logic is not the only system of logic of interest: there are many extensions and variations of first-order logic. A logic typically consists of the formal specification of a language, usually, but not always, a deductive system, and usually, but not always, an intended semantics. But the technical use of the term raises an obvious question: what do logics that are not first-order logic have to do with the word “logic,” used in the intuitive or philosophical sense? All of the systems described below are designed to model reasoning of some form or another; can we say what makes them logical?

No easy answers are forthcoming. The word “logic” is used in different ways and in different contexts, and the notion, like that of “truth,” has been analyzed from numerous philosophical stances. For example, one might take the goal of logical reasoning to be the determination of which statements are necessarily true, true a priori, true independent of the interpretation of the nonlogical terms, true by virtue of their form, or true by linguistic convention; and each of these conceptions requires a good deal of clarification. Even if one restricts one’s attention to the kind of logic used in mathematical, there is little agreement as to its scope. For example, in the *Principia Mathematica*, Russell and Whitehead tried to develop mathematics on the basis of logic, in the *logician* tradition begun by Frege. Their system of logic was a form of higher-type logic similar to the one described below. In the end they were forced to introduce axioms which, by most standards, do not seem purely logical (notably, the axiom of infinity, and the axiom of reducibility), but one might nonetheless hold that some forms of higher-order reasoning should be accepted as logical. In contrast, Quine, whose ontology does not admit “propositions” as legitimate objects of discourse, argues that second-order and higher-order logic are really manifestations of set theory in sheep’s clothing; in other words, systems involving quantification over predicates are not purely logical.

For now, it is best to leave such philosophical issues for a rainy day, and

simply think of the systems below as formal idealizations of various kinds of reasoning, logical or otherwise.

9.2 Many-Sorted Logic

In first-order logic, variables and quantifiers range over a single domain. But it is often useful to have multiple (disjoint) domains: for example, you might want to have a domain of numbers, a domain of geometric objects, a domain of functions from numbers to numbers, a domain of abelian groups, and so on.

Many-sorted logic provides this kind of framework. One starts with a list of “sorts”—the “sort” of an object indicates the “domain” it is supposed to inhabit. One then has variables and quantifiers for each sort, and (usually) an identity predicate for each sort. Functions and relations are also “typed” by the sorts of objects they can take as arguments. Otherwise, one keeps the usual rules of first-order logic, with versions of the quantifier-rules repeated for each sort.

For example, to study international relations we might choose a language with two sorts of objects, French citizens and German citizens. We might have a unary relation, “drinks wine,” for objects of the first sort; another unary relation, “eats wurst,” for objects of the second sort; and a binary relation, “forms a multinational married couple,” which takes two arguments, where the first argument is of the first sort and the second argument is of the second sort. If we use variables a, b, c to range over French citizens and x, y, z to range over German citizens, then

$$\forall a \forall x [(MarriedTo(a, x) \rightarrow (DrinksWine(a) \vee \neg EatsWurst(x)))]$$

asserts that if any French person is married to a German, either the French person drinks wine or the German doesn’t eat wurst.

Many-sorted logic can be embedded in first-order logic in a natural way, by lumping all the objects of the many-sorted domains together into one first-order domain, using unary predicate symbols to keep track of the sorts, and relativizing quantifiers. For example, the first-order language corresponding to the example above would have unary predicate symbols “*German*” and “*French*,” in addition to the other relations described, with the sort requirements erased. A sorted quantifier $\forall x A$, where x is a variable of the German sort, translates to

$$\forall x (German(x) \rightarrow A).$$

We need to add axioms that insure that the sorts are separate—e.g., $\forall x \neg(German(x) \wedge French(x))$ —as well as axioms that guarantee that “drinks wine” only holds of objects satisfying the predicate $French(x)$, etc. With these conventions and axioms, it is not difficult to show that many-sorted sentences translate to first-order sentences, and many-sorted derivations translate to first-order deriva-

tions. Also, many-sorted structures “translate” to corresponding first-order structures and vice-versa, so we also have a completeness theorem for many-sorted logic.

9.3 Second-Order logic

The language of second-order logic allows one to quantify not just over a domain of individuals, but over relations on that domain as well. Given a first-order language \mathcal{L} , for each k one adds variables R which range over k -ary relations, and allows quantification over those variables. If R is a variable for a k -ary relation, and t_1, \dots, t_k are ordinary (first-order) terms, $R(t_1, \dots, t_k)$ is an atomic formula. Otherwise, the set of formulas is defined just as in the case of first-order logic, with additional clauses for second-order quantification. Note that we only have the identity predicate for first-order terms: if R and S are relation variables of the same arity k , we can define $R = S$ to be an abbreviation for

$$\forall x_1 \dots \forall x_k (R(x_1, \dots, x_k) \leftrightarrow S(x_1, \dots, x_k)).$$

The rules for second-order logic simply extend the quantifier rules to the new second order variables. Here, however, one has to be a little bit careful to explain how these variables interact with the predicate symbols of \mathcal{L} , and with formulas of \mathcal{L} more generally. At the bare minimum, relation variables count as terms, so one has inferences of the form

$$A(R) \vdash \exists R A(R)$$

But if \mathcal{L} is the language of arithmetic with a constant relation symbol $<$, one would also expect the following inference to be valid:

$$x < y \vdash \exists R R(x, y)$$

or for a given formula A ,

$$A(x_1, \dots, x_k) \vdash \exists R R(x_1, \dots, x_k)$$

More generally, we might want to allow inferences of the form

$$A[\lambda \vec{x}. B(\vec{x})/R] \exists R A$$

where $A[\lambda \vec{x}. B(\vec{x})/R]$ denotes the result of replacing every atomic formula of the form Rt_1, \dots, t_k in A by $B(t_1, \dots, t_k)$. This last rule is equivalent to having a *comprehension schema*, i.e., an axiom of the form

$$\exists R \forall x_1, \dots, x_k (A(x_1, \dots, x_k) \leftrightarrow R(x_1, \dots, x_k)),$$

one for each formula A in the second-order language, in which R is not a free variable. (Exercise: show that if R is allowed to occur in A , this schema is inconsistent!)

When logicians refer to the “axioms of second-order logic” they usually mean the minimal extension of first-order logic by second-order quantifier rules together with the comprehension schema. But it is often interesting to study weaker subsystems of these axioms and rules. For example, note that in its full generality the axiom schema of comprehension is *impredicative*: it allows one to assert the existence of a relation $R(x_1, \dots, x_k)$ that is “defined” by a formula with second-order quantifiers; and these quantifiers range over the set of all such relations—a set which includes R itself! Around the turn of the twentieth century, a common reaction to Russell’s paradox was to lay the blame on such definitions, and to avoid them in developing the foundations of mathematics. If one prohibits the use of second-order quantifiers in the formula A , one has a *predicative* form of comprehension, which is somewhat weaker.

From the semantic point of view, one can think of a second-order structure as consisting of a first-order structure for the language, coupled with a set of relations on the domain over which the second-order quantifiers range (more precisely, for each k there is a set of relations of arity k). Of course, if comprehension is included in the proof system, then we have the added requirement that there are enough relations in the “second-order part” to satisfy the comprehension axioms—otherwise the proof system is not sound! One easy way to insure that there are enough relations around is to take the second-order part to consist of *all* the relations on the first-order part. Such a structure is called *full*, and, in a sense, is really the “intended structure” for the language. If we restrict our attention to full structures we have what is known as the *full* second-order semantics. In that case, specifying a structure boils down to specifying the first-order part, since the contents of the second-order part follow from that implicitly.

