Introduction to Mathematical Logic

A. Vasudevan

Contents

CHAPTER 1

Introduction to Mathematical Logic

1. What is 'Mathematical Logic'?

Logic, in the most general sense of the term, refers to the study of the norms that govern the activity of reasoning. Thus understood, logic comprehends not only the sort of reasoning that is expressed in mathematical proofs, but also such reasoning as underwrites statistical inference or scientific induction. As such, the first and most obvious function of the qualifier 'mathematical', when prefixed to the term logic, is to restrict the subject matter of the inquiry to just such norms as govern the particular sort of reasoning that is done in mathematics (e.g., that by which one deduces various geometrical theorems from basic axioms concerning points, lines, and other such elementary notions).

There is, however, a second interpretation of the phrase 'mathematical logic', according to which, the term 'mathematical' functions as an adverbial modifier, indicating something about the methodology by which the study is to be pursued rather than delimiting its object. In particular, the term implies that the methods to be used are themselves mathematical in nature. Thus, just as the phrase 'mathematical psychology' does not refer to the psychological study of mathematicians, but instead to research which employs mathematical techniques to investigate formal models of psychological processes, 'mathematical logic', in this latter sense of the phrase, does not refer to the study of such reasoning as is done by mathematicians, but instead to the study of reasoning, *done mathematically*.

In this course, the phrase 'mathematical logic' will be understood in *both* of the above senses. On the one hand, we will be exclusively concerned with demonstrative reasoning of the sort which ensures that the truth of a conclusion follows from a given set of premises in just the way that the truth of a theorem in geometry follows from a given set of axioms. Moreover, in Part II of the course, we will consider such reasoning in its specific application to the particular structures studied in mathematics. We will, for instance, consider systems of reasoning designed for the express purpose of validating arithmetical inferences, such as: "from $3x + 4y = -2$ and $y = 1$, it follows that $x = -2$."[1] At the same time, however, our study of logic will be 'mathematical' in that the systems of reasoning that we will consider will themselves be represented as mathematically well-defined structures, and the claims that we will make about these systems will be established using standard mathematical techniques.

As will become more readily apparent in Part II of the course, it is this duality of meaning which endows mathematical logic not only with its peculiar charm, but also with the greater part of its philosophical value: if mathematical logic is concerned to study the sort of reasoning that is done in mathematics and if, at the same time, it is itself just a branch of mathematics, then one may meaningfully ask what follows when the claims established in the study of mathematical logic are applied to itself. The answer, as it turns out, is quite a lot! The essentially self-reflective nature of mathematical logic is what allows for the precise formulation of such provocative theses as "arithmetical truth cannot be defined arithmetically," and "if a computer can simulate all of our mathematical reasoning, then, in principle, we cannot fully grasp how it works."[2] Unfortunately, a more careful analysis of these claims must be left for Part II of the course. For now, let us consider in more detail how the study of logic can be made mathematical.

2. Making Logic Mathematical

Logic seeks to investigate the norms that govern the activity of reasoning. As such, it is concerned to answer questions of the form:

[1] Indeed, by the end of Part I of the course, we will be in a position to formulate a single system of reasoning, involving only a handful of axiom schemas and a single rule of inference, which is sufficient to represent all but an extremely small fraction of the proofs that have ever been given in mathematics!

[2] To the reader: If such claims inspire in you anything more than an initial, vague sense of wonder, then the material covered in the first part of this course may be too basic for you.

- Is a given method of reasoning such as to only result in valid inferences or, by employing this method, will one sometimes be led to draw a conclusion which does not follow from one's initial assumptions?
- If an inference is valid, will one be able to draw this inference by employing a given method of reasoning, or will there be some valid inferences which, in principle, cannot result from employing this method.

In order to render such questions mathematically precise, the following tasks must first be performed:

(1) 'Inferences' must be represented as well-defined mathematical structures.

(2) The standard for normative assessment by which a given inference is judged to be 'valid' or 'invalid', must be expressed in terms of the possession by that inference of a mathematically well-defined property.

(3) The notion of a 'method of reasoning' must be described in such a way that it becomes a mathematical fact whether or not by means of a given method of reasoning, one is able to draw a given inference.

Once these three tasks have been performed, then the two questions listed above become no less mathematically well-defined than the question: Is 4,753,234,127 a prime number?

2.1. Arguments. In informal terms, an argument is a collection of propositions, one of which is distinguished as that 'for the sake of which' the others are set forth. The distinguished proposition is referred to as the conclusion of the argument and the propositions offered in its support as the argument's premises.

Although arguments are typically expressed by a collection of sentences in a language, it is natural to regard an argument as something distinct from its linguistic expression (one may, for example, wish to allow for the possibility that the same argument can be expressed in either German or French). Nevertheless, for the purpose of formalization, we will disregard the intuitive difference between an argument and its expression, and construe arguments as purely *syntactical* constructs. In other words, we will represent an argument by a particular collection of sentences expressed in a given syntax.[3]

The syntax in which we will formulate arguments will not be that of a natural language such as English, but will instead be introduced in the context of an artificial or 'formal' language devised for the express purpose of representing certain argument forms. We adopt this approach so as to minimize the syntactic complexity of the arguments that we will consider. The rules determining when a sequence of symbols in the English alphabet (or a sequence of words in the English lexicon) constitute a grammatically well-formed sentence are exceedingly complex, and much of this complexity is unnecessary for capturing the differences in meaning that are relevant for assessing the particular sort of arguments with which we will be concerned.[4]

In each of the formal languages that we will consider, the syntax of the language will be introduced in two distinct stages:

(1) First, we will specify an *alphabet* for the language, i.e., a set of symbols. Any finite string of symbols from this alphabet will be referred to as an *expression* of the language.

(2) Second, we will specify *formation rules* determining when a given expression constitutes a *sentence* of the language.

When they are defined in this way, sentences are purely syntactical constructs – they are strings of symbols, devoid of any meaning. Nevertheless, our choice of the syntax for a language will, in all cases, be guided by semantical considerations. In other words, we will always have in mind a certain conception of how various syntactical items are to be interpreted, and we will choose our syntax so as to ensure that the sentences in the language consist only of expressions, which, when given such an interpretation, express meaningful propositions.

[3]We will later give a formal account of when two sentences express the same proposition, or, in the terminology we will adopt, when two sentences are *logically equivalent*, and we will define the validity of arguments in such a way as to ensure that validity is preserved through substitution by equivalents.

[4]Thus, for example, while there may be some subtle difference in meaning (or emphasis) between the sentence 'John gave the letter to Mary,' and the sentence 'Mary was given the letter by John,' this difference in meaning (while relevant perhaps for producing a desired poetic effect) is altogether irrelevant for judging whether or not it follows from this claim that 'Mary was given something by someone.'

DEFINITION 1.1. An *argument* in a formal language consists of (i) a (possibly empty[5]) set of sentences of the language, whose members are referred to as the argument's *premises*; and (ii) a sentence of the language referred to as the argument's *conclusion*.

We denote the argument whose premises comprise the set Γ and whose conclusion is ψ by

$$\frac{\Gamma}{\psi}$$

(for the sake of readability, we will often omit the outer braces from the premise list).

EXAMPLE 1. In this example, we will devise a formal language $\mathscr{L}_\times$ suitable for expressing a very simple class of arithmetical arguments, namely, those which only consist of equations involving a single variable, x, and in which both sides of the equality are expressed as the product of numbers belonging to the set $\{1, 0\}$.

The alphabet for our language is given by the set of symbols:

$$\{x, 1, 0, \times, =\}.$$

We refer to the symbols 'x', '1', and '0' as *numerals*.

DEFINITION 1.2. A *numerical expression* is any expression which can be judged as such on the basis of the following two facts:

(1) Every expression consisting of a single numeral is a numerical expression.

(2) If α and β are numerical expressions, then the expression $\alpha \times \beta$ is a numerical expression.[6]

It follows from this definition that the numerical expressions are all and only the expressions of the form:

$$n_1 \times n_2 \times \cdots \times n_k,$$

where $n_1, \ldots, n_k$ are numerals.

The formation rule for the language is as follows:

Formation Rule: An expression is a *sentence* of the language iff it is of the form $\alpha = \beta$, where α and β are numerical expressions.

In accordance with this rule, the following expressions are all sentences of the language:

(i) $x = 0$

(ii) $1 = x$

(iii) $1 \times 0 = 0 \times 1$

(iv) $x \times x = 1$

(v) $x \times 0 \times 1 \times x = 1$

(vi) $x \times 1 \times 1 \times 1 = 1 \times x \times 0$

An example of an argument expressed in this language is

[5] In what follows, it will be convenient to assume that there exists a unique set containing no members, which we will denote by $\emptyset$ and refer to as the empty set. We will henceforth assume that (i) for every set S, $\emptyset \subset S$; (ii) for every property P, it is (vacuously) true that for all $x \in \emptyset$, x has the property P; and (iii) for every property P, it is (vacuously) false that for some $x \in \emptyset$, x has the property P.

$$\frac{x \times 0 = 0 \times 1, \quad 1 \times x \times 1 = 1}{1 = x}$$

Exercise 1.1. We refer to the number of symbols comprising an expression as the *length* of that expression.

(a) Show that the length of every sentence is odd.

(b)* If k is an odd number greater than 2, how many distinct expressions are there of length k? How many distinct sentences are there of length k? Show that the probability with which a randomly chosen expression of length k is a sentence goes to 0 as k increases.

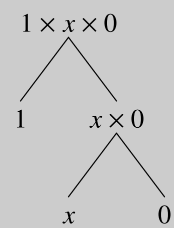
Exercise 1.2 (**). A *binary tree* is a tree structure in which every node in the tree has either exactly two direct descendants (referred to as its 'left' and 'right' children, respectively), or is a terminal node (i.e., a node with no direct descendants). By a *formation tree* for an expression α, we mean a binary tree the nodes of which are all expressions, and for which:

(1) The root node of the tree (i.e., the unique node with no parents) is α.

(2) If γ and γ' are the left and right children of the node β, then β is the expression $\gamma \times \gamma'$.

(3) Every terminal node is an expression consisting of a single numeral.

For example, the following tree is a formation tree for the expression $1 \times x \times 0$:



It is easy to see that an expression φ is a numerical expression iff there is a formation tree for φ, but a single numerical expression can have multiple distinct formation trees. Show that there are exactly

$$\frac{(2k-2)!}{k!(k-1)!} = \frac{(2k-2)\cdot(2k-3)\cdots(k+1)}{(k-1)\cdot(k-2)\cdots 1}$$

distinct formation trees for any numerical expression involving k numerals. (Hint: Prove this claim by induction on k. First, show that the claim holds for $k = 2$. Then show that if the claim holds for all k, such that $2 \le k \le n$, then the claim must also hold for $k = n + 1$.)

2.2. Validity. Consider the following hackneyed syllogism:

[6]A definition of this type is called an *inductive* (or recursive) definition. To define a class of objects inductively, one first stipulates that a certain set of objects belong to the class. This stipulation is referred to as the *base clause* of the definition. One then specifies a set of laws, each of which asserts that if an object stands in a certain relation to certain objects belonging to the class, then it too belongs to the class. These laws are referred to as the *inductive clauses* of the definition. The class itself is then defined as consisting of *only* those objects which can be judged to belong to the class by appeal to both the base clause and finitely many instances of the inductive clauses of the definition.

$$\frac{\text{All men are mortal, Socrates is a man.}}{\text{Socrates is mortal.}}$$

Intuitively, this is a valid argument, but what is it about this argument that *makes* it valid? The most immediate answer is that the argument is valid because the truth of the argument's conclusion follows necessarily from the truth of its premises. But, in what sense of 'necessity' does Socrates' mortality necessarily follow from Socrates's being a man and from the fact that all men are mortal?

In general, to say that a claim is necessarily true is to assert not only that the claim itself is true, but moreover that its truth is independent of certain features of the context of utterance in which the claim is made. Thus, when we say that it is necessarily the case that if the premises of the above argument are true, so too is its conclusion, not only do we claim that:

(P) If 'All men are mortal' and 'Socrates is a man' are both true, then 'Socrates is mortal' is true.

but, further, that (P) *would* be true, regardless of certain features of the actual context in which this claim is made. The precise sense of the implicated modality is determined by which features of the context of utterance are deemed accidental. In this case, when we say that (P) is necessarily true, we mean that it would be true no matter who (or what) the name 'Socrates' refers to, and no matter what properties are designated by the phrases 'man' and 'mortal'. If the name 'Socrates' were to refer to Kurt Gödel, the phrase 'man', to the property of being German, and the phrase 'mortal' to the property of being a poor logician, it would no less follow from the truth of the sentences 'All men are mortal' and 'Socrates is a man,' that 'Socrates is mortal' is true (for if all Germans are poor logicians, and Gödel is a German, then Gödel, too, is a poor logician).

Note that it is crucial to this line of reasoning that the meaning of the linguistic constructs 'All $\cdots$ are $--$' and '$\cdots$ is a $--$,' as they appear in the premises and conclusion of the argument, are taken to operate in their usual way, i.e., the former asserts that all objects possessing the property designated by '$\cdots$' also possess the property designated by '$--$', and the latter asserts that the object named by '$\cdots$' possesses the property designated by '$--$'. If the meaning of these constructs were allowed to vary along with 'Socrates', 'mortal' and 'man' (e.g., if 'All $\cdots$ are $--$' were interpreted as asserting that only *some* objects possessing property '$\cdots$', possess property '$--$') then there may be ways of interpreting the premises and conclusion of the argument according to which the premises are true and the conclusion false.

As a second example, consider the following elementary arithmetical argument:

$$\frac{x + 2y = -2,\ y > 1}{8 < -2x}$$

To say that this argument is valid is to assert that no matter what numbers are referred to by the variables x and y, if it is true that $x + 2y = -2$ and $y > 1$, then it is also true that $8 < -2x$. Here we are assuming that the numerals '1', '2' and '8' refer to the numbers 1, 2 and 8, respectively; that the symbol '+' refers to the operation of addition; and that the symbols '=', '<' and '>' refer to the relations of equality, being less than and being greater than, respectively.

As these examples illustrate, in assessing the validity of an argument, we must first distinguish between those elements of the syntax that are taken to have a fixed meaning for the purpose of the assessment, and those elements of the syntax whose meanings are allowed to vary. We refer to the former as the *logical* elements of the syntax and the latter as its *non-logical* elements. To say that an argument is *valid* is to assert that (under the assumption that the logical elements of the syntax are given their standard meaning) the truth of the argument's conclusion follows from the truth of its premises *no matter how the non-logical elements of the syntax are interpreted*.

Let us now try to formulate this idea in mathematically precise terms. We first define a *valuation* of a formal language to be any function *val* which assigns to each of the sentences of the language one of the values in the set $\{\mathbf{T}, \mathbf{F}\}$ (we refer to the members of this set as *truth-values*).[7] $val(\varphi) = \mathbf{T}$ means that the sentence φ is true under the valuation *val*, and $val(\varphi) = \mathbf{F}$ means that it is false under this valuation.

When we assert that an argument is valid, we will always do so relative to a certain class of valuations. The validity of an argument expresses the fact that for every valuation in this class, if all the premises of the argument are true under this valuation, then the conclusion is true under this valuation as well.

[7] There is nothing special about the symbols **T** and **F** (excepting their pnemonic value). We could have just as well used 1 and 0, or any other two distinct objects to serve as names for truth and falsity.

Definition 1.3. The argument

$$\frac{\Gamma}{\psi}$$

is *valid* with respect to the class of valuations V if, for all valuations $val \in V$, the following is true: if $val(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$, then $val(\psi) = \mathbf{T}$. To express this fact, we write:

$$\Gamma \models_V \psi.$$

If Γ is the singleton set $\{\varphi\}$, we write $\varphi \models_V \psi$ to express the fact that $\Gamma \models_V \psi$.

Exercise 1.3. Let $\varphi_1, \ldots, \varphi_n, \psi$ $(n \geq 2)$ be distinct sentences.

(a) Construct a two-member set of valuations V such that:

(i) $\{\varphi_1, \ldots, \varphi_n\} \models_V \psi$; and

(ii) $\varphi_i \not\models_V \psi$, for $i = 1, \ldots, n$.

(b) Show that if V is a set consisting of m valuations, where $m < n$, then $\varphi_i \models_V \varphi_j$, for some $i \neq j$.

Theorem 1.1. *Let V be the set of all possible valuations of the language. Then $\Gamma \models_V \psi$ if and only if $\psi \in \Gamma$.*

Proof. Suppose that $\psi \in \Gamma$. Then, clearly it is impossible for a valuation to assign to all of the sentences in Γ the value $\mathbf{T}$ and to ψ the value $\mathbf{F}$. Hence, $\Gamma \models_V \psi$. Now, suppose that $\psi \notin \Gamma$. Then, since V consists of *all* possible valuations, there exists some valuation in V which assigns to all the sentences in Γ the value $\mathbf{T}$ and to ψ the value $\mathbf{F}$. Hence, $\Gamma \not\models_V \psi$. □

Theorem 1.1 asserts that an argument is valid with respect to the class of all valuations of a language if and only if its conclusion appears among its premises. Consequently, without somehow restricting the class of admissible valuations, we are left with a trivial standard of argumentative validity.