To summarize, there is some ambiguity when talking about second-order logic. In terms of the proof system, one might have in mind either

1. A “minimal” second-order proof system, together with some comprehension axioms.
2. The “standard” second-order proof system, with full comprehension.

In terms of the semantics, one might be interested in either

1. The “weak” semantics, where a structure consists of a first-order part, together with a second-order part big enough to satisfy the comprehension axioms.
2. The “standard” second-order semantics, in which one considers full structures only.

When logicians do not specify the proof system or the semantics they have in mind, they are usually referring to the second item on each list. The advantage to using this semantics is that, as we will see, it gives us categorical descriptions of many natural mathematical structures; at the same time, the proof system is quite strong, and sound for this semantics. The drawback is that the proof system is *not* complete for the semantics; in fact, *no* effectively given proof system is complete for the full second-order semantics. On the other hand, we will see that the proof system *is* complete for the weakened semantics; this implies that if a sentence is not provable, then there is *some* structure, not necessarily the full one, in which it is false.

The language of second-order logic is quite rich. One can identify unary relations with subsets of the domain, and so in particular you can quantify over these sets; for example, one can express induction for the natural numbers with a single axiom

$$\forall R ((R(0) \wedge \forall x (R(x) \rightarrow R(x')))) \rightarrow \forall x R(x).$$

If one takes the language of arithmetic to have symbols $0, ', +, \times$ and $<$, one can add the following axioms to describe their behavior:

1. $\forall x \neg x' = 0$
2. $\forall x \forall y (s(x) = s(y) \rightarrow x = y)$
3. $\forall x (x + 0) = x$
4. $\forall x \forall y (x + y') = (x + y)'$
5. $\forall x (x \times 0) = 0$
6. $\forall x \forall y (x \times y') = ((x \times y) + x)$
7. $\forall x \forall y (x < y \leftrightarrow \exists z y = (x + z'))$

It is not difficult to show that these axioms, together with the axiom of induction above, provide a categorical description of the structure \mathbb{N} , the standard model of arithmetic, provided we are using the full second-order semantics. Given any structure A in which these axioms are true, define a function f from \mathbb{N} to the domain of A using ordinary recursion on \mathbb{N} , so that $f(0) = 0^A$ and $f(x + 1) = \iota^A(f(x))$. Using ordinary induction on \mathbb{N} and the fact that axioms (1) and (2) hold in A , we see that f is injective. To see that f is surjective, let P be the set of elements of $|A|$ that are in the range of f . Since A is full, P is in the second-order domain. By the construction of f , we know that 0^A is in P , and that P is closed under ι^A . The fact that the induction axiom holds in A (in particular, for P) guarantees that P is equal to the entire first-order domain of A . This shows that f is a bijection. Showing that f is a homomorphism is no more difficult, using ordinary induction on \mathbb{N} repeatedly.

In set-theoretic terms, a function is just a special kind of relation; for example, a unary function f can be identified with a binary relation R satisfying $\forall x \exists y R(x, y)$. As a result, one can quantify over functions too. Using the full semantics, one can then define the class of infinite structures to be the class of structures A for which there is an injective function from the domain of A to a proper subset of itself:

$$\exists f (\forall x \forall y (f(x) = f(y) \rightarrow x = y) \wedge \exists y \forall x f(x) \neq y).$$

The negation of this sentence then defines the class of finite structures.

In addition, one can define the class of well-orderings, by adding the following to the definition of a linear ordering:

$$\forall P (\exists x P(x) \rightarrow \exists x (P(x) \wedge \forall y (y < x \rightarrow \neg P(y)))).$$

This asserts that every nonempty set has a least element, modulo the identification of “set” with “one-place relation”. For another example, one can express the notion of connectedness for graphs, by saying that there is no non-trivial separation of the vertices into disconnected parts:

$$\neg \exists A (\exists x A(x) \wedge \exists y \neg A(y) \wedge \forall [w][\forall z ((A(w) \wedge \neg A(z)) \rightarrow \neg R(w, z))]).$$

For yet another example, you might try as an exercise to define the class of finite structures whose domain has even size. More strikingly, one can provide a categorical description of the real numbers as a complete ordered field containing the rationals.

In short, second-order logic is much more expressive than first-order logic. That’s the good news; now for the bad. We have already mentioned that there is no effective proof system that is complete for the full second-order semantics. For better or for worse, many of the properties of first-order logic are absent, including compactness and the Löwenheim-Skolem theorems.

On the other hand, if one is willing to give up the full second-order semantics in terms of the weaker one, then the minimal second-order proof system is complete for this semantics. In other words, if we read \vdash as “proves in the minimal system” and \models as “logically implies in the weaker semantics”, we can show that whenever $\Gamma \models A$ then $\Gamma \vdash A$. If one wants to include specific comprehension axioms in the proof system, one has to restrict the semantics to second-order structures that satisfy these axioms: for example, if Δ consists of a set of comprehension axioms (possibly all of them), we have that if $\Gamma \cup \Delta \models A$, then $\Gamma \cup \Delta \vdash A$. In particular, if A is not provable using the comprehension axioms we are considering, then there is a model of $\neg A$ in which these comprehension axioms nonetheless hold.

The easiest way to see that the completeness theorem holds for the weaker semantics is to think of second-order logic as a many-sorted logic, as follows. One sort is interpreted as the ordinary “first-order” domain, and then for each

k we have a domain of “relations of arity k .” We take the language to have built-in relation symbols “ $true_k(R, x_1, \dots, x_k)$ ” which is meant to assert that R holds of x_1, \dots, x_k , where R is a variable of the sort “ k -ary relation” and x_1, \dots, x_k are objects of the first-order sort.

With this identification, the weak second-order semantics is essentially the usual semantics for many-sorted logic; and we have already observed that many-sorted logic can be embedded in first-order logic. Modulo the translations back and forth, then, the weaker conception of second-order logic is really a form of first-order logic in disguise, where the domain contains both “objects” and “relations” governed by the appropriate axioms.

9.4 Higher-Order logic

Passing from first-order logic to second-order logic enabled us to talk about sets of objects in the first-order domain, within the formal language. Why stop there? For example, third-order logic should enable us to deal with sets of sets of objects, or perhaps even sets which contain both objects and sets of objects. And fourth-order logic will let us talk about sets of objects of that kind. As you may have guessed, one can iterate this idea arbitrarily.

In practice, higher-order logic is often formulated in terms of functions instead of relations. (Modulo the natural identifications, this difference is inessential.) Given some basic “sorts” A, B, C, \dots (which we will now call “types”), we can create new ones by stipulating

If σ and τ are finite types then so is $\sigma \rightarrow \tau$.