In general, the class of valuations that we will consider will be restricted by means of the following three-stage process:

(1) First, a certain set of expressions will be identified as the non-logical elements of the syntax.

(2) Second, with each non-logical element of the syntax we will associate a set of possible values (or interpretants). An *interpretation* of the language will then be any function which assigns to each non-logical element of the syntax a particular value from this set.

(3) Third, we will formulate a set of *semantical rules* which associate with each possible interpretation, σ, a unique valuation of the language, val^{σ}.[8]

Henceforth, whenever we assess the validity of an argument, it will always be assumed that our assessment occurs in a setting in which these three tasks have already been completed, and that the assessment is to be made with respect to the class of all valuations of the form val^{σ}, where σ is a possible interpretation of the language.[9] Thus, the more general Definition 1.3, may be replaced with the following:

[8] To express the fact that $val^{\sigma}(\varphi) = \mathbf{T}$ (resp: $\mathbf{F}$), we say that φ is 'true (resp: false) under the interpretation σ'.

[9] In general, the semantical rules for the language (which serve to encode the meanings of the logical elements of the syntax) will be chosen in such a way that not every valuation will be of the form val^{σ}. Hence, it is these rules that will render our standard of validity nontrivial.

DEFINITION 1.4. The argument

$$\frac{\Gamma}{\psi}$$

is *valid* if, for all interpretations σ, the following is true: If $val^{\sigma}(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$, then $val^{\sigma}(\psi) = \mathbf{T}$. To express this fact, we write:

$$\Gamma \models \psi.$$

In this case, we say that ψ is a *logical consequence* of Γ. If Γ is the singleton set $\{\varphi\}$, we write $\varphi \models \psi$ to express the fact that $\Gamma \models \psi$.

The following two important semantic notions can be defined in terms of validity.

DEFINITION 1.5. A sentence φ is *logically true* (written $\models \varphi$) iff $\emptyset \models \varphi$.

DEFINITION 1.6. Two sentences φ and ψ are *logically equivalent* (written $\varphi \equiv \psi$) iff $\varphi \models \psi$ and $\psi \models \varphi$.

Note that it follows from definitions 1.4, 1.5 and 1.6 that $\models \varphi$ iff $val^{\sigma}(\varphi) = \mathbf{T}$, for all interpretations σ, and that $\varphi \equiv \psi$ iff $val^{\sigma}(\varphi) = val^{\sigma}(\psi)$, for all interpretations σ.

EXERCISE 1.4. We say that a sentence φ is a *logically false* if $val^{\sigma}(\varphi) = \mathbf{F}$, for all interpretations σ. We say that sentence is *logically contingent* if it is neither logically false nor logically true.

(a) Suppose that $\bot$ is a logically false sentence. Define logical falsehood in terms of $\bot$ and $\models$.

(b) Under what conditions do the following claims entail that φ is logically contingent:
   (a) For some sentence ψ, $\varphi \not\models \psi$.
   (b) For some sentence ψ, $\psi \not\models \varphi$.
   (c) For some sentence ψ, $\varphi \models \psi$, but ???

The following theorem states a few elementary facts that follow from our definition of logical consequence:

THEOREM 1.2.

(i) If $\Gamma \models \psi$ and for every $\varphi \in \Gamma$, $\Delta \models \varphi$, then $\Delta \models \psi$.

(ii) If $\Gamma \models \psi$ and $\Gamma \subset \Delta$, then $\Delta \models \psi$.

(iii) If $\Gamma \subset \Delta$ and, for every $\varphi \in \Delta - \Gamma$, $\models \varphi$, then $\Delta \models \psi$ iff $\Gamma \models \psi$.

PROOF.

(i) Let σ be an interpretation such that $val^{\sigma}(\varphi) = \mathbf{T}$ is true for all sentences $\varphi \in \Delta$. Then, since $\Delta \models \varphi$, for all $\varphi \in \Gamma$, $val^{\sigma}(\varphi) = \mathbf{T}$ is true for all $\varphi \in \Gamma$. But since, by assumption, $\Gamma \models \psi$, $val^{\sigma}(\psi) = \mathbf{T}$. Hence, any interpretation under which all of the sentences in Δ are true, is one in which ψ is true as well, i.e., $\Delta \models \psi$.

(ii) Let σ be an interpretation such that $val^{\sigma}(\varphi) = \mathbf{T}$ is true for all sentences $\varphi \in \Delta$. Since $\Gamma \subset \Delta$, $val^{\sigma}(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$, and so, since $\Gamma \models \psi$, $val^{\sigma}(\psi) = \mathbf{T}$. Hence, any interpretation under which all of the sentences in Δ are true, is one in which ψ is true as well, i.e., $\Delta \models \psi$.

(iii) The if-part of the claim follows from (ii) and the fact that $\Gamma \subset \Delta$. To prove the only-if part of the claim, suppose that $\Delta \models \psi$, and let σ be an interpretation such that $val^{\sigma}(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$. Since every sentence in $\Delta - \Gamma$ is a logical truth, $val^{\sigma}(\varphi) = \mathbf{T}$, for all $\varphi \in \Delta - \Gamma$, and so, since $\Delta = \Gamma \cup (\Delta - \Gamma)$, $val^{\sigma}(\varphi) = \mathbf{T}$, for all $\varphi \in \Delta$. But since $\Delta \models \psi$, $val^{\sigma}(\psi) = \mathbf{T}$. Hence, any interpretation under which all of the sentences in Γ are true, is one in which ψ is true as well, i.e., $\Gamma \models \psi$.

□

Note that by putting $\Gamma = \{\varphi\}$ and $\Delta = \{\rho\}$ in (i), we obtain the result that if $\rho \models \varphi$ and $\varphi \models \psi$, then $\rho \models \psi$, i.e., the relation of logical consequence, when restricted to sentences, is a transitive relation. Also note that by putting $\Gamma = \emptyset$ in (ii), we obtain the result that any argument whose conclusion is a logical truth is valid.

Example 2. In this example, we devise a semantical framework within which to assess the validity of arguments expressed in the formal language $\mathscr{L}_{\times}$ introduced in Example 1 (see page 6).

The only non-logical element of the syntax is the variable symbol x. We stipulate that the possible semantic values of this symbol are given by the set $\{0, 1\}$.[10] Thus, there are only two possible interpretations of the language, one in which x is assigned the value 0 and one in which x is assigned the value 1. We refer to these interpretations as σ_0 and σ_1, respectively.

The semantical rules (or, in this case, rule) for the language must define a unique valuation for each of these two interpretations. In stating this rule, we will take for granted that we know what it is for a sentence in the language which only involves the numerals '0' and '1' to be a true (or false) arithmetical claim.

Semantical Rule: For $i = 0, 1$, $val^{\sigma_i}(\varphi) = \mathbf{T}$ iff the sentence which results from subtituting $\sigma_i(x)$ for all occurrences of x in φ expresses a true arithmetical claim.

To see how this rule is to be applied, let φ be the sentence:

$$x \times 1 = 1 \times 0$$

According to the semantical rule for our language, $val^{\sigma_0}(\varphi) = \mathbf{T}$ just in case the sentence which results from substituting $\sigma_0(x)$ for every occurrence of x in φ is a true arithmetical claim. In this instance, since $\sigma_0(x)$ is the symbol 0, the sentence resulting from this substitution is:

$$0 \times 1 = 1 \times 0.$$

This, of course, is a true arithmetical claim, hence $val^{\sigma_0}(\varphi) = \mathbf{T}$. On the other hand, $val^{\sigma_1}(\varphi) = \mathbf{F}$, since

$$1 \times 1 = 1 \times 0,$$

is *not* a true arithmetical claim.

As a simple example, in the context of this semantical framework, it is easy to verify that:

$$\{x \times 0 = 0 \times 1, \quad 1 \times x \times 1 = 1\} \models 1 = x,$$

for the only intepretation under which all the premises of the argument are true is σ_1, and the sentence $1 = x$ is true under this interpretation, as well.

Exercise 1.5. Show that for every sentence φ of $\mathscr{L}_{\times}$, φ is logically equivalent to one and only one of the following four sentences: (i) $x = 0$, (ii) $x = 1$, (iii) $0 = 0$, (iv) $0 = 1$. (In what follows, we will refer to this sentence as the *reduced* form of φ, and denote it φ^R).

For the purpose of formulating semantical rules and assessing the truth-values of sentences under a given interpretation it will often be convenient to assign semantical values not just to the non-logical elements of the syntax but also to certain expressions containing these elements (or, equivalently, to treat these expressions themselves as the non-logical elements of the syntax). For example, with respect to the language $\mathscr{L}_{\times}$, it will be convenient to be able to refer not just to the value of x under a given interpretation, but, more generally, to the value of any numerical expression under a given interpretation.

Let α be a numerical expression. For $i = 0, 1$, we let $\sigma_i(\alpha) = 0$ if the result of substituting $\sigma_i(x)$ for each occurrence of x in α is an arithmetical expression which evaluates to 0. Otherwise, $\sigma_i(\alpha) = 1$. Adopting this notation, the above semantical rule can now be expressed as follows:

Semantical Rule: Let α and β be any numerical expressions. For $i = 0, 1$, $val^{\sigma_i}(\alpha = \beta) = \mathbf{T}$ iff $\sigma_i(\alpha) = \sigma_i(\beta)$.

2.3. Deductive Systems. A method of reasoning is any method by which one proposes to assess the validity of an argument. Based on the definition of validity offered in the previous section, the most direct method of reasoning consists in evaluating the truth-values of the sentences comprising a given argument under all the various possible interpretations of the language. The argument is then judged to be valid if and only if no interpretation is found under which all of the premises of the argument are true and the conclusion is false.

Note that while the correctness of this method follows immediately from the definition of validity, it clearly does *not* represent the way in which reasoning ordinarily proceeds. To see this, consider, again, the arithmetical argument:

$$\frac{x + 2y = -2,\ y > 1}{8 < -2x}$$

If one were asked to assess the validity of this argument, one would most certainly *not* proceed by substituting in particular numerical values for x and y in an effort to ascertain whether or not there exists a particular interpretation of these variables under which the premises of the argument are true and its conclusion false.[11] Rather, to assess this argument's validity, one would most likely proceed by attempting to construct a proof which shows that the conclusion can be deduced from the premises by appeal to elementary arithmetical facts and simple rules governing the operation of numbers of the sort which are introduced in any first course in algebra.

In the typical language in which arguments in mathematics are expressed, such a proof might take the following form: If $y > 1$ and $x + 2y = -2$, it follows that

$$-2 > x + 2.$$

Subtracting 2 from both sides of this inequality yields

$$-4 > x,$$

and multiplying both sides of this inequality by -2 yields

$$8 < -2x,$$

which is the desired result.

Since the methods employed in standard mathematical reasoning are methods of proof, it is with these methods that we will be exclusively concerned. To provide a suitable characterization of these methods, we must first offer a more precise definition of the notion of a *proof*.

To begin with, let us express the above proof in slightly more detail. In the abbreviated form given above, a number of intermediate steps in the reasoning were taken for granted. To help us to make explicit these steps, we will impose strict constraints on what we are allowed to assume. In particular, we will assume only that (1) given any two numbers, we are able to judge which of the two is the greater, and so judge true such simple inequalities as $4 > 2$ and $-3 > -6$; and (2) that we are able to evaluate simple arithmetical expressions, and so judge true such simple equalities as $2 + 3 = 5$ and $-2 \cdot 3 = -6$.

Every other law or principle appealed to in the argument must be stated explicitly. Specifically, in order to carry out the proof, we will need to appeal to the following laws of arithmetic. For any numerical expressions α, β, γ:

[10] To be more precise, this set consists of the *symbols* '0' and '1'. It may seem odd to allow the semantic value of a symbol in the language to be another symbol in the language, but, formally, there is nothing to prevent us from doing so. Moreover, adopting this approach will allow us to refer directly to the expression which results from substituting the value of x for all occurrences of x in a given expression.

[11] This is, at least in part, because there are an *infinite* number of possible interpretations of x and y, so that, regardless of how many interpretations we consider, we could only ever claim inductive grounds for judging the argument to be valid.

(R1) If $\alpha > 0$ and $\beta > \gamma$, then $\alpha\beta > \alpha\gamma$.

(R2) If $\alpha < 0$ and $\beta > \gamma$, then $\alpha\beta < \alpha\gamma$.

(R3) If $\alpha > \beta$, then both $\gamma + \alpha > \gamma + \beta$ and $\alpha + \gamma > \beta + \gamma$.

(R4) $\alpha + 0 = \alpha$

(R5) If φ and ψ are statements which differ only in that α appears in φ where β appears in ψ, then if $\alpha = \beta$ and φ is true, then ψ is true.

Having laid down these principles, the proof may now be expressed as follows:

(1) Since $2 > 0$ and (by assumption) $y > 1$, it follows from (R1) that $2y > 2 \cdot 1$.

(2) Since $2 \cdot 1 = 2$ and (from 1) $2y > 2 \cdot 1$, it follows from (R5) that $2y > 2$.

(3) Since (from 2) $2y > 2$, it follows from (R3) that $x + 2y > x + 2$.

(4) Since (by assumption) $x + 2y = -2$ and (from 3) $x + 2y > x + 2$, it follows from (R5) that $-2 > x + 2$.

(5) Since (from 4) $-2 > x + 2$, it follows from (R3) that $-2 + (-2) > x + 2 + (-2)$.

(6) Since $-2 + (-2) = -4$ and (from 5) $-2 + (-2) > x + 2 + (-2)$, it follows from (R5) that $-4 > x + 2 + (-2)$.

(7) Since $2 + (-2) = 0$ and (from 6) $-4 > x + 2 + (-2)$, it follows from (R5) that $-4 > x + 0$.

(8) Since (from R4) $x + 0 = x$ and (from 7) $-4 > x + 0$, it follows from (R5) that $-4 > x$.

(9) Since $-2 < 0$ and (from 8) $-4 > x$, it follows from (R2) that $-2 \cdot -4 < -2x$.

(10) Since $-2 \cdot -4 = 8$ and (from 9) $-2 \cdot -4 < -2x$, it follows from (R5) that $8 < -2x$. □

What is important to note about this proof is that all of the claims that are inferred in the course of the reasoning fall into one of the following three categories:

(i) Elementary truths (or axioms) of arithmetic (e.g., $2 > 0$ or $-2 \cdot -4 = 8$);

(ii) Premises of the argument (e.g., $x + 2y = -2$); or

(iii) The result of applying one of the principles (R1)-(R5) to claims already established in the proof.

This can be emphasized by omitting from the proof such descriptive locutions as 'since ... it follows that −−', and instead simply listing the claims that are appealed to in the proof, indicating, for each such claim, into which of the above three categories it falls:

(1) $2 > 0$ [Axiom]

(2) $y > 1$ [Premise]

(3) $2y > 2 \cdot 1$ [R1 - 1,2]

(4) $2 \cdot 1 = 2$ [Axiom]

(5) $2y > 2$ [R5 - 4,3]

(6) $x + 2y > x + 2$ [R3 - 5]

(7) $x + 2y = -2$ [Premise]

(8) $-2 > x + 2$ [R5 - 7,6]

(9) $-2+(-2) > x+2+(-2)$ [R3 - 8]

(10) $-2+(-2) = -4$ [Axiom]

(11) $-4 > x+2+(-2)$ [R5 -10,9]

(12) $2+(-2) = 0$ [Axiom]

(13) $-4 > x+0$ [R5 - 12,11]

(14) $x+0 = x$ [R4]

(15) $-4 > x$ [R5 - 14,13]

(16) $-2 < 0$ [Axiom]

(17) $-2 \cdot -4 < -2x$ [R2 - 16,15]

(18) $-2 \cdot -4 = 8$ [Axiom]

(19) $8 < -2x$ [R5 - 18,17]

We will take this expanded arithmetical proof as a model of the proofs that will be given in our formal languages. Based on this model, we may first observe that a proof is always given relative to both a set of axioms, which we can take for granted in our reasoning, and certain rules of inferences by which we can infer new claims from claims already proven. Taken together, these axioms and rules form what we shall call a deductive system:[12]

Definition 1.7. A *deductive system* consists of (i) a set of sentences, the members of which are referred to as *axioms*; and (ii) a set of *rules of inference*, each of which states that certain sentences can be inferred from others.

We can now define what we mean by a proof (in a deductive system):

Definition 1.8. Let T be a deductive system. If Γ is a set of sentences and ψ a sentence, then a *proof of* ψ *from* Γ *in* T is a finite sequence of sentences $\varphi_1, \varphi_2 \ldots, \varphi_k$, satisfying the following two properties:

(1) $\varphi_k = \psi$.

(2) For all $i = 1, \ldots, k$, either:

   (i) $\varphi_i \in \Gamma$

   (ii) φ_i is an axiom of T, or

   (iii) φ_i follows from one of the rules of T applied to certain sentences among $\varphi_1, \varphi_2 \ldots, \varphi_{i-1}$.[13]

To express the fact that there exists a proof of ψ from Γ in T, we write:

$$\Gamma \vdash_T \psi.$$

In this case, we say that ψ is a *deductive consequence* of Γ in T, or simply that ψ is *provable* from Γ in T.

[12]Such a system is sometimes referred to as a Hilbert style deductive system, in honor of the mathematician David Hilbert, who was the first to propose the mathematical treatment of such systems. The term Hilbert system has also come to possess a more narrow meaning, indicating the further fact that the system includes a relatively large number of axiom schemes and relatively few rules of inference.

Note that the class of deductive consequences of Γ in T can be given the following inductive definition:

(1) If ψ is an axiom of T, then $\Gamma \vdash_T \psi$.

(2) If $\psi \in \Gamma$, then $\Gamma \vdash_T \psi$.

(3) If $\Gamma \vdash_T \varphi_i$ for $i = 1, \dots, k$, and if ψ can be inferred from $\varphi_1, \dots, \varphi_k$ by some rule of T, then $\Gamma \vdash_T \psi$.

Exercise 1.6. Show that provability in T is monotonic in the following sense: if $\Gamma \vdash_T \psi$ and for every $\varphi \in \Gamma$, $\Gamma' \vdash_T \varphi$, then $\Gamma' \vdash_T \psi$.

The proof-theoretic analogue of logical truth is ‘theoremhood’.

Definition 1.9. A sentence φ is a *theorem* of T (written $\vdash_T \varphi$) iff $\emptyset \vdash_T \varphi$.

We will henceforth represent methods of reasoning by deductive systems. When we say that a method of reasoning is able to ‘produce’ the argument

$$\frac{\Gamma}{\psi}$$

what we will mean by this claim is that one is able to prove ψ from Γ in a given deductive system.

Example 3. In this example, we introduce a deductive system, $T_\times$, in the context of which we may construct proofs in the formal language $\mathscr{L}_\times$ introduced in Example 1 (see page 6). T has three axioms:

$$x = x \qquad 0 = 0 \qquad 1 = 1,$$

and ten rules of inference.

The first six rules concern the introduction and elimination of numerals from numerica expressions. The first four of these rules can be stated schematically as follows (here, and in what follows, α and β are any numerical expressions):

(1–) $\dfrac{1 \times \alpha = \beta}{\alpha = \beta}$ (0–) $\dfrac{0 \times \alpha = \beta}{0 = \beta}$

(1+) $\dfrac{\alpha = \beta}{1 \times \alpha = \beta}$ (0+) $\dfrac{0 = \beta}{0 \times \alpha = \beta}$

The fifth and sixth rules are:

(x–) If x occurs in α, then:

$$\frac{x \times \alpha = \beta}{\alpha = \beta}$$

(x+) If x occurs in α, then:

$$\frac{\alpha = \beta}{x \times \alpha = \beta}$$

[13] We adopt the notational convention that if j does not appear in the continuation of the pattern indicated by i_1 and i_2, then the expression $x_{i_1}, x_{i_2}, \dots, x_j$ denotes the empty set. Thus, e.g., if $i = 1$, condition (2.iii) in Definition 1.8 states that φ_1 follows from one of the rules of T applied to sentences in the empty set.

The seventh rule expresses the symmetricality of the equality relation,

$$(\mathrm{S}) \qquad \frac{\alpha = \beta}{\beta = \alpha}\ ,$$

and the eigth rule, its transitivity

$$(\mathrm{T}) \qquad \frac{\alpha = \beta\ ,\ \beta = \gamma}{\alpha = \gamma}\ .$$

The ninth rule asserts that from $0 = 1$, anything follows:

$$(\neq) \qquad \frac{0 = 1}{\varphi}\ .$$

The tenth and final rule expresses the interchangeability of terms in a numerical expression:

$$(\times) \qquad \frac{n_1 \times n_2 \times \cdots \times n_k = \alpha}{n_{x_1} \times n_{x_2} \times \cdots \times n_{x_k} = \alpha}\ ,$$

Here $x_1, \ldots, x_k$ are numbers in the set $\{1, \ldots, k\}$ such that $x_i \neq x_j$, for all $i \neq j$.

It is easy to show that:

$$\{x \times 0 = 0 \times 1, \quad 1 \times x \times 1 = 1\} \vdash_{T_\times} 1 = x.$$

The proof proceeds as follows:

(1) $1 \times x \times 1 = 1$ [Premise]

(2) $x \times 1 = 1$ [1–,1]

(3) $1 \times x = 1$ [×, 2]

(4) $x = 1$ [1–, 3]

(5) $1 = x$ [S, 4]

Exercise 1.7. Show that if a sentence is of any of the following three forms:

(a) $0 \times \alpha = 0 \times \beta$

(b) $\alpha = \alpha$

(c) $1 \times \alpha = \alpha$

then it is a theorem of $T_\times$.

It may be noted that with the exception of (T) and (≠), all of the other rules in the system are 'reversible', in the sense that if ψ can be inferred from φ by one of these rules, then φ can be inferred from ψ by one of these rules, as well. To capture this idea, we introduce the following notion:

Definition 1.10. A *direct proof* of ψ from φ is a sequence of sentences $\varphi_1, \ldots, \varphi_k$, such that:

(1) $\varphi_1 = \varphi$

(2) $\varphi_k = \psi$

(3) For $i = 2, \ldots, k$, φ_i can be inferred from φ_{i-1} by means of one of the rules (1−), (1+), (0−), (0+), (x−), (x+), (S) or (×).

Clearly, every direct proof of ψ from φ is a proof of ψ from φ in $T_\times$. Moreover, if $\varphi_1, \varphi_2, \ldots, \varphi_k$ is a direct proof of ψ from φ, then $\varphi_k, \varphi_{k-1}, \ldots, \varphi_1$ is a direct proof of φ from ψ.

2.4. Soundness and Completeness. This completes our project of formalizing the central notions of logic. We are now in a position to express the questions stated at the beginning of this section in purely mathematical terms. These questions were:

(1) Is a given method of reasoning such as to only produce valid arguments or, by employing this method, will one sometimes be led to offer an invalid argument in support of a claim?

(2) If an argument is valid, will one be able to judge that it is valid by employing a given method of reasoning, or will there be some valid arguments which, in principle, cannot be produced by this method?

Now that we have made precise the notions of argument, validity, and method of reasoning, we can interpret these questions in strictly mathematical terms. When we ask, for example, "does a given method of reasoning ever produce invalid arguments?", interpreted formally, we are asking of a given deductive system whether it is possible for a sentence ψ to be provable from Γ, despite the fact that ψ is not a logical consequence of Γ. We refer to the property of the deductive system which denies this possibility as soundness:

Definition 1.11. A deductive system T is *sound* if, for all sets of sentences Γ and all sentences ψ:

$$\text{If } \Gamma \vdash_T \psi, \text{ then } \Gamma \models \psi.$$

Conversely, when we ask "do there exist valid arguments which cannot be produced by a given method of reasoning?" we are asking of a given deductive system whether it is possible for a sentence ψ to be a logical consequence of Γ, despite the fact that ψ is not provable from Γ. The property which denies this possibility is referred to as completeness:

Definition 1.12. A deductive system T is *complete* if, for all sets of sentences Γ and all sentences ψ:

$$\text{If } \Gamma \models \psi, \text{ then } \Gamma \vdash_T \psi.$$

In Part I of the course our aim will be to construct deductive systems for sentential and first-order logic that are both sound and complete.[14] The fact that such deductive systems exist (especially in the case of first-order logic) represents

[14] As we shall see, it is, in general, much easier to prove the soundness of a deductive system, than it is to prove its completeness. The reason for this is straightforward. Soundness asserts that, for any given set of sentences Γ, a certain property holds of all the deductive consequences of Γ (*viz.*, the property of being a logical consequence of Γ). As noted above, the set of deductive consequences of Γ admits of a simple inductive definition, so that the soundness of a deductive system can generally be established by means of a straightforward proof by induction. Completeness, on the other hand, asserts that a certain property holds of the set of all logical consequences of Γ (*viz.*, the property of being a deductive consequence of Γ). However, this latter set, as initially described, does not have an obvious inductive definition. Thus, to prove completeness, we will generally have to employ alternative non-inductive techniques, which, in general, will be less straightforward. (Of course, once we have proven that a given deductive system is both sound and complete, the inductive definition of the set of deductive consequences of Γ is itself an inductive definition of the set of logical consequences Γ, since soundness and completeness amount to the claim that these two sets are identical. Nevertheless, on pain of circularity, we cannot appeal to this definition as part of a proof of this system's completeness).

the first fundamental theorem of mathematical logic.[15]

Exercise 1.8. We say that a deductive system T is *sound with respect to a class of valuations* V if, for all sets of sentences Γ and all sentences ψ:

$$\text{If } \Gamma \vdash_T \psi, \text{ then } \Gamma \models_V \psi,$$

and we say that T is *complete with respect to a class of valuations* V, if, for all sets of sentences Γ and all sentences ψ:

$$\text{If } \Gamma \models_V \psi, \text{ then } \Gamma \vdash_T \psi.$$

Prove the following claims:

(a) If T is sound with respect to V' and $V \subset V'$, then T is sound with respect to V.

(b) If T is complete with respect to V and $V \subset V'$, then T is complete with respect to V'.

(c) If V is the set of all possible valuations of a formal language with more than one sentence, then the only deductive systems that are sound with respect to V are those which contain no axioms and only rules of inference of the form:

$$\frac{\varphi_1, \varphi_2, \ldots, \psi, \ldots, \varphi_{k-1}, \varphi_k}{\psi}$$

(d) If V is the set of all possible valuations of a formal language, then every deductive system is complete with respect to V.

Example 4. In this example, we will prove the soundness and completeness of the deductive system $T_\times$ introduced in Example 3 (see page 15) (Since the only deductive system with which we will be concerned is $T_\times$, we will henceforth omit any explicit reference to this system, writing $\Gamma \vdash \psi$ in place of $\Gamma \vdash_{T_\times} \psi$). We begin with soundness:

Theorem 1.3 (Soundness Theorem for $T_\times$). *If* $\Gamma \vdash \psi$, *then* $\Gamma \models \psi$.

Proof. To prove the soundness of $T_\times$, we appeal to the fact that the set of all deductive consequences of Γ (in $T_\times$) can be defined inductively along the lines of the definition given on page 15. It is thus sufficient to show that:

(i) If $\psi \in \Gamma$, then $\Gamma \models \psi$;

(ii) If ψ is an axiom of $T_\times$, then $\Gamma \models \psi$; and

(iii) If $\Gamma \models \varphi_i$, for $i = 1, 2, \ldots, k$ and if ψ can be inferred from $\varphi_1, \varphi_2, \ldots, \varphi_k$ by some rule of $T_\times$, then $\Gamma \models \psi$.

Claim (i) follows trivially from the definition of validity. To prove claim (ii), we first recall that $T_\times$ has three axioms: $x = x$, $0 = 0$, and $1 = 1$. It is easy to verify that each of these axioms is a logical truth, and since any argument with a logically true conclusion is valid (see Exercise **??**), claim (ii) follows.

All that remains is to prove claim (iii). For this, it is enough to show that if ψ follows from $\varphi_1, \varphi_2, \ldots, \varphi_k$ by some rule of $T_\times$, then $\{\varphi_1, \varphi_2, \ldots, \varphi_k\} \models \psi$ (see Exercise **??**). We must, therefore, consider each of the

[15] The existence of a sound and complete deductive system for first-order logic was first proven by Kurt Gödel in 1929.

rules of $T_\times$ and show that this is true. For instance, with respect to rule (1−), it must be shown that for any numerical expressions α, β:

$$1 \times \alpha = \beta \models \alpha = \beta$$

To see that this is so, suppose that $val^{\sigma_i}(1 \times \alpha = \beta) = \mathbf{T}$. This means that $\sigma_i(1 \times \alpha) = \sigma_i(\beta)$. But since $\sigma_i(1 \times \alpha) = \sigma_i(\alpha)$, it follows that $\sigma_i(\alpha) = \sigma_i(\beta)$, i.e., $val^{\sigma_i}(\alpha = \beta) = \mathbf{T}$. Thus, the above implication holds.

The corresponding arguments for the rest of the rules are equally straightforward:

(1+): Suppose that $val^{\sigma_i}(\alpha = \beta) = \mathbf{T}$. Then:

$$\sigma_i(1 \times \alpha) = \sigma_i(\alpha) = \sigma_i(\beta).$$

Hence, $val^{\sigma_i}(1 \times \alpha = \beta) = \mathbf{T}$.

(0−): Suppose that $val^{\sigma_i}(0 \times \alpha = \beta) = \mathbf{T}$. Then:

$$\sigma_i(0) = \sigma_i(0 \times \alpha) = \sigma_i(\beta).$$

Hence, $val^{\sigma_i}(0 = \beta) = \mathbf{T}$.

(0+): Suppose that $val^{\sigma_i}(0 = \beta) = \mathbf{T}$. Then:

$$\sigma_i(0 \times \alpha) = \sigma_i(0) = \sigma_i(\beta).$$

Hence, $val^{\sigma_i}(0 \times \alpha = \beta) = \mathbf{T}$.

(x−): Suppose that $val^{\sigma_i}(x \times x \times \alpha = \beta) = \mathbf{T}$. Then:

$$\sigma_i(x \times \alpha) = \sigma_i(x \times x \times \alpha) = \sigma_i(\beta).$$

Hence, $val^{\sigma_i}(x \times \alpha = \beta) = \mathbf{T}$.

(x+): Suppose that $val^{\sigma_i}(x \times \alpha = \beta) = \mathbf{T}$. Then:

$$\sigma_i(x \times x \times \alpha) = \sigma_i(x \times \alpha) = \sigma_i(\beta).$$

Hence, $val^{\sigma_i}(x \times x \times \alpha = \beta) = \mathbf{T}$.

For the rules (S), (T) and (≠):

(S): Suppose that $val^{\sigma_i}(\alpha = \beta) = \mathbf{T}$. Then:

$$\sigma_i(\beta) = \sigma_i(\alpha).$$

Hence, $val^{\sigma_i}(\beta = \alpha) = \mathbf{T}$.

(T): Suppose that $val^{\sigma_i}(\alpha = \beta) = \mathbf{T}$ and $val^{\sigma_i}(\beta = \gamma) = \mathbf{T}$. Then:

$$\sigma_i(\alpha) = \sigma_i(\beta) = \sigma_i(\gamma).$$

Hence, $val^{\sigma_i}(\alpha = \gamma) = \mathbf{T}$.

(≠): Since there is no interpretation under which 0 = 1 is true, it follows that, for any sentence φ:

$$0 = 1 \models \varphi.$$

Lastly, for rule (×):

(×): Suppose that $val^{\sigma_i}(n_1 \times n_2 \times \cdots \times n_k = \alpha) = \mathbf{T}$. Then:

$$\sigma_i(n_{x_1} \times n_{x_2} \times \cdots n_{x_k}) = \sigma_i(n_1 \times n_2 \times \cdots n_k) = \sigma_i(\alpha).$$

Hence, $val^{\sigma_i}(n_{x_1} \times n_{x_2} \times \cdots \times n_{x_k} = \alpha) = \mathbf{T}$

This completes the proof of the soundness of $T_\times$. □

Before we proceed to prove the completeness of $T_\times$, we first take note of the following trivial corollary of soundness:

Corollary 1.4. *If there exists a direct proof of ψ from φ, then $\varphi \equiv \psi$.*

This follows from the soundness of $T_\times$, in conjunction with the fact that if there exists a direct proof of ψ from φ, then there exists a direct proof of φ from ψ.

We now turn to the completeness of $T_\times$.

Theorem 1.5 (Completeness Theorem for $T_\times$). *If $\Gamma \models \psi$, then $\Gamma \vdash \psi$.*

Our proof of this theorem will proceed in three distinct stages: (1) First, we will show that if an argument expressed in $\mathscr{L}_\times$ is valid, so too is the argument that results from substituing for each of the premises and conclusion of the argument, its reduced form (see Exercise 1.5, p. 11). (2) Next, we will show that for any valid argument consisting of only sentences in reduced form, there exists a proof of its conclusion from its premises. (3) Lastly, we will show that if there exists a proof of the reduced form of the conclusion of an argument from the reduced forms of its premises, then the conclusion of the argument is provable from its premises. Claims (1) and (3) together show that the completeness of $T_\times$ can be reduced to its completeness on the tiny fragment of the language consisting only of sentences in reduced form. Claim (2) then shows that $T_\times$ is complete, in this restricted sense.

In what follows, we write Γ^R for the set of all reduced forms of sentences in Γ, i.e.:

$$\Gamma^R = \{\varphi^R : \varphi \in \Gamma\}.$$

Our proof of Theorem 1.5 can thus be represented schematically as follows:

$$\begin{array}{ccc} \Gamma \models \psi & & \Gamma \vdash \psi \\ \Downarrow & & \Uparrow \\ \Gamma^R \models \psi^R & \Longrightarrow & \Gamma^R \vdash \psi^R \end{array}$$

Lemma 1.6. *If $\Gamma \models \psi$, then $\Gamma^R \models \psi^R$.*

Proof. Suppose that $\Gamma \models \psi$. Let σ_i be an interpretation such that $val^{\sigma_i}(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma^R$. Since, for every sentence $\varphi' \in \Gamma$, there exists some $\varphi \in \Gamma^R$ such that $\varphi \equiv \varphi'$, it follows that $val^{\sigma_i}(\varphi') = \mathbf{T}$, for all $\varphi' \in \Gamma$. But since $\Gamma \models \psi$, $val^{\sigma_i}(\psi) = \mathbf{T}$, and since $\psi \equiv \psi^R$, $val^{\sigma_i}(\psi^R) = \mathbf{T}$. So, if $val^{\sigma_i}(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma^R$, $val^{\sigma_i}(\psi^R) = \mathbf{T}$, i.e., $\Gamma^R \models \psi^R$. □

Lemma 1.7. *If $\Gamma^R \models \psi^R$, then $\Gamma^R \vdash \psi^R$.*

Proof. If ψ^R is $0 = 0$, then, since $0 = 0$ is an axiom of $T_\times$, $\Gamma^R \vdash \psi^R$. If, on the other hand, $0 = 1$ is in Γ^R, then, from $(\neq)$, $\Gamma^R \vdash \psi^R$. Thus, we may assume without loss of generality that (i) ψ^R is not $0 = 0$; and that (2) $0 = 1$ is not in Γ^R.