Think of types as syntactic “labels,” which classify the objects we want in our domain; $\sigma \rightarrow \tau$ describes those objects that are functions which take objects of type σ to objects of type τ . For example, we might want to have a type Ω of truth values, “true” and “false,” and a type \mathbb{N} of natural numbers. In that case, you can think of objects of type $\mathbb{N} \rightarrow \Omega$ as unary relations, or subsets of \mathbb{N} ; objects of type $\mathbb{N} \rightarrow \mathbb{N}$ are functions from natural numbers to natural numbers; and objects of type $(\mathbb{N} \rightarrow \mathbb{N}) \rightarrow \mathbb{N}$ are “functionals,” that is, higher-type functions that take functions to numbers.

As in the case of second-order logic, one can think of higher-order logic as a kind of many-sorted logic, where there is a sort for each type of object we want to consider. But it is usually clearer just to define the syntax of higher-type logic from the ground up. For example, we can define a set of finite types inductively, as follows:

1. \mathbb{N} is a finite type.
2. If σ and τ are finite types, then so is $\sigma \rightarrow \tau$.
3. If σ and τ are finite types, so is $\sigma \times \tau$.

Intuitively, \mathbb{N} denotes the type of the natural numbers, $\sigma \rightarrow \tau$ denotes the type of functions from σ to τ , and $\sigma \times \tau$ denotes the type of pairs of objects, one from σ and one from τ . We can then define a set of terms inductively, as follows:

1. For each type σ , there is a stock of variables x, y, z, \dots of type σ
2. 0 is a term of type \mathbb{N}
3. S (successor) is a term of type $\mathbb{N} \rightarrow \mathbb{N}$
4. If s is a term of type σ , and t is a term of type $\mathbb{N} \rightarrow (\sigma \rightarrow \sigma)$, then R_{st} is a term of type $\mathbb{N} \rightarrow \sigma$
5. If s is a term of type $\tau \rightarrow \sigma$ and t is a term of type τ , then $s(t)$ is a term of type σ
6. If s is a term of type σ and x is a variable of type τ , then $\lambda x. s$ is a term of type $\tau \rightarrow \sigma$.
7. If s is a term of type σ and t is a term of type τ , then $\langle s, t \rangle$ is a term of type $\sigma \times \tau$.
8. If s is a term of type $\sigma \times \tau$ then $p_1(s)$ is a term of type σ and $p_2(s)$ is a term of type τ .

Intuitively, R_{st} denotes the function defined recursively by

$$\begin{aligned} R_{st}(0) &= s \\ R_{st}(x + 1) &= t(x, R_{st}(x)), \end{aligned}$$

$\langle s, t \rangle$ denotes the pair whose first component is s and whose second component is t , and $p_1(s)$ and $p_2(s)$ denote the first and second elements (“projections”) of s . Finally, $\lambda x. s$ denotes the function f defined by

$$f(x) = s$$

for any x of type σ ; so item (6) gives us a form of comprehension, enabling us to define functions using terms. Formulas are built up from identity predicate statements $s = t$ between terms of the same type, the usual propositional connectives, and higher-type quantification. One can then take the axioms of the system to be the basic equations governing the terms defined above, together with the usual rules of logic with quantifiers and identity predicate.

If one augments the finite type system with a type Ω of truth values, one has to include axioms which govern its use as well. In fact, if one is clever, one can get rid of complex formulas entirely, replacing them with terms of type Ω ! The proof system can then be modified accordingly. The result is essentially the *simple theory of types* set forth by Alonzo Church in the 1930s.

As in the case of second-order logic, there are different versions of higher-type semantics that one might want to use. In the full version, variables of type $\sigma \rightarrow \tau$ range over the set of *all* functions from the objects of type σ to objects of type τ . As you might expect, this semantics is too strong to admit a complete, effective proof system. But one can consider a weaker semantics, in which a structure consists of sets of elements T_τ for each type τ , together with appropriate operations for application, projection, etc. If the details are carried out correctly, one can obtain completeness theorems for the kinds of proof systems described above.

Higher-type logic is attractive because it provides a framework in which we can embed a good deal of mathematics in a natural way: starting with \mathbb{N} , one can define real numbers, continuous functions, and so on. It is also particularly attractive in the context of intuitionistic logic, since the types have clear “constructive” interpretations. In fact, one can develop constructive versions of higher-type semantics (based on intuitionistic, rather than classical logic) that clarify these constructive interpretations quite nicely, and are, in many ways, more interesting than the classical counterparts.

9.5 Intuitionistic logic

In contrast to second-order and higher-order logic, intuitionistic first-order logic represents a restriction of the classical version, intended to model a more “constructive” kind of reasoning. The following examples may serve to illustrate some of the underlying motivations.

Suppose someone came up to you one day and announced that they had determined a natural number x , with the property that if x is prime, the Riemann hypothesis is true, and if x is composite, the Riemann hypothesis is false. Great news! Whether the Riemann hypothesis is true or not is one of the big open questions of mathematics, and here they seem to have reduced the problem to one of calculation, that is, to the determination of whether a specific number is prime or not.

What is the magic value of x ? They describe it as follows: x is the natural number that is equal to 7 if the Riemann hypothesis is true, and 9 otherwise.

Angrily, you demand your money back. From a classical point of view, the description above does in fact determine a unique value of x ; but what you really want is a value of x that is given *explicitly*.

To take another, perhaps less contrived example, consider the following question. We know that it is possible to raise an irrational number to a rational power, and get a rational result. For example, $\sqrt{2}^2 = 2$. What is less clear is whether or not it is possible to raise an irrational number to an *irrational* power, and get a rational result. The following theorem answers this in the affirmative:

Theorem 9.1. *There are irrational numbers a and b such that a^b is rational.*

Proof. Consider $\sqrt{2}^{\sqrt{2}}$. If this is rational, we are done: we can let $a = b = \sqrt{2}$. Otherwise, it is irrational. Then we have

$$(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2} \cdot \sqrt{2}} = \sqrt{2}^2 = 2,$$

which is certainly rational. So, in this case, let a be $\sqrt{2}^{\sqrt{2}}$, and let b be $\sqrt{2}$. \square

Does this constitute a valid proof? Most mathematicians feel that it does. But again, there is something a little bit unsatisfying here: we have proved the existence of a pair of real numbers with a certain property, without being able to say *which* pair of numbers it is. It is possible to prove the same result, but in such a way that the pair a, b is given in the proof: take $a = \sqrt{3}$ and $b = \log_3 4$. Then

$$a^b = \sqrt{3}^{\log_3 4} = 3^{1/2 \cdot \log_3 4} = (3^{\log_3 4})^{1/2} = 4^{1/2} = 2,$$

since $3^{\log_3 x} = x$.

Intuitionistic logic is designed to model a kind of reasoning where moves like the one in the first proof are disallowed. Proving the existence of an x satisfying $A(x)$ means that you have to give a specific x , and a proof that it satisfies A , like in the second proof. Proving that A or B holds requires that you can prove one or the other.

Formally speaking, intuitionistic first-order logic is what you get if you omit restrict a proof system for first-order logic in a certain way. Similarly, there are intuitionistic versions of second-order or higher-order logic. From the mathematical point of view, these are just formal deductive systems, but, as already noted, they are intended to model a kind of mathematical reasoning. One can take this to be the kind of reasoning that is justified on a certain philosophical view of mathematics (such as Brouwer's intuitionism); one can take it to be a kind of mathematical reasoning which is more "concrete" and satisfying (along the lines of Bishop's constructivism); and one can argue about whether or not the formal description captures the informal motivation. But whatever philosophical positions we may hold, we can study intuitionistic logic as a formally presented logic; and for whatever reasons, many mathematical logicians find it interesting to do so.

There is an informal constructive interpretation of the intuitionist connectives, usually known as the Brouwer-Heyting-Kolmogorov interpretation. It runs as follows: a proof of $A \wedge B$ consists of a proof of A paired with a proof of B ; a proof of $A \vee B$ consists of either a proof of A , or a proof of B , where we have explicit information as to which is the case; a proof of $A \rightarrow B$ consists of a procedure, which transforms a proof of A to a proof of B ; a proof of $\forall x A(x)$ consists of a procedure which returns a proof of $A(x)$ for any value of x ; and a proof of $\exists x A(x)$ consists of a value of x , together with a proof that this value satisfies A . One can describe the interpretation in computational terms known as the "Curry-Howard isomorphism" or the "formulas-as-types

paradigm”: think of a formula as specifying a certain kind of data type, and proofs as computational objects of these data types that enable us to see that the corresponding formula is true.

Intuitionistic logic is often thought of as being classical logic “minus” the law of the excluded middle. This following theorem makes this more precise.

Theorem 9.2. *Intuitionistically, the following axiom schemata are equivalent:*

1. $(A \rightarrow \perp) \rightarrow \neg A$.
2. $A \vee \neg A$
3. $\neg\neg A \rightarrow A$

Obtaining instances of one schema from either of the others is a good exercise in intuitionistic logic.

The first deductive systems for intuitionistic propositional logic, put forth as formalizations of Brouwer’s intuitionism, are due, independently, to Kolmogorov, Glivenko, and Heyting. The first formalization of intuitionistic first-order logic (and parts of intuitionist mathematics) is due to Heyting. Though a number of classically valid schemata are not intuitionistically valid, many are.

The *double-negation translation* describes an important relationship between classical and intuitionist logic. It is defined inductively follows (think of A^N as the “intuitionist” translation of the classical formula A):

$$\begin{aligned} !A^N &\equiv \neg\neg A \quad \text{for atomic formulas } A \\ (A \wedge B)^N &\equiv (A^N \wedge B^N) \\ (A \vee B)^N &\equiv \neg\neg(A^N \vee B^N) \\ (A \rightarrow B)^N &\equiv (A^N \rightarrow B^N) \\ (\forall x A)^N &\equiv \forall x A^N \\ (\exists x A)^N &\equiv \neg\neg\exists x A^N \end{aligned}$$

Kolmogorov and Glivenko had versions of this translation for propositional logic; for predicate logic, it is due to Gödel and Gentzen, independently. We have

Theorem 9.3. 1. $A \leftrightarrow A^N$ is provable classically

2. If A is provable classically, then A^N is provable intuitionistically.

We can now envision the following dialogue. Classical mathematician: “I’ve proved A !” Intuitionist mathematician: “Your proof isn’t valid. What you’ve really proved is A^N .” Classical mathematician: “Fine by me!” As far as

the classical mathematician is concerned, the intuitionist is just splitting hairs, since the two are equivalent. But the intuitionist insists there is a difference.

Note that the above translation concerns pure logic only; it does not address the question as to what the appropriate *nonlogical* axioms are for classical and intuitionistic mathematics, or what the relationship is between them. But the following slight extension of the theorem above provides some useful information:

Theorem 9.4. *If Γ proves A classically, Γ^N proves A^N intuitionistically.*

In other words, if A is provable from some hypotheses classically, then A^N is provable from their double-negation translations.

To show that a sentence or propositional formula is intuitionistically valid, all you have to do is provide a proof. But how can you show that it is not valid? For that purpose, we need a semantics that is sound, and preferably complete. A semantics due to Kripke nicely fits the bill.

We can play the same game we did for classical logic: define the semantics, and prove soundness and completeness. It is worthwhile, however, to note the following distinction. In the case of classical logic, the semantics was the “obvious” one, in a sense implicit in the meaning of the connectives. Though one can provide some intuitive motivation for Kripke semantics, the latter does not offer the same feeling of inevitability. In addition, the notion of a classical structure is a natural mathematical one, so we can either take the notion of a structure to be a tool for studying classical first-order logic, or take classical first-order logic to be a tool for studying mathematical structures. In contrast, Kripke structures can only be viewed as a logical construct; they don’t seem to have independent mathematical interest.

A Kripke structure for a propositional language consists of a partial order $\text{Mod}(P)$ with a least element, and an “monotone” assignment of propositional variables to the elements of $\text{Mod}(P)$. The intuition is that the elements of $\text{Mod}(P)$ represent “worlds,” or “states of knowledge”; an element $p \geq q$ represents a “possible future state” of q ; and the propositional variables assigned to p are the propositions that are known to be true in state p . The forcing relation $\mathbf{P}, p \Vdash A$ then extends this relationship to arbitrary formulas in the language; read $\mathbf{P}, p \Vdash A$ as “ A is true in state p .” The relationship is defined inductively, as follows:

1. $\mathbf{P}, p \Vdash p_i$ iff p_i is one of the propositional variables assigned to p .
2. $\mathbf{P}, p \not\Vdash \perp$.
3. $\mathbf{P}, p \Vdash (A \wedge B)$ iff $\mathbf{P}, p \Vdash A$ and $\mathbf{P}, p \Vdash B$.
4. $\mathbf{P}, p \Vdash (A \vee B)$ iff $\mathbf{P}, p \Vdash A$ or $\mathbf{P}, p \Vdash B$.
5. $\mathbf{P}, p \Vdash (A \rightarrow B)$ iff, whenever $q \geq p$ and $\mathbf{P}, q \Vdash A$, then $\mathbf{P}, q \Vdash B$.

It is a good exercise to try to show that $\neg(p \wedge q) \rightarrow (\neg p \vee \neg q)$ is not intuitionistically valid, by cooking up a Kripke structure that provides a counterexample.

9.6 Modal Logics

Consider the following example of a conditional sentence:

If Jeremy is alone in that room, then he is drunk and naked and dancing on the chairs.

This is an example of a conditional assertion that may be materially true but nonetheless misleading, since it seems to suggest that there is a stronger link between the antecedent and conclusion other than simply that either the antecedent is false or the consequent true. That is, the wording suggests that the claim is not only true in this particular world (where it may be trivially true, because Jeremy is not alone in the room), but that, moreover, the conclusion *would have* been true *had* the antecedent been true. In other words, one can take the assertion to mean that the claim is true not just in this world, but in any “possible” world; or that it is *necessarily* true, as opposed to just true in this particular world.

Modal logic was designed to make sense of this kind of necessity. One obtains modal propositional logic from ordinary propositional logic by adding a box operator; which is to say, if A is a formula, so is $\Box A$. Intuitively, $\Box A$ asserts that A is *necessarily* true, or true in any possible world. $\Diamond A$ is usually taken to be an abbreviation for $\neg\Box\neg A$, and can be read as asserting that A is *possibly* true. Of course, modality can be added to predicate logic as well.