Now, suppose that $0 = 0$ is in Γ^R, and let Γ' be the set which results from omitting $0 = 0$ from Γ^R. Since the interpretations under which all of the sentences in Γ^R are true are exactly those in which all of the sentences in Γ' are true, $\Gamma^R \models \psi^R$ iff $\Gamma' \models \psi^R$. Moreover, since $0 = 0$ is an axiom, every sentence that is provable from Γ' is also provable from Γ^R, and vice-versa, and so $\Gamma^R \vdash \psi^R$ iff $\Gamma' \vdash \psi^R$. Thus, for every instance of the claim, there is an equivalent instance in which the sentence $0 = 0$ is not in Γ^R, and so we may assume without loss of generality that this is the case.

All that is left to show then is that the claim holds whenever $\Gamma^R \subset \{x = 0, x = 1\}$ and $\psi^R \in \{x = 0, x = 1, 0 = 1\}$. If $\psi^R \in \Gamma^R$, then $\Gamma^R \vdash \psi^R$, and so the claim holds. Thus, there are only four cases left to consider:

(i) $\Gamma^R = \emptyset$.

(ii) Γ^R is the set $\{x = 0\}$ and ψ^R is $x = 1$.

(iii) Γ^R is the set $\{x = 1\}$ and ψ^R is $x = 0$.

(iv) Γ^R is the set $\{x = 0, x = 1\}$ and ψ^R is $0 = 1$.

In case (i)-(iii), $\Gamma^R \not\models \psi^R$, and so the claim holds. In case (iv), $\{x = 0, x = 1\} \models 0 = 1$. We must therefore show that $\{x = 0, x = 1\} \vdash 0 = 1$. The proof is as follows:

(1) $x = 0$ [Premise]

(2) $0 = x$ [S,1]

(3) $x = 1$ [Premise]

(4) $0 = 1$ [T, 2,3]

□

Lemma 1.8. *For any sentence φ, $\varphi \vdash \varphi^R$.*

Proof. The proof proceeds by induction on the number of numerals in φ. If φ contains only two numerals, then either (i) φ is equal to φ^R; (ii) φ^R can be proven from φ by a single application of rule (S); or (iii) φ is either $1 = 1$ or $x = x$. In cases (i) and (ii), the claim obviously holds. In case (iii), φ^R is $0 = 0$, and so, since $0 = 0$ is an axiom, the claim holds as well.

Now, suppose that the claim holds for all sentences containing fewer than k numerals ($k > 2$), and let φ be a sentence containing k numerals. We must show that the claim holds for φ. To prove this, it is sufficient to show that there exists a direct proof from φ to some sentence ψ which contains fewer than k numerals, for if this is so, then it follows from Corollary 1.4 that $\varphi^R = \psi^R$, and so, by the induction hypothesis, $\psi \vdash \varphi^R$. But since $\varphi \vdash \psi$, it follows that $\varphi \vdash \varphi^R$.

Let φ be the sentence $\alpha = \beta$. Since, by rule (S), there is a direct proof of ψ from $\alpha = \beta$ iff there is a direct proof of ψ from $\beta = \alpha$, we may assume without loss of generality that α contains at least as many numerals as β. Hence, α contains at least two numerals.

If one of the numerals in α is 1, then, for some numerical expression γ, $1 \times \gamma$ is the result of reordering the numerals in α. So,

(1) $\alpha = \beta$

(2) $1 \times \gamma = \beta$ [×,1]

(3) $\gamma = \beta$ [1−, 2]

is a direct proof of $\gamma = \beta$ from $\alpha = \beta$, and $\gamma = \beta$ contains fewer numerals than $\alpha = \beta$.

If one of the numerals in α is 0, then, for some numerical expression γ, $0 \times \gamma$ is the result of reordering the numerals in α. Hence,

(1) $\alpha = \beta$

(2) $0 \times \gamma = \beta$ [×,1]

(3) $0 = \beta$ [0−, 2]

is a direct proof of $0 = \beta$ from $\alpha = \beta$, and $0 = \beta$ contains fewer numerals than $\alpha = \beta$.

Finally if neither 0 nor 1 appear as a numeral in α, then α is a numerical expression of the form $x \times x \times \gamma$, where γ is a (possibly empty) numerical expression. Thus:

(1) $\alpha = \beta$

(2) $x \times \gamma = \beta$ [x–,1]

is a direct proof of $x \times \gamma = \beta$ from $\alpha = \beta$, and $x \times \gamma = \beta$ contains fewer numerals than $\alpha = \beta$. This completes the inductive step of the proof. □

Lemma 1.9. *For any sentence φ, $\varphi^R \vdash \varphi$.*

Since the proof of this lemma is nearly identical to the proof of Lemma 1.8, we leave it as an exercise for the reader.

Lemma 1.10. *If $\Gamma^R \vdash \psi^R$, then $\Gamma \vdash \psi$.*

Proof. Suppose that $\Gamma^R \vdash \psi^R$ and that Γ^R is equal to the set $\{\varphi_1, \ldots, \varphi_k\}$. For each $i = 1, \ldots, k$, let ψ_i be some sentence in Γ such that $\psi_i^R = \varphi_i$. Then, by Lemma 1.8, for each i, $\psi_i \vdash \varphi_i$, and so $\Gamma \vdash \varphi_i$. Thus, it follows from the fact that provability is monotonic (see Exercise 1.6) that $\Gamma \vdash \psi^R$. But, by Lemma 1.9, $\psi^R \vdash \psi$. Hence, $\Gamma \vdash \psi$. □

Theorem 1.5 follows directly from lemmas 1.6, 1.7, and 1.10.

Exercise 1.9. Let T_0 be the deductive system obtained from $T_\times$ by omitting the rule (T).

(a) Show that $\{x = 0, x = 1\} \vdash_{T_0} 0 = 1$ only if there exists a direct proof of $0 = 1$ from either $x = 0$, $x = 1$, or some axiom of T_0.

(b) Conclude from (a) that the deductive system T_0 is incomplete.

Example 5. In this example, we will consider an extended version of the formal language $\mathscr{L}_\times$ (see Example 1, page 6), which we will denote $\mathscr{L}_\neq$. The alphabet for $\mathscr{L}_\neq$ is given by:

$$\{x, 0, 1, \times, =, \neq\}$$

Numerals and numerical expressions are defined exactly as in $\mathscr{L}_\times$. We refer to the symbols $=$ and $\neq$ as *relation symbols*. The formation rule for the language is:

Formation Rule: An expression is a *sentence* of the language iff it is of the form $\alpha * \beta$, where α and β are numerical expressions and $*$ is a relation symbol.

The semantics for the language $\mathscr{L}_\neq$ is identical to that of $\mathscr{L}_\times$ (see Example 2, page 11). In particular, there are only two possible interpretations of the language, σ_0 and σ_1, and the semantical rule for the language is:

Semantical Rule: For $i = 0, 1$, $val^{\sigma_i}(\varphi) = \mathbf{T}$ iff the sentence which results from subtituting $\sigma_i(x)$ for all occurrences of x in φ expresses a true arithmetical claim.

We introduce a deductive system $T_\neq$ for the language $\mathscr{L}_\neq$, which is a modified version of $T_\times$ (see Example 3, page 15). The axioms of $T_\neq$ are the same as the axioms of $T_\times$. $T_\neq$ includes all the rules of $T_\times$, but in all of the rules except (T) and ($\neq$), the rules are extended so as to apply to inequalities as well as equalities. Thus, for example, the rule (1-) in $T_{\neq}$ is now:

$$\frac{1 \times \alpha * \beta}{\alpha * \beta}$$

where α and β are any numerical expressions and $*$ is any relation symbol.

In addition, $T_{\neq}$ has five more rules of inference. The first four are:

$$(0\neq) \quad \frac{0 \neq \alpha}{1 = \alpha} \qquad\qquad (1\neq) \quad \frac{1 \neq \alpha}{0 = \alpha}$$

$$(1{=}) \quad \frac{1 = \alpha}{0 \neq \alpha} \qquad\qquad (0{=}) \quad \frac{0 = \alpha}{1 \neq \alpha}$$

The fifth rule is:

$$(x\neq) \qquad \frac{x \neq x}{\varphi}$$

where φ is any sentence.

Exercise 1.10.

(a) Show that if ψ can be inferred from φ by any one of the rules $(0\neq)$, $(1{=})$, $(1\neq)$, $(0{=})$ or $(x\neq)$, then $\varphi \models \psi$. Convince yourself that $T_{\neq}$ is sound.

(b) Prove that $T_{\neq}$ is complete. (Hint: First prove that any sentence in $\mathscr{L}_{\neq}$ is equivalent to one of the reduced forms of $\mathscr{L}_{\times}$. Then, consider which of the lemmas in the proof of the completeness of $T_{\times}$ (see Example 4, page 18) hold without alteration. Lastly, prove that the remaining lemma(s) hold in the case of $T_{\neq}$.)

CHAPTER 2

Sentential Logic

1. Introduction

Sentential logic[1] is the logic which applies to arguments in which the truth conditions for both the premises and the conclusion can be expressed as a function of the truth or falsity of a given set of elementary (or atomic) sentences. In this chapter, we will introduce a formal language in which to represent such arguments, a semantical framework in which to assess their validity, and a deductive system in which to carry out proofs in the language.

We begin by introducing a formal language based on a relatively large number of syntactic operations. By means of these operations compound sentences can be built up recursively from more elementary ones. After describing the semantical rules for this language, we will prove that it is both semantically unambiguous (in a sense to be made precise below) and sufficiently expressive to allow for the formulation of any claim whose truth supervenes on that of finitely many atomic sentences. We will then proceed to simplify the syntax of our language in a manner which preserves its expressive power. In this simplified setting, we will introduce a deductive system, which we will prove to be both sound and complete.

2. The Syntax and Semantics of Sentential Logic

2.1. The Formal Language $\mathscr{L}_S^+$. In this section we define a formal language for sentential logic, which we refer to as $\mathscr{L}_S^+$. The alphabet for the language consists, first, of an infinite number of symbols

$$\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n, \ldots$$

referred to as *atoms*.

In addition, the alphabet also includes the five symbols:

$$\neg\ ,\ \wedge\ ,\ \vee\ ,\ \rightarrow\ ,\ \leftrightarrow$$

referred to collectively as *sentential connectives*, and, respectively, as the negation operator, the conjunction operator, the disjunction operator, the conditional operator, and the biconditional operator. The latter four connectives (i.e., $\wedge, \vee, \rightarrow$ and $\leftrightarrow$) are referred to as *binary connectives*, whereas $\neg$ is referred to as a *monadic* connective.

The alphabet also includes the left and right parentheses:

$$(\ ,\ )$$

The sentences of the language are generated inductively through the recursive application of the five connectives:

Definition 2.1. The sentences of $\mathscr{L}_S^+$ are defined inductively as follows:

(1) Any expression consisting of a single atom is a sentence. (We refer to these sentences as *atomic sentences*.)

(2) If φ is a sentence, then the expression $(\neg\varphi)$ is a sentence.

(3) If φ and ψ are sentences and $*$ is a binary connective, then the expression $(\varphi * \psi)$ is a sentence. We refer to $*$ as the *main connective* of the sentence $(\varphi * \psi)$.

[1] Sentential logic is also sometimes referred to as ‘Propositional logic’ or ‘Boolean logic’.

Examples of sentences in $\mathscr{L}_S^+$ include:

(i) $\mathbf{P}_7$

(ii) $(\neg \mathbf{P}_2)$

(iii) $(\mathbf{P}_4 \leftrightarrow \mathbf{P}_{21})$

(iv) $(\neg(\neg \mathbf{P}_{14}))$

(v) $((\mathbf{P}_3 \to (\neg \mathbf{P}_3)) \wedge (\mathbf{P}_{21} \vee \mathbf{P}_3))$

(vi) $(\mathbf{P}_{16} \leftrightarrow (\mathbf{P}_{16} \leftrightarrow (\mathbf{P}_{16} \leftrightarrow \mathbf{P}_{16})))$

The sentence $(\neg\varphi)$ is referred to as a *negation*, and, more specifically, as the negation of φ. If φ is either an atomic sentence or the negation of an atomic sentence, we refer to φ as a *literal*.

The sentence $(\varphi \wedge \psi)$ [$(\varphi \vee \psi)$] is referred to as a *conjunction* [*disjunction*], with left conjunct [disjunct] φ and right conjunct [disjunct] ψ. The sentence $(\varphi \to \psi)$ is referred to as a *conditional*, with antecedent φ and consequent ψ. The sentence $(\varphi \leftrightarrow \psi)$ is referred to as a *biconditional*, with left term φ and right term ψ. Conjunctions, disjunctions, conditional and biconditionals are referred to, collectively, as *binary compounds*.

Note that we have not yet shown that the labels negation, conjunction, disjunction, etc., are mutually exclusive. That is, we have not yet shown that a given sentence cannot, for example, be both a conjunction and a conditional. Nor have we shown that a sentence which is a conjunction must have unique left and right conjuncts. We will later confirm that this is indeed the case, but only after we have introduced the semantical rules for the language.

2.2. The Truth Functional Interpretation of the Sentential Connectives. The non-logical elements of the syntax of $\mathscr{L}_S^+$ are the atomic sentences, which are to be interpreted as elementary and logically independent sentences. The atomic sentences are elementary in the sense that, consistent with the meanings of the sentential connectives, any particular atomic sentence can be either true or false, and they are independent, in that the truth or falsity of any one atomic sentence, in combination with the meanings of the sentential connectives, imposes no constraints on the truth or falsity of any other. In short, the interpretations of the language include all possible assignments of truth-values to the atomic sentences.

Definition 2.2. An *interpretation* of the language $\mathscr{L}_S^+$ is a function σ which assigns to each of the atomic sentences a truth-value, i.e., an object from the set $\{\mathbf{T}, \mathbf{F}\}$.

The semantical rules for the language must specify how a given interpretation determines the truth-values for all sentences in the language. For the atomic sentences, the truth-value is given directly by the interpretation, i.e.:

(S1) If φ is an atomic sentence, then $val^\sigma(\varphi) = \sigma(\varphi)$.

To assign truth-values to non-atomic sentences, we must introduce additional semantical rules, specifying the meanings of the sentential connectives. We will do so with the aim of interpreting the sentential connectives as effecting roughly the same semantic operations as the ordinary English locutions to which they correspond in the following table:

Sentential Connective	Operation	English Operator
$\neg$	Negation	It is not the case that ....
$\wedge$	Conjunction	... and −−.
$\vee$	Disjunction	... or −−.
$\to$	Conditional	If ..., then −−.
$\leftrightarrow$	Biconditional	..., if and only if −−.

What defines each of these operations is the specific way in which the truth or falsity of compound sentences resulting from their application is related to the truth-values of the component sentences to which they are applied. For example, negation is an operation which takes a sentence and returns a sentence with the opposite truth-value (i.e., true if the component sentence is false, and false if it is true), conjunction is an operation which takes two sentences and returns a true sentence if and only if both of its component sentences are true, etc..

To formalize the behavior of these semantic operations, we introduce the notion of a truth-function:

DEFINITION 2.3. A *truth-function* is a function which takes truth-values and returns a truth-value.

EXERCISE 2.1.

(a) How many k-ary truth-functions are there?

(b) A k-ary truth-function f is *symmetrical* if:

$$f(\alpha_1, \ldots, \alpha_k) = f(\alpha_{n_1}, \ldots, \alpha_{n_k}),$$

where $n_1, \ldots, n_k \in \{1, \ldots, k\}$ and $n_i \neq n_j$ for $i \neq j$.

A 2-ary truth-function f is *associative* if:

$$f(f(\alpha, \beta), \gamma) = f(\alpha, f(\beta, \gamma))$$

Let g be a 2-ary truth-function that is both symmetrical and associative. Show that, for $k \geq 2$, the k-ary function:

$$g^k(\alpha_1, \ldots, \alpha_k) = g(\cdots g(g(\alpha_1, \alpha_2), \alpha_3), \alpha_4) \cdots), \alpha_k)$$

is symmetrical.