Kripke structures can be used to provide a semantics for modal logic; in fact, Kripke first designed this semantics with modal logic in mind. Rather than restricting to partial orders, more generally one has a set of “possible worlds,” P , and a binary “accessibility” relation $R(x, y)$ between worlds. Intuitively, $R(p, q)$ asserts that the world q is compatible with p ; i.e., if we are “in” world p , we have to entertain the possibility that the world could have been like q .

Modal logic is sometimes called an “intensional” logic, as opposed to an “extensional” one. The intended semantics for an extensional logic, like classical logic, will only refer to a single world, the “actual” one; while the semantics for an “intensional” logic relies on a more elaborate ontology. In addition to structuring necessity, one can use modality to structure other linguistic constructions, reinterpreting \Box and \Diamond according to the application. For example:

1. In provability logic, $\Box A$ is read “ A is provable” and $\Diamond A$ is read “ A is consistent.”

2. In epistemic logic, one might read $\Box A$ as “I know A ” or “I believe A .”
3. In temporal logic, one can read $\Box A$ as “ A is always true” and $\Diamond A$ as “ A is sometimes true.”

One would like to augment logic with rules and axioms dealing with modality. For example, the system **S4** consists of the ordinary axioms and rules of propositional logic, together with the following axioms:

$$\begin{aligned}\Box(A \rightarrow B) &\rightarrow (\Box A \rightarrow \Box B) \\ \Box A &\rightarrow A \\ \Box A &\rightarrow \Box \Box A\end{aligned}$$

as well as a rule, “from A conclude $\Box A$.” **S5** adds the following axiom:

$$\Diamond A \rightarrow \Box \Diamond A$$

Variations of these axioms may be suitable for different applications; for example, **S5** is usually taken to characterize the notion of logical necessity. And the nice thing is that one can usually find a semantics for which the proof system is sound and complete by restricting the accessibility relation in the Kripke structures in natural ways. For example, **S4** corresponds to the class of Kripke structures in which the accessibility relation is reflexive and transitive. **S5** corresponds to the class of Kripke structures in which the accessibility relation is *universal*, which is to say that every world is accessible from every other; so $\Box A$ holds if and only if A holds in every world.

9.7 Other Logics

As you may have gathered by now, it is not hard to design a new logic. You too can create your own a syntax, make up a deductive system, and fashion a semantics to go with it. You might have to be a bit clever if you want the proof system to be complete for the semantics, and it might take some effort to convince the world at large that your logic is truly interesting. But, in return, you can enjoy hours of good, clean fun, exploring your logic’s mathematical and computational properties.

Recent decades have witnessed a veritable explosion of formal logics. Fuzzy logic is designed to model reasoning about vague properties. Probabilistic logic is designed to model reasoning about uncertainty. Default logics and nonmonotonic logics are designed to model defeasible forms of reasoning, which is to say, “reasonable” inferences that can later be overturned in the face of new information. There are epistemic logics, designed to model reasoning about knowledge; causal logics, designed to model reasoning about causal relationships; and even “deontic” logics, which are designed to model reasoning about moral and ethical obligations. Depending on whether the primary

motivation for introducing these systems is philosophical, mathematical, or computational, you may find such creatures studies under the rubric of mathematical logic, philosophical logic, artificial intelligence, cognitive science, or elsewhere.

The list goes on and on, and the possibilities seem endless. We may never attain Leibniz' dream of reducing all of human reason to calculation—but that can't stop us from trying.

Part III

Turing Machines

Chapter 10

Turing Machine Computations

10.1 Introduction

Even though the term “Turing machine” evokes the image of a physical machine with moving parts, strictly speaking a Turing machine is a purely mathematical construct. It is perhaps best to think of a Turing machine as a program for a special kind of imaginary mechanism. This mechanism consists of a *tape* and a *read-write head*. In our version of Turing machines, the tape is infinite in one direction (to the right), and it is divided into *squares*, each of which may contain a symbol from a finite *alphabet*. Such alphabets can contain any number of different symbols, but we will mainly make do with three: \triangleright , \sqcup , and $|$. When the mechanism is started, the tape is empty (i.e., each square contains the symbol \sqcup) except for the leftmost square, which contains \triangleright , and a finite number of squares which contain the *input*. At any time, the mechanism is in one of a finite number of *states*. At the outset, the head scans the leftmost square and is in a specified *initial state*. At each step of the mechanism’s run, the content of the square currently scanned together with the state the mechanism is in and the Turing machine program determine what happens next. The Turing machine program consists of a list of 5-tuples $\langle q_i, \sigma, q_j, \sigma', D \rangle$. Whenever the mechanism is in state q_i and reads symbol σ , it replaces the symbol on the current square with σ' , the head moves left, right, or stays put according to whether D is L , R , or N , and the mechanism goes into state q_j . When the mechanism enters state h we say it *halts*, and the contents of the tape at that point is its *output*.

10.2 Turing Machines

The formal definition of what constitutes a Turing machine looks abstract, but is actually simple: it merely packs into one mathematical structure all the information needed to specify the workings of a Turing machine. This includes (1) which states the machine can be in, (2) which symbols are allowed

to be on the tape, (3) which state the machine should start in, and (4) what the instruction set of the machine is.

Definition 10.1. A Turing machine $T = \langle Q, \Sigma, s, I \rangle$ consists of

1. a finite set of states Q which includes the *halting state* h ,
2. a finite alphabet Σ which includes \triangleright and \sqcup ,
3. an initial state $s \in Q$,
4. a finite instruction set $I \subseteq Q \times \Sigma \times Q \times \Sigma \times \{L, R, N\}$.

We assume that the tape is infinite in one direction only. For this reason it is useful to designate a special symbol \triangleright as a marker for the left end of the tape. This makes it easier for Turing machine programs to tell when they're "in danger" of running off the tape. Other definitions of Turing machines are possible, including one where the tape is infinite in both directions. In that case, marker for the left end of the tape is not necessary.

10.3 Configurations and Computations

The imaginary mechanism consisting of tape, read/write head, and Turing machine program is really just in intuitive way of visualizing what a Turing machine computation is. Formally, we can define the computation of a Turing machine on a given input as a sequence of *configurations*—and a configuration in turn is a sequence of symbols (corresponding to the contents of the tape at a given point in the computation), a number indicating the position of the read/write head, and a state. Using these, we can define what the Turing machine M computes on a given input.

Definition 10.2. A *configuration* of Turing machine $M = \langle Q, \Sigma, s, I \rangle$ is a triple $\langle C, n, q \rangle$ where

1. $C \in \Sigma^*$ is a finite sequence of symbols from Σ ,
2. $n \in \mathbb{N}$ is a number $\leq \text{len}(C)$, and
3. $q \in Q$

The potential input for a Turing machine is a sequence of symbols, usually a sequence that encodes a number in some form. The initial configuration of the Turing machine is that configuration in which we start the Turing machine to work on that input: the tape contains the tape end marker immediately followed by the input written on the squares to the right, the read/write head is scanning the leftmost square of the tape (i.e., the left end marker), and the mechanism is in the designated start state s .

Definition 10.3. The *initial configuration* of M for input $I \in \Sigma^*$ is

$$\langle \triangleright \frown I, 0, s \rangle$$

Definition 10.4. We say that a configuration $\langle C, n, q \rangle$ *yields* $\langle C', n', q' \rangle$ in one step (according to M), iff

1. the n -th symbol of C is σ ,
2. the instruction set of M contains a tuple $\langle q, \sigma, q', \sigma', D \rangle$,
3. the n -th symbol of C' is σ' ,
4.
 - a) $D = L$ and $n' = n - 1$, or
 - b) $D = R$ and $n' = n$, or
 - c) $D = N$ and $n' = n$,
5. for all $i \neq n$, $C'(i) = C(i)$,
6. if $n' > \text{len}(C)$, then $\text{len}(C') = \text{len}(C) + 1$ and the n' -th symbol of C' is \dots .

Definition 10.5. A *run* of M on input I is a sequence C_i of configurations of M , where C_0 is the initial configuration of M for input I , and each C_i yields C_{i+1} in one step.

We say that M *halts on input* I after k steps if $C_k = \langle \triangleright \frown O, n, h \rangle$. In that case the *output* of M for input I is O .

10.4 Unary Representation of Numbers

Turing machines work on sequences of symbols written on their tape. Depending on the alphabet a Turing machine uses, these sequences of symbols can represent various inputs and outputs. Of particular interest, of course, are Turing machines which compute *arithmetical* functions, i.e., functions of natural numbers. A simple way to represent positive integers is by coding them as sequences of a single symbol $|$.

Definition 10.6. A Turing machine M *computes* the function $f: \mathbb{N}^n \rightarrow \mathbb{N}$ iff M halts on input

$$|^{k_1} \sqcup |^{k_2} \sqcup \dots \sqcup |^{k_n}$$

with output $|^{f(k_1, \dots, k_n)}$.

Chapter 11

Undecidability

11.1 Decision Problems

We have a number of important logical notions, such as those of satisfiability, validity, and consequence, which are properties of (sets) formulas or relations between them. The definitions of these logical notions provide *criteria* for when, say, a sentence is valid. (In this case, it is valid iff it is satisfied in every first-order structure.) These criteria, however, do not in general provide a *method for deciding* if a sentence or set of sentences has the property in question. The question of whether there is such a method for a given notion and class of formulas or sentences is called a *decision problem*.

Example 11.1. The *decision problem for validity of sentences in a first-order language* is the question of whether there is a procedure for deciding, given a sentence in a first-order language, if it is valid or not.

Example 11.2. The *decision problem for satisfiability of sentences in a first-order language* is the question of whether there is a procedure for deciding, given a sentence in a first-order language, if it is satisfiable or not.

We say that a decision problem is *solvable* if there is such a procedure, and *unsolvable* otherwise.

To show that a decision problem is solvable, you typically simply write down the procedure that solves it (and prove that it in fact solves it correctly in every case). Proving that a decision problem is unsolvable is a lot harder: you have to show that there can *in principle* be no procedure whatsoever that solves it. This can only be done rigorously if there is a precise definition of what a “procedure” is.

One such precise definition is provided by the notion of a *Turing machine*. A decision problem is solvable by a Turing machine if there is a Turing machine which, when started on the description of an instance of the decision problem as input, eventually halts with simply “1” or “0” on the tape, representing

“yes” or “no”, respectively. To show that a decision problem is not solvable it suffices to show that no such Turing machine exists.

There are decision problems that can relatively easily be shown to be unsolvable by Turing machines, such as the Halting problem. To show logical decision problems undecidable, you show that a known unsolvable problem can be “reduced” to it. A decision problem A can be reduced to a decision problem B iff the answer to an instance of A can be obtained by an answer to an instance of B. For instance, the problem of validity of a sentence can be reduced to that of satisfiability: Given a sentence A , form the sentence $\neg A$: the former is valid iff the latter is not satisfiable. So the decision problem for validity can be solved by taking an instance, transforming that instance into its negation by putting “ \neg ” in front of it, obtaining the answer to “Is $\neg A$ satisfiable?,” and switching the answer. This shows that the problem of validity can be reduced to the problem of satisfiability.

A reduction of a decision problem A to a decision problem B shows two things:

1. If B is solvable, then A is solvable.
2. If A is unsolvable, then B is unsolvable.

A decision problem can therefore be shown to be unsolvable by reducing a known unsolvable problem, such as the Halting Problem, to it.

11.2 Representing Turing Machines

In order to represent Turing machines and their behavior by a sentence of first-order logic, we have to define a suitable language. The language consists of two parts: predicates for describing configurations of the machine, and expressions for counting execution steps (“moments”) and positions on the tape. The latter require an initial moment, o , a “successor” function which is traditionally written as a postfix $'$, and an ordering $x < y$ of “before.”

Definition 11.3. Given a Turing machine $M = \langle Q, \Sigma, \delta, s \rangle$, the language \mathcal{L}_M consists of:

1. A two-place predicate $Q_q(x, y)$ for every state $q \in Q$.
2. A two-place predicate $S_\sigma(x, y)$ for every symbol $\sigma \in \Sigma$
3. A constant o
4. A one-place function $'$
5. A two-place predicate $<$

For each number n there is a canonical term \bar{n} , the *numeral* for n , which represents it in \mathcal{L}_M . $\bar{0}$ is o , $\bar{1}$ is o' , $\bar{2}$ is o'' , and so on. More formally:

$$\begin{aligned}\bar{0} &= o \\ \overline{n+1} &= \bar{n}'\end{aligned}$$

The sentences describing the operation of the Turing machine M on input w are the following:

I Axioms describing numbers:

a) A sentence that says that the successor function is injective:

$$\forall x \forall y (x' = y' \rightarrow x = y)$$

b) A sentence that says that every number is less than its successor:

$$\forall x (x < x')$$

c) A sentence that ensures that $<$ is transitive:

$$\forall x \forall y \forall z ((x < y \wedge y < z) \rightarrow x < z)$$

II. Axioms describing the input configuration:

a) M is in the initial state s at time 0, scanning square 0:

$$Q_s(o, o)$$

b) The first n squares contain the symbols $\triangleright, \sigma_{i_1}, \dots, \sigma_{i_n}$:

$$S_{\triangleright}(o, o) \wedge S_{i_1}(\bar{1}, o) \wedge \dots \wedge S_{i_n}(\bar{n}, o)$$

c) Otherwise, the tape is empty:

$$\forall x (\bar{n} < x \rightarrow S_{\perp}(x, o))$$

III. Axioms describing the transition from one configuration to the next:

For the following, let $A(x, y)$ be the conjunction of all sentences of the form

$$\forall z ((z < x \vee x < z \wedge S_{\sigma}(z, y)) \rightarrow S_{\sigma}(z, y'))$$

where $\sigma \in \Sigma$.

1. For every instruction $\langle q_i, \sigma, q_j, \sigma', L \rangle$, the sentence:

$$\forall x \forall y ((Q_{q_i}(x', y) \wedge S_{\sigma}(x, y)) \rightarrow (Q_{q_j}(x, y) \wedge S_{\sigma'}(x, y) \wedge A(x, y)))$$

2. For every instruction $\langle q_i, \sigma, q_j, \sigma', R \rangle$, the sentence:

$$\forall x \forall y ((Q_{q_i}(x, y) \wedge S_{\sigma}(x, y)) \rightarrow (Q_{q_j}(x', y) \wedge S_{\sigma'}(x, y) \wedge A(x, y)))$$

(Probably also need axioms saying every square as exactly one symbol on it at all times, machine always in exactly one state.)