The following five truth-functions describe, respectively, the operation of each of the five sentential connectives:

$$f_\neg(\alpha) = \begin{cases} \mathbf{F} & \text{if } \alpha = \mathbf{T} \\ \mathbf{T} & \text{if } \alpha = \mathbf{F} \end{cases}$$

$$f_\wedge(\alpha, \beta) = \begin{cases} \mathbf{T} & \text{if } \alpha = \mathbf{T} \text{ and } \beta = \mathbf{T} \\ \mathbf{F} & \text{otherwise} \end{cases}$$

$$f_\vee(\alpha, \beta) = \begin{cases} \mathbf{F} & \text{if } \alpha = \mathbf{F} \text{ and } \beta = \mathbf{F} \\ \mathbf{T} & \text{otherwise} \end{cases}$$

$$f_\rightarrow(\alpha, \beta) = \begin{cases} \mathbf{F} & \text{if } \alpha = \mathbf{T} \text{ and } \beta = \mathbf{F} \\ \mathbf{T} & \text{otherwise} \end{cases}$$

$$f_\leftrightarrow(\alpha, \beta) = \begin{cases} \mathbf{T} & \text{if } \alpha = \beta \\ \mathbf{F} & \text{otherwise} \end{cases}$$

EXERCISE 2.2. Show that the following claims hold for any truth-values α, β.

(i) $f_\neg(f_\neg(\alpha)) = \alpha$

(ii) $f_\neg(f_\wedge(\alpha, \beta)) = f_\vee(f_\neg(\alpha), f_\neg(\beta))$

(iii) $f_\neg(f_\vee(\alpha, \beta)) = f_\wedge(f_\neg(\alpha), f_\neg(\beta))$

(iv) $f_\rightarrow(\alpha, \beta) = f_\vee(f_\neg(\alpha), \beta)$

(v) $f_\leftrightarrow(\alpha, \beta) = f_\wedge(f_\rightarrow(\alpha, \beta), f_\rightarrow(\beta, \alpha))$

To say that these truth-functions define the operations of the sentential connectives is to assert that:

(S2) For any sentence φ:

$$val^{\sigma}((\neg\varphi)) = f_{\neg}(val^{\sigma}(\varphi))$$

(S3) For any sentences φ and ψ and any binary connective $*$:

$$val^{\sigma}((\varphi * \psi)) = f_{*}(val^{\sigma}(\varphi), val^{\sigma}(\psi))$$

Rules (S1), (S2) and (S3) express semantical rules for the language $\mathscr{L}_S^+$. We have not, however, yet shown that these rules are mutually consistent, for we have not yet shown that every non-atomic sentence in $\mathscr{L}_S^+$ has a *unique* decomposition in terms of more elementary sentences. If, for example, a sentence φ is both a negation and a conjunction, then it may be that rules (S2) and (S3) will assign different truth-values to φ. In the following section we will prove that the rules (S1), (S2) and (S3) do, in fact, constitute a consistent set of semantical rules for the language by showing that the sentences of $\mathscr{L}_S^+$ are uniquely decomposable.

2.3. Unique Decomposition. In order to show that the rules (S1)-(S3) determine a unique valuation of the sentences of $\mathscr{L}_S^+$, we must show that every non-atomic sentence admits of a unique decomposition into its component parts, i.e., we must prove the following theorem:

THEOREM 2.1 (Unique Decomposition). *For every sentence φ, exactly one of the following holds:*

(1) φ is an atomic sentence.

(2) There exists a unique sentence ψ such that $\varphi = (\neg\psi)$.

(3) There exist unique sentences ψ and ρ and a unique binary connective $*$, such that $\varphi = (\psi * \rho)$.

To prove this theorem, we begin by observing the following trivial syntactic facts:

PROPOSITION 2.1. *No negation is an atomic sentence.*

PROOF. Every negation begins with the symbol (, but every atomic sentence begins with an atom. □

PROPOSITION 2.2. *No binary compound is an atomic sentence.*

PROOF. Every binary compound begins with the symbol (, but every atomic sentence begins with an atom. □

PROPOSITION 2.3. *No negation is a binary compound.*

PROOF. Every negation begins with the symbols (¬, but every binary compound begins with either the symbols ($\mathbf{P}_i$ or ((. □

From these three propositions it follows that, for every sentence φ, exactly one of the following holds:

(1) φ is an atomic sentence.

(2) φ is a negation.

(3) φ is a binary compound.

Moreover, from the obvious fact that

PROPOSITION 2.4. *If $(\neg\varphi) = (\neg\psi)$, then $\varphi = \psi$,*

it follows that if φ is a negation, then there is a unique sentence ψ, such that $\varphi = (\neg\psi)$. All then, that remains to be shown is that if φ is a binary compound, it can be decomposed in a unique way, i.e.:

PROPOSITION 2.5. *If $(\varphi * \psi) = (\varphi' *' \psi')$, then $\varphi = \varphi'$, $* = *'$ and $\psi = \psi'$.*

To prove this proposition, it will be helpful to first introduce the following terminology:

DEFINITION 2.4. An expression φ is a *proper prefix* of ψ (written $\varphi < \psi$) iff there exists an expression φ' such that ψ is the expression $\varphi\varphi'$.

DEFINITION 2.5. Let w be the following function assigning numbers to the symbols in the alphabet of $\mathscr{L}_S^+$:

$$w(x) = \begin{cases} 1 & \text{if } x \text{ is the symbol (} \\ -1 & \text{if } x \text{ is the symbol)} \\ 0 & \text{otherwise} \end{cases}$$

let φ be the expression $x_1x_2\ldots x_k$. The *weight* of φ is the quantity:

$$W(\varphi) = \sum_{i=1}^{k} w(x_i)$$

An expression φ is *well-balanced* if the following two conditions hold:

(1) $W(\varphi) = 0$

(2) For all $\psi < \varphi$, $W(\psi) \neq 0$.

LEMMA 2.2. *Every sentence of $\mathscr{L}_S^+$ is a well-balanced expression.*

PROOF. The proof proceeds by induction on the structure of the sentences. If φ is an atomic sentence, then since $w(\mathbf{P}_i) = 0$ $(i = 1, 2, \ldots)$, it follows that $W(\varphi) = 0$, and since φ has no proper prefixes, condition (2) is trivially satisfied.

Let φ be a well-balanced sentence. Then:

$$W((\neg\varphi)) = w(() + w(\neg) + W(\varphi) + w()) = 1 + 0 + 0 + (-1) = 0$$

Now, suppose, for contradiction, that there exists a $\psi < (\neg\varphi)$, with $W(\psi) = 0$. Since $W(() = W((\neg) = W((\neg\varphi) = 1$ there must exist some $\psi' < \varphi$, such that $W(\psi') = -1$. Thus, φ is not an atomic sentence since atomic sentences have no proper prefixes. It follows then that there exists a ψ'', namely, (, such that $\psi'' < \psi'$, $W(\psi'') = 1$, and $W(\psi') = -1$. But if this is the case, then there exists a ρ, such that $\psi'' < \rho < \psi'$ and $W(\rho) = 0$. But since $\rho < \varphi$, this contradicts our assumption that φ is well-balanced. Hence, if φ is a well-balanced sentence, then $(\neg\varphi)$ is well-balanced.

Let φ and ψ be well-balanced sentences, and let $*$ be a binary connective. Then:

$$W((\varphi * \psi)) = w(() + W(\varphi) + w(*) + W(\psi) + w()) = 1 + 0 + 0 + 0 + (-1) = 0$$

Now, suppose, for contradiction, that there exists a $\psi' < (\varphi * \psi)$, with $W(\psi') = 0$. Since $W(() = W((\varphi) = W((\varphi*) = W((\varphi * \psi) = 1$, there must either exist a ψ' such that either $\psi' < \varphi$ or $\psi' < \psi$ and $W(\psi) = 0$. In either case, we get a contradiction following the same reasoning given above. Hence, if φ and ψ are well-balanced sentences, then $(\varphi * \psi)$ is well-balanced. □

COROLLARY 2.3. *No sentence of $\mathscr{L}_S^+$ is a proper prefix of another.*

The proof of proposition 2.5 can now proceed as follows: suppose that $(\varphi * \psi) = (\varphi' *' \psi')$. If $\varphi \neq \varphi'$, then, since $\varphi * \psi) = \varphi' *' \psi')$, either $\varphi < \varphi'$ or $\varphi' < \varphi$, either of which contradicts 2.3. Hence, $\varphi = \varphi'$. But, then $*\psi) = *'\psi')$, and so $* = *'$. From this, however, it follows that $\psi) = \psi')$ and so $\psi = \psi'$, which completes the proof.

By appeal to Theorem 2.1, we can now show that conditions (S1)-(S3) do indeed comprise a consistent set of semantical rules.

THEOREM 2.4. *There is one and only one valuation val^σ satisfying conditions (S1), (S2) and (S3)*

PROOF. We will first prove that there is at most one valuation satisfying (S1)-(S3). For each sentence φ, let $\kappa(\varphi)$ be the number of connectives appearing in φ. Suppose, for contradiction, that $val_1^\sigma \neq val_2^\sigma$, but both are valuations satisfying (S1)-(S3). Then,

$$\kappa^* = \min\{\kappa(\varphi) : val_1^\sigma(\varphi) \neq val_2^\sigma(\varphi)\} \geq 0.$$

If $\kappa^* = 0$, then for some atomic sentence φ, $val_1^\sigma(\varphi) \neq val_2^\sigma(\varphi)$. But, by (S1), $val_1^\sigma(\varphi) = val_2^\sigma(\varphi) = \sigma(\varphi)$. Contradiction.

If $\kappa^* > 0$, then there exists a sentence φ such that $\kappa(\varphi) = \kappa^*$ and $val_1^\sigma(\varphi) \neq val_2^\sigma(\varphi)$. But, since $\kappa(\varphi) > 0$, either (i) $\varphi = (\neg\psi)$; or (ii) $\varphi = (\psi * \rho)$, for some binary connective $*$. In case (i), it follows from (S2) that $f_\neg(val_1^\sigma(\psi)) \neq f_\neg(val_2^\sigma(\psi))$, and so $val_1^\sigma(\psi) \neq val_2^\sigma(\psi)$. But since $\kappa(\psi) < \kappa^*$, this contradicts the definition of κ^*. In case (ii), it follows from (S3) that $f_*(val_1^\sigma(\psi), val_1^\sigma(\rho)) \neq f_*(val_2^\sigma(\psi), val_2^\sigma(\rho))$, and so either $val_1^\sigma(\psi) \neq val_2^\sigma(\psi)$, or $val_1^\sigma(\rho) \neq val_2^\sigma(\rho)$. But since both $\kappa(\psi) < \kappa^*$ and $\kappa(\rho) < \kappa^*$, this contradicts the definition of κ^*.

We will prove that there exists one valuation satisfying conditions (S1)-(S3) by induction on κ, i.e., we will first show that there exists a partial valuation consistent with (S1)-(S3) that assigns truth-values to all the atomic sentences, and we will then show that if there is a valuation consistent with (S1)-(S3) that assigns truth-values to all sentences φ, with $\kappa(\varphi) < n$, this valuation can be extended in a consistent way to apply to all sentences φ, with $\kappa(\varphi) = n$. Since, by theorem 2.1, no atomic sentence is a negation or binary compound, (S2) and (S3) put no constraints the truth-values assigned to atomic sentences. Thus, putting $val^\sigma(\varphi) = \sigma(\varphi)$, for every atomic φ, we obtain a partial valuation of the atomic sentences, consistent with (S1)-(S3). Now, suppose that we have defined a partial valuation consistent with (S1)-(S3) that applies to all sentences with $\kappa < n$ ($n \geq 1$) and let φ be a sentence such that $\kappa(\varphi) = n$. By theorem 2.1, φ is either a negation or a binary compound, but not both. In the first case, by theorem 2.1, (S1) and (S3) put no constraints on the truth-value assigned to φ. Moreover, since there is a unique ψ such that $\varphi = (\neg(\psi))$, it is consistent with (S2) to put $val^\sigma(\varphi) = f_\neg(val^\sigma(\psi))$. If φ is a binary compound, then by theorem 2.1, (S1) and (S2) put no constraints on the truth-value assigned to φ, and since there is a unique ψ, ρ, $*$, such that $\varphi = (\psi * \rho)$, it is consistent with (S3) to put $val^\sigma(\varphi) = f_*(val^\sigma(\psi), val^\sigma(\rho))$. This completes the proof. □

When writing sentences of $\mathscr{L}_S^+$, we will henceforth omit the outermost parentheses, as well as any parentheses pairs enclosing a negation. Thus, for example, instead of $((\mathbf{P}_1 \wedge \mathbf{P}_2) \rightarrow (\neg(\neg\mathbf{P}_3)))$, we will write $(\mathbf{P}_1 \wedge \mathbf{P}_2) \rightarrow \neg\neg\mathbf{P}_3$.

Exercise 2.3. Suppose that we omit the parentheses from our language and instead define the sentences inductively as follows:

(1) Any expression consisting of a single atom is a sentence.

(2) If φ is a sentence, then the expression $\neg\varphi$ is a sentence.

(3) If φ and ψ are sentences and $*$ is a binary connective, then the expression $\varphi * \psi$ is a sentence.

Show that the three conditions:

(S1)* If φ is an atomic sentence, then $val^\sigma(\varphi) = \sigma(\varphi)$.

(S2)* For any sentence φ:

$$val^\sigma(\neg\varphi) = f_\neg(val^\sigma(\varphi))$$

(S3)* For any sentences φ and ψ and any binary connective $*$:

$$val^\sigma(\varphi * \psi) = f_*(val^\sigma(\varphi), val^\sigma(\psi))$$

are mutually inconsistent, i.e., show that they entail that, for some interpretation, the same sentence is assigned both **T** and **F**.

Exercise 2.4. If we wish to omit the parentheses from our language in a manner which preserves unique decomposability, we can adopt a prefix (rather than the standard infix) notation for expressing sentential compounds.[2] Suppose we define the sentences of the language inductively as follows:

(1) Any expression consisting of a single atom is an (atomic) sentence.

(2) If φ is a sentence then $\neg\varphi$ is a sentence.

(3) If φ and ψ are sentences and $*$ is a binary connective, then the expression $*\varphi\psi$ is a sentence.

(a) Show that if $*\varphi\psi = *'\varphi'\psi'$, then $* = *'$, $\varphi = \varphi'$ and $\psi = \psi'$.

(b) Convince yourself that the sentences satisfy the unique decomposition theorem, i.e., for each sentence φ exactly one of the following holds:

(1) φ is an atomic sentence.

(2) There exists a unique sentence ψ, such that $\varphi = \neg\psi$.

(3) There exists unique sentences ψ and ρ, and a unique binary connective $*$, such that $\varphi = *\psi\rho$.

Example 6. The issue of unique decomposition naturally arises in the context of questions related to the retrieval of encoded information. Suppose for example that we have an information source, which outputs messages consisting of finite strings of symbols from the alphabet,

$$\{A, B, C\}.$$

We would like to send these messages to be processed by a digital computer, but in order to do so, each message must first be encoded into a binary string, i.e., a sequence of 0's and 1's.

Definition 2.6. An *encoding function* (or simply an encoding) is a function which assigns to each of the symbols A, B, C, a fixed binary string. If μ is an encoding, and $X_1X_2\ldots X_n$ a sequence of symbols from the set $\{A, B, C\}$, then the *μ-code* of this sequence is the binary string:

$$\mu(X_1)\mu(X_2)\ldots\mu(X_n)$$

The set of all μ-codes can be given the following inductive definition:

(1) The strings $\mu(A)$, $\mu(B)$ and $\mu(C)$ are all μ-codes.

(2) If φ is a μ-code and $X \in \{A, B, C\}$, then $\varphi\mu(X)$ is a μ-code.

Ideally, we would like our encoding to be such that no information is lost in the coding process. In other words, knowledge of the encoding μ, should allow us to associate with each μ-code the exact message which it encodes.

To decode an encoded message, first we identify a sequence of the form $\mu(A)$, $\mu(B)$ or $\mu(C)$, at the end of the encoded message. We then decode the rest of the message, appending A, B or C to the end of the decoded message, depending on which of the above three sequences was identified. To show that this process yields a unique result, we must show that there exists a *decoding function*, g, satisfying the following properties:

(1) $g(\mu(X)) = X$, where $X \in \{A, B, C\}$

(2) $g(\varphi\mu(X)) = g(\varphi)g(\mu(X))$.

If there exists such a decoding function, we say that μ is a *lossless* code.

From Theorem **??** it suffices to prove that μ is lossless to show that the inductive definition of μ-codes given above possesses the unique decomposition property.

[2]Such a prefix notation is sometimes referred to as Polish notation.

EXERCISE 2.5.

(a) Let μ be the encoding function defined by:

$$\mu(X) = \begin{cases} 0 & \text{if } X = A. \\ 10 & \text{if } X = B. \\ 01 & \text{if } X = C. \end{cases}$$

Show that μ is *not* a lossless code.

(b) Provide an example of a lossless code and prove that it is lossless.

2.4. Complete Sets of Connectives. The semantical rules for the language $\mathscr{L}_S^+$ are:

(S1) If φ is an atomic sentence, then $val^\sigma(\varphi) = \sigma(\varphi)$.

(S2) For any sentence φ:

$$val^\sigma(\neg\varphi) = f_\neg(val^\sigma(\varphi))$$

(S3) For any sentences φ and ψ and any binary connective $*$:

$$val^\sigma(\varphi * \psi) = f_*(val^\sigma(\varphi), val^\sigma(\psi))$$

All of the standard semantical notions (e.g., logical implication, logical truth, logical equivalence) are defined just as before (see Chapter 1, §2.2).