Let $T(M, w)$ be the conjunction of all the above sentences for Turing machine M and input w

The sentence $H(M, w)$:

$$\exists x \exists y Q_h(x, y)$$

expresses that the Turing machine M halts on input w .

11.3 Verifying the Representation

In order to verify that our representation works, we first have to make sure that if M halts on input w , then $T(m, w) \rightarrow H(M, w)$ is valid. We can do this simply by proving that $T(m, w)$ implies a description of the configuration of M for each step of the execution of M on input w . If M halts on input w , then for some n , M will be in a halting configuration at step n (and be scanning square m , for some m). Hence, $T(M, w)$ implies $Q_h(\bar{m}, \bar{n})$.

Definition 11.4. Let $C(M, w, n)$ be the sentence

$$Q_q(\bar{m}, \bar{n}) \wedge S_{\sigma_0}(\bar{0}, \bar{n}) \wedge \dots \wedge S_{\sigma_k}(\bar{k}, \bar{n})$$

where q is the state of M at time n , M is scanning square m at time n , square i contains symbol σ_i at time n for $0 \leq i \leq k$ and k is the right-most non-blank square of the tape at time m .

Lemma 11.5. For each n , $T(M, w)$ implies $C(M, w, n)$.

Proof. By induction on n .

If $n = 0$, then $C(M, w, n)$ is a conjunct of $T(M, w)$, so implied by it.

Suppose $n > 0$ and at time n , M started on w is in state q scanning square m , and the content of the tape is $\sigma_0, \dots, \sigma_k$.

... to be completed

□

To complete the verification of our claim, we also have to establish the reverse direction: if $T(M, w) \rightarrow H(M, w)$ is valid, then M does in fact halt when started on input m .

Lemma 11.6. If $T(M, w)$ entails $H(M, w)$, then M halts on input w .

Proof. Consider the \mathcal{L}_M -structure \mathbf{M} with domain \mathbb{N} which interprets o as 0 , $'$ as the successor function, and $<$ as the less-than relation, and the predicates Q_q and S_σ as follows:

$$Q_q^{\mathbf{M}} = \{(m, n) : \text{after } n \text{ steps, } M \text{ started on } w \text{ is in state } q \text{ scanning square } m\}$$

$$S_\sigma^{\mathbf{M}} = \{(m, n) : \text{after } n \text{ steps, } M \text{ started on } w \text{ has symbol } \sigma \text{ on square } m\}$$

Clearly, $\mathbf{M} \models T(M, w)$. If $T(M, w)$ implies $H(M, w)$, then $\mathbf{M} \models H(M, w)$, i.e.,

$$\mathbf{M} \models \exists x \exists y Q_h(x, y).$$

As $|\mathbf{M}| = \mathbb{N}$, there must be $m, n \in \mathbb{N}$ so that $\mathbf{M} \models Q_h(\bar{n}, \bar{m})$. By the definition of \mathbf{M} , this means that M started on input w is in state h after m steps, i.e., has halted. \square

Problems

Problems for Chapter 1

Problem 1.1. List all subsets of $\{a, b, c, d\}$.

Problem 1.2. Prove in detail that $X \cup (X \cap Y) = X$. Then compress it into a “textbook proof.” (Hint: for the $X \cup (X \cap Y) \subseteq X$ direction you will need proof by cases, aka \vee Elim.)

Problems for Chapter 2

Problem 2.1. List the elements of the relation \subseteq on the set $\wp(\{a, b, c\})$.

Problem 2.2. Give examples of relations that are (a) reflexive and symmetric but not transitive, (b) reflexive and anti-symmetric, (c) anti-symmetric, transitive, but not reflexive, and (d) reflexive, symmetric, and transitive. Do not use relations on numbers or sets.

Problem 2.3. Show that if R is a weak partial order on X , then $R^- = R \setminus \text{Id}_X$ is a strict partial order and Rxy iff R^-xy for all $x \neq y$.

Problem 2.4. Show that the transitive closure of R is in fact transitive.

Problems for Chapter 3

Problem 3.1. Show that if $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ are both injective, then $g \circ f: X \rightarrow Z$ is injective.

Problem 3.2. Show that if $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ are both surjective, then $g \circ f: X \rightarrow Z$ is surjective.

Problem 3.3. A function $g: Y \rightarrow X$ is the *inverse* of a function $f: X \rightarrow Y$ if $f(g(y)) = y$ and $g(f(x)) = x$ for all $x \in X$ and $y \in Y$.

Show that if f is bijective, such a function g exists, i.e., define such a function and show that it is a function. Then show that it is also bijective.

Problem 3.4. Show that if the functions $g: X \rightarrow Y$ and $f: Y \rightarrow Z$ are both bijections then $(g \circ f): X \rightarrow Z$ is also a bijection.

Problems for Chapter 4

Problem 4.1. Give an enumeration of the set of all ordered pairs of positive rational numbers.

Problem 4.2. Recall from your introductory logic course that each possible in a truth table expresses a truth function. In other words, the truth functions are all functions from $\mathbb{B}^k \rightarrow \mathbb{B}$ for some k . Prove that the set of all truth functions is enumerable.

Problem 4.3. Show that the set of all finite subsets of an arbitrary infinite enumerable set is enumerable.

Problem 4.4. Show that if X and Y are countable, so is $X \cup Y$.

Problem 4.5. A set of positive integers is said to be *cofinite* iff it is the complement of a finite set of positive integers. Let I be the set that contains all the finite and cofinite sets of positive integers. Show that I is enumerable.

Problem 4.6. Show that the countable union of countable sets is countable. That is, whenever X_1, X_2, \dots are sets, and each X_i is countable, then the union $\bigcup_{i=1}^{\infty} X_i$ of all of them is also countable.

Problem 4.7. Show that $\wp(\mathbb{N})$ is uncountable.

Problem 4.8. Show that the set of functions $f: \mathbb{Z}^+ \rightarrow \mathbb{Z}^+$ is uncountable by a direct diagonal argument.

Problem 4.9. Show that the set of all sets of pairs of positive integers is uncountable.

Problem 4.10. Show that \mathbb{N}^{ω} , the set of infinite sequences of natural numbers, is uncountable.

Problem 4.11. Let P be the set of total functions from the set of positive integers to the set $\{0\}$, and let Q be the set of partial functions from the set of positive integers to the set $\{0\}$. Show that P is countable and Q is not.

Problem 4.12. Let S be the set of all total surjective functions from the set of positive integers to the set $\{0,1\}$. Show that S is uncountable.

Problem 4.13. Show that the set \mathbb{R} of all real numbers is uncountable.

Problem 4.14. Show that if X is equinumerous with U and Y is equinumerous with V , and the intersections $X \cap Y$ and $U \cap V$ are empty, then the unions $X \cup Y$ and $U \cup V$ are equinumerous.

Problem 4.15. Given an enumeration of a set X , show that if X is not finite then it is equinumerous with the positive integers \mathbb{Z}^+ .

Problems for Chapter 5

Problem 5.1. Prove [Lemma 5.10](#).

Problem 5.2. Prove [Proposition 5.11](#) (Hint: Formulate and prove a version of [Lemma 5.10](#) for terms.)