DEFINITION 2.7. Let φ be any sentence containing only atoms in the sequence $\mathbf{P}_1, \ldots, \mathbf{P}_k$. The *k-ary truth-function for* φ is the k-ary truth-function $f_{\varphi,k}$ defined by:

$$f_{\varphi,k}(\alpha_1, \ldots, \alpha_k) = val^\sigma(\varphi),$$

for any interpretation σ, such that:

$$\sigma(\mathbf{P}_i) = \alpha_i,$$

for $i = 1, \ldots, k$.

A particularly convenient way to represent the k-ary truth function for φ is in terms of a *truth-table*. Excluding column headers, the first k columns of each of the 2^k rows of the table specify a distinct possible assignment of truth-values to the atoms $\mathbf{P}_1, \ldots, \mathbf{P}_k$, and the $(k+1)$-th column gives the value of the function $f_{\varphi,k}$ for this particular assignment:

$\mathbf{P}_1$	$\mathbf{P}_2$	$\cdots$	$\mathbf{P}_k$	φ
T	**T**	$\cdots$	**T**	$f_{\varphi,k}(\mathbf{T}, \mathbf{T}, \ldots, \mathbf{T})$
T	**T**	$\cdots$	**F**	$f_{\varphi,k}(\mathbf{T}, \mathbf{T}, \ldots, \mathbf{F})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
F	**F**	$\cdots$	**F**	$f_{\varphi,k}(\mathbf{F}, \mathbf{F}, \ldots, \mathbf{F})$

If φ and ψ are any two sentences containing only atoms in the sequence $\mathbf{P}_1, \ldots, \mathbf{P}_k$, then it is easily seen that $\varphi \equiv \psi$ if and only if $f_{\varphi,k} = f_{\psi,k}$. Hence, we can establish that two sentences are equivalent by showing that they have the same k-ary truth-table.

When writing the truth-table for φ, it is often convenient to omit columns corresponding to atoms not appearing in φ since the truth-value of these atoms has no effect on the truth-value of φ. In addition, one often includes additional columns listing the truth-values of certain component sentences of φ (indicated in the column header) which may aid in the computation of φ's truth-value. Thus, for example, a truth-table for the sentence $\neg(\mathbf{P}_3 \wedge \mathbf{P}_1) \to (\mathbf{P}_1 \vee \mathbf{P}_2)$ may be written:

$\mathbf{P}_1$	$\mathbf{P}_2$	$\mathbf{P}_3$	$\mathbf{P}_3 \wedge \mathbf{P}_1$	$\neg(\mathbf{P}_3 \wedge \mathbf{P}_1)$	$\mathbf{P}_1 \vee \mathbf{P}_2$	$\neg(\mathbf{P}_3 \wedge \mathbf{P}_1) \rightarrow (\mathbf{P}_1 \vee \mathbf{P}_2)$
T	**T**	**T**	**T**	**F**	**T**	**T**
T	**T**	**F**	**F**	**T**	**T**	**T**
T	**F**	**T**	**T**	**F**	**T**	**T**
T	**F**	**F**	**F**	**T**	**T**	**T**
F	**T**	**T**	**F**	**T**	**T**	**T**
F	**T**	**F**	**F**	**T**	**T**	**T**
F	**F**	**T**	**F**	**T**	**F**	**F**
F	**F**	**F**	**F**	**T**	**F**	**F**

We may also use truth-tables to indicate how the truth of a given compound sentence depends on its (not necessarily atomic) components. For example, if φ and ψ are any two sentences, we may write the truth-table for the sentence $\neg\varphi \rightarrow \psi$ as follows:

φ	ψ	$\neg\varphi$	$\neg\varphi \rightarrow \psi$
T	**T**	**F**	**T**
T	**F**	**F**	**T**
F	**T**	**T**	**T**
F	**F**	**T**	**F**

In a schematic truth-table like this, we countenance all possible truth-value assignments to the component sentences φ and ψ. In fact, depending on the sentences φ and ψ, certain of these assignments may not be possible. Nevertheless, if we wish to establish for two sentence schemas that *all* possible instantiations of these schemas are equivalent, we must consider all possible valuations of φ and ψ, since the equivalency is meant to hold even when φ and ψ are both atomic.

Using truth-tables, we may establish the following standard equivalencies, which we will take for granted in the sequel. For all sentences φ, ψ, ρ:

(i) $\varphi \wedge \varphi \equiv \varphi$

(ii) $\varphi \vee \varphi \equiv \varphi$

(iii) $\varphi \wedge \psi \equiv \psi \wedge \varphi$

(iv) $\varphi \vee \psi \equiv \psi \vee \varphi$

(v) $(\varphi \wedge \psi) \wedge \rho \equiv \varphi \wedge (\psi \wedge \rho)$

(vi) $(\varphi \vee \psi) \vee \rho \equiv \varphi \vee (\psi \vee \rho)$

(vii) $\varphi \wedge (\psi \vee \rho) \equiv (\varphi \wedge \psi) \vee (\varphi \wedge \rho)$

(viii) $\varphi \vee (\psi \wedge \rho) \equiv (\varphi \vee \psi) \wedge (\varphi \vee \rho)$

(ix) $\neg(\varphi \wedge \psi) \equiv \neg\varphi \vee \neg\psi$

(x) $\neg(\varphi \vee \psi) \equiv \neg\varphi \wedge \neg\psi$

(xi) $(\varphi \vee \neg\varphi) \wedge \psi \equiv \psi$

(xii) $(\varphi \wedge \neg\varphi) \vee \psi \equiv \psi$

(xiii) $\neg\neg\varphi \equiv \varphi$.

(xiv) $\varphi \rightarrow \psi \equiv \neg\psi \rightarrow \neg\varphi$

(xv) $\varphi \rightarrow (\psi \rightarrow \rho) \equiv (\varphi \wedge \psi) \rightarrow \rho$

(xvi) $\varphi \leftrightarrow \psi \equiv (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$

In contexts in which we are only concerned with the truth-values of sentences, we allow for the use of certain ambiguous expressions provided all possible disambiguations of these expressions lead to logically equivalent sentences. In particular, we will often omit the parentheses from nested conjunctions, disjunctions and biconditionals, since the parentheses in this case only serve to indicate the order in which the terms are to be combined, and, regardless of this order, the truth function for the resulting sentence is the same (see Exercise 2.1). Thus, for example, we write:

$$\varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_k$$

to express some nested conjunction of the sentences $\varphi_1, \ldots, \varphi_k$, despite the fact that it is underdetermined which specific sentence this expression denotes.

DEFINITION 2.8. Let C be a set of sentential connectives, i.e., $C \subset \{\neg, \wedge, \vee, \rightarrow, \leftrightarrow\}$. The set Γ_C is the set of sentences of $\mathscr{L}_S^+$ defined inductively as follows:

(1) If φ is an atomic sentence, then $\varphi \in \Gamma_C$.

(2) If $\varphi \in \Gamma_C$ and $\neg \in C$, then $(\neg\varphi) \in \Gamma_C$.

(3) If $\varphi, \psi \in \Gamma_C$ and $* \in C$, where $*$ is any binary connective, then $(\varphi * \psi) \in \Gamma_C$.

We say that C is a *truth-functionally complete* set of connectives if every k-ary truth function is the k-ary truth function for some sentence $\varphi \in \Gamma_C$.

EXERCISE 2.6. For any truth-value α, let:

$$\neg\alpha = \begin{cases} \mathbf{T} & \text{if } \alpha = \mathbf{F} \\ \mathbf{F} & \text{if } \alpha = \mathbf{T} \end{cases}$$

If f is a k-ary truth function, we refer to the function:

$$f^d(\alpha_1, \ldots, \alpha_k) = \neg(f(\neg\alpha_1, \ldots, \neg\alpha_k))$$

as the *dual* of f. With each sentence $\varphi \in \Gamma_{\{\neg,\wedge,\vee\}}$ we associate a sentence φ^d (referred to as the *dual* of φ) as follows:

(1) If φ is an atomic sentence, then $\varphi^d = \varphi$.

(2) $(\neg\varphi)^d = (\neg\varphi^d)$.

(3) $(\varphi \wedge \psi)^d = (\varphi^d \vee \psi^d)$

(4) $(\varphi \vee \psi)^d = (\varphi^* \wedge \psi^d)$

(a). Show that if f is the k-ary truth function for some $\varphi \in \Gamma_{\{\neg,\wedge,\vee\}}$, then f^d is the k-ary truth function for φ^d. Convince yourself that for any $\varphi, \psi \in \Gamma_{\{\neg,\wedge,\vee\}}$:

$$\varphi \equiv \psi \Leftrightarrow \varphi^d \equiv \psi^d$$

(b).(*) A sentence $\varphi \in \Gamma_{\{\neg,\wedge,\vee\}}$ is *self-dual* if $\varphi \equiv \varphi^d$. Provide an example of a non-atomic, self-dual sentence belonging to the set $\Gamma_{\{\wedge,\vee\}}$.

DEFINITION 2.9. Let $\{f_1, f_2, \ldots\}$ be a set of truth-functions. We say that the truth-function g *is definable in terms of* $\{f_1, f_2, \ldots\}$ if it belongs to the set F defined inductively as follows:

(1) $f_i \in F$, for $i = 1, 2, \ldots$.

(2) If $g_1, \ldots, g_k$ are j-ary functions in F and h is a k-ary function in F, then the j-ary function:

$$h(g_1, \ldots, g_k)$$

is in F.

It can easily be verified that definability is monotonic in the following sense: If f is definable in terms of G and, for every $g \in G$, g is definable in terms of H, then f is definable in terms of H.[3]

[3]In some texts, the following alternative definition of definability is adopted: the truth-function g is definable in terms of $\{f_1, f_2, \ldots\}$ if g has a definition:

$$g(\alpha_1, \ldots, \alpha_k) = \cdots,$$

Definition 2.10. Let C be a set of sentential connectives. A *C-function* is a truth-function which is definable in terms of the set:

$$\{f_* : * \in C\} \cup \{f_{\mathbf{P}_j,k} : k = 1, 2, \ldots; j \leq k\}$$

Exercise 2.7.

(a) Show that every function which is definable in terms of a set of symmetrical functions (see Exercise 2.1) is itself symmetrical.

(b) Prove that $f_{\mathbf{P}_1,2}$ is *not* definable in terms of $\{f_\wedge, f_\vee, f_\leftrightarrow\}$.

If f is a k-ary function, we say that f is j-independent ($1 \leq j \leq k$) if:

$$f(\alpha_1, \ldots, \alpha_j, \ldots, \alpha_k) = f(\alpha_1, \ldots, \neg\alpha_j, \ldots, \alpha_k),$$

for all $\alpha_1, \ldots, \alpha_k$. We say that f is a *real* k-ary function if, for $j = 1, \ldots, k$, f is *not* j-independent.

(c) Show that every function which is definable in terms of a set of real functions is real.

(d) Prove that $f_{\mathbf{P}_1,2}$ is *not* definable in terms of $\{f_\neg, f_\wedge, f_\vee, f_\rightarrow, f_\leftrightarrow\}$.

In the remainder of this section, we will examine which sets of sentential connectives are truth-functionally complete.

Theorem 2.5. *Let C be a set of sentential connectives. g is a C-function iff $g = f_{\varphi,k}$, for some sentence $\varphi \in \Gamma_C$.*

Proof. For the only-if part of the claim, the proof proceeds by induction on g. If $g = f_{\mathbf{P}_j,k}$, the claim obviously holds. Moreover, since

$$f_\neg = f_{(\neg\mathbf{P}_1),1}$$

and

$$f_* = f_{(\mathbf{P}_1 * \mathbf{P}_2),2}$$

for any binary connective $*$, the claim holds in the base case.

The inductive step of the proof follows from the fact that for any sentences $\varphi_1, \ldots, \varphi_k$ and any $j \leq k$:

$$f_{\mathbf{P}_j,k}(f_{\varphi_1,k}, \ldots, f_{\varphi_k,k}) = f_{\varphi_j,k}$$

$$f_\neg(f_{\varphi_1,k}) = f_{(\neg\varphi_1),k} \tag{2.1}$$

$$f_*(f_{\varphi_1,k}, f_{\varphi_2,k}) = f_{(\varphi_1 * \varphi_2),k}. \tag{2.2}$$

For the if-part of the claim, the proof proceeds by induction on φ. If φ is the atomic sentence $\mathbf{P}_j$ ($j \leq k$), then $g = f_{\mathbf{P}_j,k}$, and hence a C-function. The inductive step follows from (2.1) and (2.2). □

Theorem 2.6. *The set of connectives $\{\neg, \wedge, \vee\}$ is truth-functionally complete.*

Proof. Given Theorem 2.5, it will suffice to show that every k-ary truth function is a $\{\neg, \wedge, \vee\}$-function. We begin by noting that the k-ary truth functions:

$$h_\vee(\alpha_1, \ldots, \alpha_k) = \begin{cases} \mathbf{T} & \text{if } \alpha_i = \mathbf{T} \text{ for some } i = 1, \ldots, k \\ \mathbf{F} & \text{otherwise} \end{cases}$$

---

where the right-hand side is some expression involving parentheses, commas, the f_i's, and $\alpha_1, \ldots, \alpha_k$. Obviously, every definable function, in our sense, is definable in this sense, but, under this definition, the projection functions:

$$g(\alpha_1, \ldots, \alpha_k) = \alpha_i,$$

for $i = 1, \ldots, k$ are definable from any set, whereas, under our definition, this is not the case (see Exercise 2.7, d).

and

$$h_{\wedge}(\alpha_1, \ldots, \alpha_k) = \begin{cases} \mathbf{T} & \text{if } \alpha_i = \mathbf{T} \text{ for all } i = 1, \ldots, k \\ \mathbf{F} & \text{otherwise} \end{cases}$$

are both $\{\neg, \wedge, \vee\}$-functions since:

$$h_{\vee} = f_{\vee}(\ldots(f_{\vee}(f_{\vee}(f_{\mathbf{P}_1,k}, f_{\mathbf{P}_2,k}), f_{\mathbf{P}_3,k}), \ldots, f_{\mathbf{P}_k,k})$$

and

$$h_{\wedge} = f_{\wedge}(\ldots(f_{\wedge}(f_{\wedge}(f_{\mathbf{P}_1,k}, f_{\mathbf{P}_2,k}), f_{\mathbf{P}_3,k}), \ldots, f_{\mathbf{P}_k,k})$$

Let ω be the sequence of truth-values $\omega_1, \ldots, \omega_k$. We write h^k_ω for the k-ary truth function defined by:

$$h^k_\omega(\alpha_1, \ldots, \alpha_k) = \begin{cases} \mathbf{T} & \text{if } \alpha_i = \omega_i \text{ for } i = 1, \ldots, k \\ \mathbf{F} & \text{otherwise} \end{cases}$$

We will show that h^k_ω is a $\{\neg, \wedge, \vee\}$-function. For $n = 1, \ldots, k$, let $g^n_{\mathbf{T}} = f_{\mathbf{P}_n,k}$ and let $g^n_{\mathbf{F}} = f_{\neg}(f_{\mathbf{P}_n,k})$. Then, $g^n_\beta(\alpha_1, \ldots, \alpha_k)$ is the k-ary truth-function which returns $\mathbf{T}$ iff $\alpha_n = \beta$. Hence:

$$h^k_\omega = h_{\wedge}(g^1_{\omega_1}, \ldots, g^k_{\omega_k}).$$

Now, let f be any k-ary truth-function and let F be the set of all k-ary sequences of truth-values which return $\mathbf{T}$ when given as arguments to f. Then, f returns $\mathbf{T}$ iff h^k_ω returns $\mathbf{T}$ for some $\omega \in F$. Hence, if $F = \{\omega_1, \ldots, \omega_n\}$, then

$$f = h_{\vee}(h^k_{\omega_1}, \ldots, h^k_{\omega_n}).$$

Thus f is a $\{\neg, \wedge, \vee\}$-function. □

Corollary 2.7. *Every sentence of $\mathscr{L}^+_S$ is logically equivalent to some sentence in* $\Gamma_{\{\neg,\wedge,\vee\}}$.

Note that the proof of Theorem 2.6, in fact, establishes the slightly stronger claim that every k-ary truth-function is expressible in the form:

$$h_{\vee}(h_{\wedge}(g^1_{1,1}, \ldots, g^k_{1,k}), \ldots, h_{\wedge}(g^1_{n,1}, \ldots, g^k_{n,k}))$$

where the g's are either of the form $f_{\mathbf{P}_n,k}$ or $f_{\neg}(f_{\mathbf{P}_n,k})$. This implies that every sentence φ of $\mathscr{L}^+_S$ is equivalent to some sentence of the form:

$$(\varphi_{1,1} \wedge \cdots \wedge \varphi_{1,k}) \vee \cdots \vee (\varphi_{n,1} \wedge \cdots \wedge \varphi_{n,k}),$$

where $\varphi_{i,j}$ is a literal. This sentence is referred to as the *disjunctive normal form* of φ.