Problem 5.3. Give an inductive definition of the bound variable occurrences along the lines of [Definition 5.17](#).

Problem 5.4. Is N , the standard model of arithmetic, covered? Explain.

Problem 5.5. Let $\mathcal{L} = \{c, f, A\}$ with one constant symbol, one one-place function symbol and one two-place predicate symbol, and let the structure M be given by

1. $|M| = \{1, 2, 3\}$
2. $c^M = 3$
3. $f^M(1) = 2, f^M(2) = 3, f^M(3) = 2$
4. $A^M = \{\langle 1, 2 \rangle, \langle 2, 3 \rangle, \langle 3, 3 \rangle\}$

(a) Let $s(v) = 1$ for all variables v . Find out whether

$$M, s \models \exists x (A(f(z), c) \rightarrow \forall y (A(y, x) \vee A(f(y), x)))$$

Explain why or why not.

(b) Give a different structure and variable assignment in which the formula is not satisfied.

Problem 5.6. Complete the proof of [Proposition 5.35](#).

Problem 5.7. Show that if A is a sentence, $M \models A$ iff *there is* a variable assignment s so that $M, s \models A$.

Problem 5.8. Prove [Proposition 5.37](#).

Problem 5.9. Suppose \mathcal{L} is a language without function symbols. Given a structure M and $a \in |M|$, define $M[a/c]$ to be the structure that is just like M , except that $c^{M[a/c]} = a$. Define $M \models A$ for sentences A by:

1. $A \equiv \perp$: $M \models A$.
2. $A \equiv R(d_1, \dots, d_n)$: $M \models A$ iff $\langle d_1^M, \dots, d_n^M \rangle \in R^M$.
3. $A \equiv d_1 = d_2$: $M \models A$ iff $d_1^M = d_2^M$.
4. $A \equiv \neg B$: $M \models A$ iff not $M \models B$.

5. $A \equiv (B \wedge C)$: $M \models A$ iff $M \models B$ and $M \models C$.
6. $A \equiv (B \vee C)$: $M \models A$ iff $M \models B$ or $M \models C$ (or both).
7. $A \equiv (B \rightarrow C)$: $M \models A$ iff not $M \models B$ or $M \models C$ (or both).
8. $A \equiv \forall x B$: $M \models A$ iff for all $a \in |M|$, $M[a/c] \models B[c/x]$, if c does not occur in B .
9. $A \equiv \exists x B$: $M \models A$ iff there is an $a \in |M|$ such that $M[a/c] \models B[c/x]$, if c does not occur in B .

Let x_1, \dots, x_n be all free variables in A , c_1, \dots, c_n constant symbols not in A , $a_1, \dots, a_n \in |M|$, and $s(x_i) = a_i$.

Show that $M, s \models A$ iff $M[a_1/c_1, \dots, a_n/c_n] \models A[c_1/x_1] \dots [c_n/x_n]$.

Problem 5.10. Suppose that f is a function symbol not in $A(x, y)$. Show that there is a M such that $M \models \forall x \exists y A(x, y)$ iff there is a M' such that $M' \models \forall x A(x, f(x))$.

Problem 5.11. Prove [Proposition 5.40](#)

- Problem 5.12.**
1. Show that $\Gamma \models \perp$ iff Γ is unsatisfiable.
 2. Show that $\Gamma, A \models \perp$ iff $\Gamma \models \neg A$.
 3. Suppose c does not occur in A or Γ . Show that $\Gamma \models \forall x A$ iff $\Gamma \models A[c/x]$.

Problems for Chapter 6

Problem 6.1. Find formulas in \mathcal{L}_A which define the following relations:

1. n is between i and j ;
2. n evenly divides m (i.e., m is a multiple of n);
3. n is a prime number (i.e., no number other than 1 and n evenly divides n).

Problem 6.2. Suppose the formula $A(x_1, x_2)$ expresses the relation $R \subseteq |M|^2$ in a structure M . Find formulas that express the following relations:

1. the inverse R^{-1} of R ;
2. the relative product $R \mid R$;

Can you find a way to express R^+ , the transitive closure of R ?

Problem 6.3. Let \mathcal{L} be the language containing a 2-place predicate symbol $<$ only (no other constant symbols, function symbols or predicate symbols—except of course $=$). Let N be the structure such that $|N| = \mathbb{N}$, and $<^N = \{\langle n, m \rangle : n < m\}$. Prove the following:

1. $\{0\}$ is definable in N ;
2. $\{1\}$ is definable in N ;
3. $\{2\}$ is definable in N ;
4. for each $n \in \mathbb{N}$, the set $\{n\}$ is definable in N ;
5. every finite subset of $|N|$ is definable in N ;
6. every co-finite subset of $|N|$ is definable in N (where $X \subseteq \mathbb{N}$ is co-finite iff $\mathbb{N} \setminus X$ is finite).

Problems for Chapter 7

Problem 7.1. Give derivations of the following formulas:

1. $\neg(A \rightarrow B) \rightarrow (A \wedge \neg B)$
2. $\forall x (A(x) \rightarrow B) \rightarrow (\exists y A(y) \rightarrow B)$

Problem 7.2. Prove [Proposition 7.11](#)

Problem 7.3. Prove [Proposition 7.12](#)

Problem 7.4. Prove [Proposition 7.18](#)

Problem 7.5. Prove [Proposition 7.19](#).

Problem 7.6. Prove [Proposition 7.20](#).

Problem 7.7. Prove [Proposition 7.21](#).

Problem 7.8. Prove [Proposition 7.22](#).

Problem 7.9. Prove [Proposition 7.23](#).

Problem 7.10. Complete the proof of [Theorem 7.26](#).

Problem 7.11. Prove that $=$ is both symmetric and transitive, i.e., give derivations of $\forall x \forall y (x = y \rightarrow y = x)$ and $\forall x \forall y \forall z ((x = y \wedge y = z) \rightarrow x = z)$

Problem 7.12. Give derivations of the following formulas:

1. $\forall x \forall y ((x = y \wedge A(x)) \rightarrow A(y))$
2. $\exists x A(x) \wedge \forall y \forall z ((A(y) \wedge A(z)) \rightarrow y = z) \rightarrow \exists x (A(x) \wedge \forall y (A(y) \rightarrow y = x))$

Problems for Chapter 8

Problem 8.1. Complete the proof of [Proposition 8.2](#).

Problem 8.2. Complete the proof of [Lemma 8.9](#).

Problem 8.3. Complete the proof of [Proposition 8.11](#).

Problem 8.4. Use [Corollary 8.17](#) to prove [Theorem 8.16](#), thus showing that the two formulations of the completeness theorem are equivalent.

Problem 8.5. In order for a derivation system to be complete, its rules must be strong enough to prove every unsatisfiable set inconsistent. Which of the rules of **LK** were necessary to prove completeness? Are any of these rules not used anywhere in the proof? In order to answer these questions, make a list or diagram that shows which of the rules of **LK** were used in which results that lead up to the proof of [Theorem 8.16](#). Be sure to note any tacit uses of rules in these proofs.

Problem 8.6. Prove (1) of [Theorem 8.19](#).

Problems for Chapter 9**Problems for Chapter 10****Problems for Chapter 11**