Exercise 2.8. Show that every sentence φ of $\mathscr{L}^+_S$ is equivalent to some sentence of the form:

$$(\varphi_{1,1} \vee \cdots \vee \varphi_{1,k}) \wedge \cdots \wedge (\varphi_{n,1} \vee \cdots \vee \varphi_{n,k}),$$

where $\varphi_{i,j}$ is a literal (i.e., either an atomic sentence or its negation). This sentence is referred to as the *conjunctive normal form* of φ.

Supposing that C is a truth-functionally complete set of connectives, it suffices to show that the set of connectives C' is truth-functionally complete, to show that every sentence in Γ_C is logically equivalent to some sentence in Γ'_C. Thus, in light of Theorem 2.6, we can prove that the set $\{\neg, \vee\}$ is truth-functionally complete by observing that:

$$\varphi \wedge \psi \equiv \neg(\neg\varphi \vee \psi).$$

Furthermore, from the fact that $\{\neg, \vee\}$ is truth-functionally complete, we can show that $\{\neg, \to\}$ is truth-functionally complete by noting that:

$$\varphi \vee \psi \equiv \neg\varphi \to \psi.$$

EXERCISE 2.9. Let $\uparrow$ be the binary sentential connective, whose corresponding truth-function is given by:

$$f_{\uparrow}(\alpha,\beta) = f_{\neg}(f_{\wedge}(\alpha,\beta))$$

and let $\downarrow$ be the binary sentential connective, whose corresponding truth-function is given by:

$$f_{\downarrow}(\alpha,\beta) = f_{\neg}(f_{\vee}(\alpha,\beta))$$

(a). Show that $\{\uparrow\}$ and $\{\downarrow\}$ are truth-functionally complete.

(b). Let $*$ be any binary sentential connective. Show that if $\{*\}$ is truth-functionally complete, then $\varphi * \psi$ is equivalent to either $\varphi \uparrow \psi$ or $\varphi \downarrow \psi$.

From a semantical point of view, it thus turns out that a number of the sentential connectives in $\mathscr{L}_S^+$ are superfluous. Even in a language with only the two connectives $\neg$ and $\rightarrow$, every sentence of $\mathscr{L}_S^+$ has an equivalent formulation.

3. A Deductive System for Sentential Logic

3.1. The Deductive System T_S. The formal language $\mathscr{L}_S$ is the sub-language of $\mathscr{L}_S^+$ which includes only the two connectives $\neg$ and $\rightarrow$.[4] In this section, we introduce a deductive system T_S for the language $\mathscr{L}_S$, which we will prove to be both sound and complete.

The axioms of T_S will be expressed schematically, i.e., we will specify certain axiom schemas and stipulate that every sentence which is an instance of one of these schemas is an axiom of the system. The three axiom schemas for T_S are:

(SL-1) $\varphi \rightarrow (\psi \rightarrow \varphi)$

(SL-2) $(\varphi \rightarrow (\psi \rightarrow \rho)) \rightarrow ((\varphi \rightarrow \psi) \rightarrow (\varphi \rightarrow \rho))$

(SL-3) $(\neg\varphi \rightarrow \neg\psi) \rightarrow (\psi \rightarrow \varphi)$

The system has only one rule of inference, namely, 'modus ponens.'[5] It asserts that from a conditional and its antecedent, one can infer its consequent:[6]

$$\text{(MP)} \qquad \frac{\varphi\ ,\ \varphi \rightarrow \psi}{\psi}\ .$$

As an example of a proof in T_S, we have:

PROPOSITION 2.6. $\vdash \varphi \rightarrow \varphi$

The proof is as follows:

(1) $\varphi \rightarrow ((\psi \rightarrow \varphi) \rightarrow \varphi)$ [SL-1]

(2) $(\varphi \rightarrow ((\psi \rightarrow \varphi) \rightarrow \varphi)) \rightarrow ((\varphi \rightarrow (\psi \rightarrow \varphi)) \rightarrow (\varphi \rightarrow \varphi))$ [SL-2]

(3) $(\varphi \rightarrow (\psi \rightarrow \varphi)) \rightarrow (\varphi \rightarrow \varphi)$ [MP - 1,2]

(4) $\varphi \rightarrow (\psi \rightarrow \varphi)$ [SL-1]

(5) $\varphi \rightarrow \varphi$ [MP - 4,3]

---

[4]More precisely, $\mathscr{L}_S$ is the language whose sentences are defined inductively precisely as in Definition 2.1, where the set of binary connectives is taken to be the singleton set $\{\rightarrow\}$. Under the latter assumption, the semantical rules for $\mathscr{L}_S$ are simply (S1), (S2) and (S3).

[5]The phrase 'modus ponens' is an abbreviation of the Latin modus ponendo ponens, which translates "the way that affirms by affirming".

[6]In the remainder of this chapter, the symbol $\vdash$ will be used as shorthand for the symbol $\vdash_{T_S}$.

Note that while this line of reasoning conforms to the specifications laid down in the formal definition of a proof, it is far from 'natural'. Indeed, if our aim were to construct a deductive system suitable for modeling actual reasoning, T_S would be a poor choice. But this, of course, is not our aim; our aim is rather to construct a deductive system for sentential logic which is both sound and complete. Nevertheless, in order to show that this is the case, we will have to make certain claims concerning provability in the system, and to this end, the following theorem is extremely useful:

THEOREM 2.8 (Deduction Theorem). $\Gamma \cup \{\varphi\} \vdash \psi$ *if and only if* $\Gamma \vdash \varphi \to \psi$.

PROOF. The if part of the claim is trivial, since if $\Gamma \vdash \varphi \to \psi$, then $\Gamma \cup \{\varphi\} \vdash \psi$ by a single application of MP.

To prove the only-if part of the claim, let $\psi_1, \ldots, \psi_k$ be a proof of ψ from $\Gamma \cup \{\varphi\}$ (i.e., $\psi = \psi_k$). To prove that $\Gamma \vdash \varphi \to \psi_k$ it will suffice to show that for all $i = 1, \ldots, k$:

$$\Gamma \cup \{\varphi \to \psi_1, \ldots \varphi \to \psi_{i-1}\} \vdash \varphi \to \psi_i \tag{3.1}$$

We prove this claim by induction on i. We first note that ψ_i is either (i) a member of Γ, (ii) equal to φ, (iii) an axiom of T_S, or (iv) the result of applying MP to two previous lines in the proof. In cases (i) and (iii), we have $\Gamma \vdash \varphi \to \psi_i$ since:

(1) $\psi_i \to (\varphi \to \psi_i)$ [SL-1]

(2) ψ_i

(3) $\varphi \to \psi_i$ [MP - 1,2]

where the inclusion of ψ_i in the second step of the proof is justified on the grounds of its being either a premise or an axiom, depending on the case under consideration. In case (ii), (3.1) follows from Proposition 2.6. In case (iv), suppose that ψ_i is the result of applying MP to ψ_j and $\psi_j \to \psi_i$. By the induction hypothesis, it follows that:

$$\Gamma \cup \{\varphi \to \psi_1, \ldots \varphi \to \psi_{i-1}\} \vdash \varphi \to \psi_j$$

and

$$\Gamma \cup \{\varphi \to \psi_1, \ldots \varphi \to \psi_{i-1}\} \vdash \varphi \to (\psi_j \to \psi_i)$$

It thus suffices to prove (3.1) to show that:

$$\{\varphi \to \psi_j, \varphi \to (\psi_j \to \psi_i)\} \vdash \varphi \to \psi_i$$

The proof is as follows:

(1) $(\varphi \to (\psi_j \to \psi_i)) \to ((\varphi \to \psi_j) \to (\varphi \to \psi_i))$ [SL-2]

(2) $\varphi \to (\psi_j \to \psi_i)$ [Premise]

(3) $(\varphi \to \psi_j) \to (\varphi \to \psi_i)$ [MP - 1,2]

(4) $\varphi \to \psi_j$ [Premise]

(5) $\varphi \to \psi_i$ [MP - 4,3]

□

Note that the deduction thoerem holds in any deductive system for $\mathscr{L}_S$ which has (SL-1) and (SL-2) as axiom schemas and MP as its only rule of inference.[7] The deduction theorem provides a remarkably powerful aid in the establishment of claims concerning provability in T_S. For example, it follows as an immediate corollary of this theorem that provability is preserved through arbitrary permutations of the antecedent sentences in a nested conditional:

[7]Indeed, the principal reason for including these axiom schemas in our system was to facilitate a proof of this theorem.

Proposition 2.7. *Let $\varphi_1, \ldots, \varphi_k, \psi$ be any sentences of $\mathscr{L}_S$ and let $n_1, \ldots, n_k \in \{1, \ldots, k\}$, where $n_i \neq n_j$, for all $i \neq j$. Then, for any set of sentences Γ:*

$$\Gamma \vdash \varphi_1 \to (\varphi_2 \to (\cdots \to (\varphi_k \to \psi)\cdots)$$

if and only if

$$\Gamma \vdash \varphi_{n_1} \to (\varphi_{n_2} \to (\cdots \to (\varphi_{n_k} \to \psi)\cdots).$$

Another claim which can be easily established by appeal to the deduction theorem is the following:

Theorem 2.9. *For any sentences φ, ψ, $\{\neg\psi, \psi\} \vdash \varphi$.*

Proof. From (SL-1) and the deduction theorem, we have $\neg\psi \vdash \neg\varphi \to \neg\psi$, and from (SL-3) and the deduction theorem, we have $\neg\varphi \to \neg\psi \vdash \psi \to \varphi$. Thus, $\neg\psi \vdash \psi \to \varphi$, and so by one application of MP, $\{\neg\psi, \psi\} \vdash \varphi$. □

In what follows, once it has been shown that a given sentence is a theorem of T_S, we will treat this sentence as a 'derived' axiom of the system, and include it in proofs without further justification.

Exercise 2.10. Use the deduction theorem to prove that the following claims hold for all sentences φ, ψ:

(a). $\varphi \vdash \neg\neg\varphi$

(b). $\neg\neg\varphi \vdash \varphi$

(c). $\varphi \vdash \psi \to \varphi$

(d). $\neg\varphi \vdash \varphi \to \psi$

(e). $\{\varphi, \neg\psi\} \vdash \neg(\varphi \to \psi)$

3.2. The Soundness Theorem for T_S. In this section, we present a proof of the soundness of the deductive system T_S:

Theorem 2.10 (Soundness of T_S). *If $\Gamma \vdash \psi$, then $\Gamma \models \psi$.*

Proof. We prove this claim by induction on the set of sentences which are provable from Γ, i.e., we will show that:

(1) If $\varphi \in \Gamma$, then $\Gamma \models \varphi$.

(2) If φ is an axiom of T_S, then $\Gamma \models \varphi$.

(3) If $\Gamma \models \varphi$ and $\Gamma \models \varphi \to \psi$, then $\Gamma \models \psi$.

Claim (1) follows immediately from the definition of $\models$. For (2), we note that if φ is an axiom of T_S, then φ is a logical truth. This can be verified by considering the truth-tables for each of the axiom schemas of T_S:

φ	ψ	$\psi \to \varphi$	$\varphi \to (\psi \to \varphi)$
T	**T**	**T**	**T**
T	**F**	**T**	**T**
F	**T**	**F**	**T**
F	**F**	**T**	**T**

For (3), it suffices to show that $\{\varphi, \varphi \to \psi\} \models \psi$. Let σ be an interpretation, such that $val^\sigma(\varphi) = val^\sigma(\varphi \to \psi) = \mathbf{T}$. Then

$$f_\to(\mathbf{T}, val^\sigma(\psi)) = \mathbf{T}.$$

Hence, $val^\sigma(\psi) = \mathbf{T}$. This completes the proof. □

φ	ψ	ρ	$\psi \to \rho$	$\varphi \to (\psi \to \rho)$	$\varphi \to \psi$	$\varphi \to \rho$	$(\varphi \to \psi) \to (\varphi \to \rho)$	$(\varphi \to (\psi \to \rho)) \to ((\varphi \to \psi) \to (\varphi \to \rho))$
T	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**
T	**T**	**F**	**F**	**F**	**T**	**F**	**F**	**T**
T	**F**	**T**	**T**	**T**	**F**	**T**	**T**	**T**
T	**F**	**F**	**T**	**T**	**F**	**F**	**T**	**T**
F	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**
F	**T**	**F**	**F**	**T**	**T**	**T**	**T**	**T**
F	**F**	**T**	**T**	**T**	**T**	**T**	**T**	**T**
F	**F**	**F**	**T**	**T**	**T**	**T**	**T**	**T**

φ	ψ	$\neg\varphi \to \neg\psi$	$\psi \to \varphi$	$(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$
T	**T**	**T**	**T**	**T**
T	**F**	**T**	**T**	**T**
F	**T**	**F**	**F**	**T**
F	**F**	**T**	**T**	**T**

The soundness of T_S expresses a fact about what is *not* provable in the system. In particular, it asserts that no claim which does not logically follow from Γ is provable from Γ. In establishing the soundness of T_S, the strategy we employed was to show that the property of truth under a given interpretation is possessed by any sentence provable from any set of sentences all of which possess this property. Since it follows from the definition of validity, that if $\Gamma \not\models \psi$, there is some interpretation such that truth under this interpretation is not preserved in passing from Γ to ψ, it can be inferred that $\Gamma \nvdash \psi$.

As a general rule, we may note that any property possessed by the axioms of a deductive system, which is preserved under its rules of inference, is possessed by every theorem of that system. This fact can be used to show that certain sentences are not provable in a given deductive system. For example, consider the deductive system which has axiom schemas (SL-1) and (SL-2) and MP as its sole rule of inference, and let $f_\neg^*$ be the 1-ary truth-function which always returns false. If we replace the semantical rule (S2), with the "non-standard" semantical rule:

(S2)* $val^\sigma(\neg\varphi) = f_\neg^*(val^\sigma(\varphi))$

it is easily verified that every instance of (SL-1) and (SL-2) has the property of being true under all interpretations of the language, and that this property is preserved under applications of MP. Hence, every theorem of this system must likewise possesses this property. Note, however, that if ψ is **T** and φ is **F** the sentence $(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$ is false, and so provided ψ is a sentence which can be **T** and φ a sentence which can be **F**, this sentence cannot be proven in the system.

Exercise 2.11.

(a). Let T_S^* be the deductive system obtained from T_S by replacing the axiom schema (SL-3) with:

$$(\psi \to \varphi) \to (\neg\varphi \to \neg\psi)$$

Show that not every instance of (SL-3) is a theorem of T_S^*.

(b).(*) Let T_S^i $(i = 1, 2)$ be the deductive system obtained from T_S by omitting the axiom schema (SL-i). Show that not every instance of (SL-i) is a theorem of T_S^i.

3.3. An Alternative Formulation of Completeness. In this section we will provide an alternative, but equivalent formulation of the completeness theorem for sentential logic:

Theorem 2.11 (Completeness of T_S). *If* $\Gamma \models \psi$, *then* $\Gamma \vdash \psi$.

Definition 2.11. Let Γ be any set of sentences of $\mathscr{L}_S$. We say that Γ is *satisfiable* iff there exists an interpretation σ, such that $val^{\sigma}(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$ (in this case, we say that σ *satisfies* Γ). We say that Γ is *unsatisfiable* if it is not satisfiable.

Definition 2.12. Let Γ be any set of sentences of $\mathscr{L}_S$. We say that Γ is *consistent* iff there is no sentence ψ, such that $\Gamma \vdash \psi$, and $\Gamma \vdash \neg\psi$. We say that Γ is *inconsistent* if it is not consistent.

If T_S is complete, then every consistent set of sentences is satisfiable, for suppose that Γ is unsatisfiable. Since there is no interpretation which satisfies Γ, it follows trivially that for every ψ, both $\Gamma \models \psi$ and $\Gamma \models \neg\psi$, and so, if completeness holds, both $\Gamma \vdash \psi$ and $\Gamma \vdash \neg\psi$, i.e., Γ is inconsistent. As we shall presently show, the converse of this claim holds as well, that is, if every consistent set of sentences is satisfiable, then the system T_S is complete. To see this, we first observe the following two facts:

Proposition 2.8. *$\Gamma \not\models \psi$ iff $\Gamma \cup \{\neg\psi\}$ is satisfiable.*

Proof. This follows straightforwardly from the definition of $\models$ and satisfiability. □

Proposition 2.9. *$\Gamma \nvdash \psi$ iff $\Gamma \cup \{\neg\psi\}$ is consistent.*

Proof. The if part of the claim follows from the definition of consistency and the trivial fact that $\Gamma \cup \{\neg\psi\} \vdash \neg\psi$.

To prove the only-if part of the claim suppose that $\Gamma \cup \{\neg\psi\}$ is inconsistent. We must show that $\Gamma \vdash \psi$. Let A be any axiom of T_S. Then, by Theorem 2.9, $\Gamma \cup \{\neg\psi\} \vdash \neg A$, and so, by the deduction theorem, $\Gamma \vdash \neg\psi \rightarrow \neg A$. Hence, by (SL-3) and one application of MP, we have $\Gamma \vdash A \rightarrow \psi$. By appeal to the axiom A and one more application of MP, we have $\Gamma \vdash \psi$. □

Proposition 2.10. *If Γ is a consistent set, then for any sentence ψ, either $\Gamma \cup \{\psi\}$ or $\Gamma \cup \{\neg\psi\}$ is consistent.*

Proof. Suppose that both $\Gamma \cup \{\psi\}$ and $\Gamma \cup \{\neg\psi\}$ are inconsistent. If $\Gamma \cup \{\psi\}$ is inconsistent, then $\Gamma \cup \{\neg\neg\psi\}$ is inconsistent, since $\neg\neg\psi \vdash \psi$ (see Exercise 2.10b). Hence, from Proposition 3.9, $\Gamma \vdash \neg\psi$. But, if $\Gamma \cup \{\neg\psi\}$ is inconsistent, then, again by Proposition 3.9, $\Gamma \vdash \psi$. This, however, contradicts the assumption that Γ is consistent. □

In its contrapositive form, the completeness theorem asserts that if $\Gamma \nvdash \psi$, then $\Gamma \not\models \psi$. Thus, from Propositions 3.8 and 3.9, it is equivalent to the claim: If $\Gamma \cup \{\neg\psi\}$ is consistent, then $\Gamma \cup \{\neg\psi\}$ is satisfiable. Hence, if every consistent set of sentences is satisfiable, the completeness theorem holds.

The completeness theorem may thus be expressed as follows:

Theorem 2.12 (Completeness Theorem for T_S). *Every consistent set of sentences is satisfiable.*

Exercise 2.12. Without appealing to the alternative formulation of completeness given above, show that the soundness theorem for T_S is equivalent to the claim that every satisfiable set of sentences is consistent.

3.4. The Completeness Theorem for T_S. The completeness theorem states that for every consistent set of sentences, there exists an interpretation which satisfies this set. The proof that we will give of this claim is "constructive" in the sense that it will not merely establish the existence of such an interpretation, but it will provide us with an effective method for identifying (or constructing) a particular interpretation of this sort.

This construction will proceed along the following lines: we will consider the sentences of the language, one by one, and determine, for each such sentence, whether or not its truth or falsity is provable by the sentences already in the set. If so, we will add to the set whichever of either the sentence or its negation is consistent with the sentences already in the set; if not, we will simply include the sentence in the set. In this way, we will extend the original set to a 'maximal' set of sentences which can be used to decide, in a consistent manner, the truth of falsity of every sentence in the language.

Such a maximal set amounts to an explicit characterization of a particular valuation of the language, with the members of the set indicating which of the sentences are true under this valuation. Thus, by considering which of the atomic sentences belong to this set, we will be able to arrive at an interpretation of the language that satisfies every sentence in the maximal set. Since the original set is a subset of this maximal set, it too will be satisfied by this interpretation.

Definition 2.13. A set of sentences Γ is *maximally consistent*, if the following two conditions hold:

(1) Γ is consistent.

(2) Every set Γ' of which Γ is a proper subset is inconsistent.

Proposition 2.11. *If Γ is a maximally consistent set of sentences, then for every sentence φ, exactly one of the two sentences φ, $\neg\varphi$ is a member of Γ.*

Proof. That at most one of these two sentences belongs to Γ follows from the fact that Γ is consistent. Now, suppose that $\varphi, \neg\varphi \notin \Gamma$. By the definition of maximal consistency, $\Gamma \cup \{\varphi\}$ is inconsistent and since $\neg\neg\varphi \vdash \varphi$, it follows that $\Gamma \cup \{\neg\neg\varphi\}$ is inconsistent. Hence, by Proposition 3.9, $\Gamma \vdash \neg\varphi$. Similarly, since $\Gamma \cup \{\neg\varphi\}$ is inconsistent, again by Proposition 3.9, $\Gamma \vdash \varphi$. But this contradicts the assumption that Γ is consistent. □

Theorem 2.13. *Every maximally consistent set of sentences is satisfiable.*

Proof. Let Γ be a maximally consistent set of sentences, and let σ_Γ be the interpretation defined by:

$$\sigma_\Gamma(\mathbf{P}_i) = \begin{cases} \mathbf{T} & \text{if } \mathbf{P}_i \in \Gamma \\ \mathbf{F} & \text{if } \neg\mathbf{P}_i \in \Gamma \end{cases}$$

By Proposition 3.11, this is a well-defined interpretation of $\mathscr{L}_S$. We will show that $val^{\sigma_\Gamma}(\varphi) = \mathbf{T}$ if and only if $\varphi \in \Gamma$.[8] The proof will proceed by induction on φ. If φ is an atomic sentence, then this follows directly from the definition of σ_Γ.

Suppose that $\varphi = \neg\psi$. If $\varphi \in \Gamma$, then $\psi \notin \Gamma$, and so by hypothesis, $val^{\sigma_\Gamma}(\psi) = \mathbf{F}$. Hence, $val^{\sigma_\Gamma}(\varphi) = \mathbf{T}$. Now suppose that $\varphi \notin \Gamma$. Then $\psi \in \Gamma$, and so by hypothesis $val^{\sigma_\Gamma}(\psi) = \mathbf{T}$. Hence $val^{\sigma_\Gamma}(\varphi) = \mathbf{F}$.

Now suppose that $\varphi = \psi \to \rho$ and that $\varphi \in \Gamma$. Then either $\psi \notin \Gamma$ or $\rho \in \Gamma$. For if $\rho \notin \Gamma$, then $\neg\rho \in \Gamma$, and so $\Gamma \vdash \neg\rho$. But if it is also the case that $\varphi, \psi \in \Gamma$, then $\Gamma \vdash \rho$ by one application of MP, contradicting the assumption that Γ is consistent. If $\psi \notin \Gamma$, then by hypothesis, $val^{\sigma_\Gamma}(\psi) = \mathbf{F}$. Hence, $val^{\sigma_\Gamma}(\varphi) = \mathbf{T}$. If, on the other hand, $\rho \in \Gamma$, then, by hypothesis, $val^{\sigma_\Gamma}(\rho) = \mathbf{T}$. Hence, $val^{\sigma_\Gamma}(\varphi) = \mathbf{T}$. Thus, if $\varphi \in \Gamma$, $val^{\sigma_\Gamma}(\varphi) = \mathbf{T}$. Now suppose that $\varphi \notin \Gamma$, then both $\psi \in \Gamma$ and $\rho \notin \Gamma$, for otherwise either $\neg\psi \in \Gamma$ or $\rho \in \Gamma$, and from either of these claims one can prove φ (see Exercise 2.10, c,d), contradicting the assumption that Γ is consistent. But if $\psi \in \Gamma$ and $\rho \notin \Gamma$, then, by hypothesis, $val^{\sigma_\Gamma}(\psi) = \mathbf{T}$ and $val^{\sigma_\Gamma}(\rho) = \mathbf{F}$. Hence $val^{\sigma_\Gamma}(\varphi) = \mathbf{F}$. □

All then that remains to be shown is that every consistent set can be extended to a maximally consistent set in the manner described above. In the context of first-order logic, this claim is typically referred to as Lindenbaum's Lemma. We will refer to it as the extension theorem.

Theorem 2.14 (Extension Theorem). *If Γ is a consistent set, then there exists a maximally consistent set $\Gamma^* \supset \Gamma$.*

Proof. Let $\varphi_1, \varphi_2, \varphi_3, \ldots$ be an enumeration of all of the sentences of $\mathscr{L}_S$ (see Example . We construct an increasing sequence of sets $\Gamma_0 \subset \Gamma_1 \subset \Gamma_2 \subset \ldots$ as follows: Let $\Gamma_0 = \Gamma$. Suppose that Γ_n ($n \geq 0$) has been constructed and that it is a consistent set. Let:

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

[8] Strictly speaking, we only need to show the if part of this claim, but to carry through with the induction it turns out to be easier to prove the claim in this slightly stronger form.

By Proposition 3.10, Γ_{n+1} is consistent, and so the construction can continue. Let Γ^* be the set of all sentences which appear in some set in this sequence, i.e.:

$$\Gamma^* = \bigcup_{i=0}^{\infty} \Gamma_i$$

We will prove that Γ^* is a maximally consistent set. Suppose, for contradiction, that Γ^* is inconsistent. Then for some φ, $\Gamma^* \vdash \varphi$ and $\Gamma^* \vdash \neg\varphi$. Given a proof of φ from Γ^* and a proof of $\neg\varphi$ from Γ^*, let Δ be the set of all of the sentences in Γ^* that appear somewhere in this proof. For each $\psi \in \Delta$, let $n(\psi)$ be the smallest n such that $\psi \in \Gamma_n$. Since proofs are, by definition, of finite length, Δ is a finite set, and so:

$$N = \max\{n(\psi) : \psi \in \Delta\}$$

is a finite number, and every sentence in $\Delta \in \Gamma_N$. But then the aforementioned proofs of φ and $\neg\varphi$ are also proofs from Γ_N, so that Γ_N is inconsistent. But this contradiction the above definition of Γ_N. Thus, Γ^* is consistent.

To see that Γ^* is maximal, note that, by the definition of Γ^*, for every sentence φ, either φ or $\neg\varphi$ belongs to the set. Hence, if a sentence not already in Γ^* were to be added to Γ^*, then both this sentence and its negation would appear in the set, and so any set which has Γ^* as a proper subset is inconsistent. □

Example 7. In our proof of the extension theorem, we took for granted that the sentences of $\mathscr{L}_S$ can be enumerated in a sequence. In this example, we verify that this is indeed the case.

> Definition 2.14. A set S is *enumerable* iff there exists a sequence $a_1, a_2, a_3, \ldots$ of members of S, such that for every $x \in S$, there exists a number n, for which $x = a_n$. We refer to such a sequence as an *enumeration* of S.

Clearly, every finite set is enumerable, for if $S = \{a_1, \ldots, a_n\}$, we can enumerate the members of S as follows:

$$a_1, a_2, \ldots, a_n, a_1, a_1, \ldots.$$

There are also infinite sets which are enumerable, e.g., the set of natural numbers itself:

$$1, 2, 3 \ldots.$$

Moreover, it is easy to see that if $T \subset S$, and S is enumerable, then T is enumerable as well, for if $a_1, a_2, \ldots,$ is an enumeration of S, and c is some member of T (if T is empty, then it is trivially enumerable), then the sequence $b_1, b_2, \ldots$, where:

$$b_i = \begin{cases} a_i & \text{if } a_i \in T \\ c & \text{otherwise} \end{cases}$$

is an enumeration of S.

EXERCISE 2.13. Let f be a function which assigns to each of the members of the set X, some value in the set Y. f is *surjective* (or a surjection) if for every $y \in Y$ there is at least one $x \in X$ such that $f(x) = y$. f is *injective* (or an injection) if for every $y \in Y$ there is at most one $x \in X$ such that $f(x) = y$. f is *bijective* (or a bijection) if it is both surjective and injective, i.e., if for every $y \in Y$, there is exactly one $x \in X$ such that $f(x) = y$.

Show that enumerability is equivalent to each of the following three conditions:

(a). There exists a surjection from the set of natural numbers to S.

(b). There exists an injection from S to the set of natural numbers.

(c). There exists a bijection from S to some subset of the set of natural numbers.

THEOREM 2.15. *Let $S_1, S_2, \ldots$ be enumerable sets. Then the set:*

$$S = \bigcup_{i=1}^{\infty} S_i$$

is enumerable.

PROOF. For each $j = 1, 2, \ldots$, let $a_{j,1}, a_{j,2}, \ldots$, be an enumeration of S_j. Then all of the members of S appear somewhere in the table:

$$\begin{array}{cccc} a_{1,1} & a_{1,2} & a_{1,3} & \cdots \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

We can enumerate the members of S by proceeding down consecutive upper-right-to-lower-left diagonals in the table as follows:

$$a_{1,1}, a_{1,2}, a_{2,1}, a_{1,3}, a_{2,2}, a_{3,1}, \ldots$$

□

THEOREM 2.16. *Let S be an enumerable set, and let S_k, $k = 1, 2, \ldots$, be the set of all k-length sequences of members of S. Then S_k is enumerable.*

PROOF. The proof proceeds by induction on k. For $k = 1$, $S_k = S$, and hence is enumerable. Now suppose that S_k ($k \geq 1$) is enumerable, and let $a_1, a_2, \ldots$ be an enumeration of S_k. Let $S^i_{k+1} = \{a_i b : b \in S\}$. Clearly, each of the sets S^i_{k+1} is enumerable and so since

$$S_{k+1} = \bigcup_{i=1}^{\infty} S^i_{k+1}$$

S_{k+1} is enumerable by Theorem 2.15. □

COROLLARY 2.17. *Let S be an enumerable set. The set of all finite sequences of members of S is enumerable.*

PROOF. Since the set of all finite sequences of members of S is the union of the sets of all k-length sequences of members of S, for $k = 1, 2, \ldots$, this claim follows from Theorems 2.16 and 2.15. □

COROLLARY 2.18. *The set of all sentences of $\mathscr{L}_S$ is enumerable.*

Proof. The alphabet of $\mathscr{L}_S$ can be enumerated as follows:

$$\neg, \rightarrow, (,), \mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \ldots.$$

Hence, by Corollary 2.17, the set of all expressions of $\mathscr{L}_S$ is enumerable, and so since the set of all sentences of $\mathscr{L}_S$ is a subset of this latter set, this set too is enumerable. □

Exercise 2.14 (*). Let p_i be the ith prime number (i.e., $p_1 = 2, p_2 = 3, p_3 = 5, \ldots$) and let $n_1, \ldots, n_k$ and $m_1, \ldots, m_k$ be any natural numbers.

(a). Show that

$$p_1^{n_1} \cdot p_2^{n_2} \cdots p_k^{n_k} = p_1^{m_1} \cdot p_2^{m_2} \cdots p_{k'}^{m_{k'}}$$

if and only if $k = k'$ and $n_i = m_i$, for $i = 1, \ldots, k$.

(b). Use (a) to give a direct proof of Corollary 2.17 by constructing an injective function from the set of all finite sequences of S to the set of natural numbers.

Are there sets which are not enumerable? We will show that the answer to this question is yes, by proving that if we admitted sentences of infinite length in our language, e.g.:

$$\mathbf{P}_1 \wedge \mathbf{P}_2 \wedge \mathbf{P}_3 \wedge \cdots$$

then Corollary 2.18 would not hold. The proof proceeds by Cantor's method of diagonalization.

Theorem 2.19. *Let S be an enumerable set with at least two members and let S^∞ be the set of all infinite sequences of members in S. Then S^∞ is not enumerable.*

Proof. Suppose for contradiction that S^∞ were enumerable, and let $a_{i,1}, a_{i,2}, \ldots$ be the ith term in an enumeration of this set. We will construct a sequence that belongs to S^∞ but does not appear in this enumeration. Let c_1 and c_2 be any two distinct elements in S, and for each $i = 1, 2, \ldots$, let

$$b_i = \begin{cases} c_1 & \text{if } a_{i,i} \neq c_1 \\ c_2 & \text{otherwise} \end{cases}$$

Then the sequence $b_1, b_2, \ldots$ is in S^∞, but does not appear in the above enumeration of the set since, by definition, $b_i \neq a_{i,i}$ for all $i = 1, 2, \ldots$. □

3.5. The Compactness Theorem. The soundness and completeness of T_S has as a corollary, the following important result:

Theorem 2.20 (The Compactness Theorem for $\mathscr{L}_S$). *If $\Gamma \models \psi$, then there exists some finite set $\Gamma_0 \subset \Gamma$, such that $\Gamma_0 \models \psi$.*

Proof. If $\Gamma \models \psi$, then, by the completeness theorem, $\Gamma \vdash \psi$. Let Γ_0 be the set of all sentences in Γ which appear in a given proof of ψ from Γ. Then $\Gamma_0 \vdash \psi$ and so, by the soundness theorem, $\Gamma_0 \models \psi$. But since a proof consists of only finitely many sentences, Γ_0 is finite. □

Exercise 2.15. Show that completeness is equivalent to compactness in conjunction with the following fact: For every sentence φ, if $\models \varphi$, then $\vdash \varphi$.

An equivalent formulation of the compactness theorem is given by the following claim:

Theorem 2.21. *Let Γ be any set of sentences. If every finite subset of Γ is satisfiable, then Γ is satisfiable.*

To see that this claim is equivalent to compactness, first suppose that Theorem 2.21 holds and that $\Gamma \models \psi$. Then by Proposition 3.8, $\Gamma \cup \{\neg\psi\}$ is not satisfiable, and so there exists some finite set $\Gamma_0' \subset \Gamma \cup \{\neg\psi\}$ which is not satisfiable. This means that there exists a finite set $\Gamma_0 \subset \Gamma$, such that $\Gamma_0' \subset \Gamma_0 \cup \{\neg\psi\}$. Hence, $\Gamma_0 \cup \{\neg\psi\}$ is not satisfiable, i.e., $\Gamma_0 \models \psi$.

Now, suppose that the compactness theorem holds and that Γ is unsatisfiable. Then, for any sentence ψ, $\Gamma \models \neg(\psi \to \psi)$ and so for some finite $\Gamma_0 \subset \Gamma$, $\Gamma_0 \models \neg(\psi \to \psi)$. This means that Γ_0 is unsatisfiable. Hence, Theorem 2.21 holds.

Example 8. When a property can be expressed in terms of the satisfiability of a certain infinite number of sentences of $\mathscr{L}_S$, then the compactness theorem guarantess that this property can be expressed in terms of other strictly finite properties. In this example, we will consider a particular consequence of this fact.

Definition 2.15. A *partial order* on the set S is a binary relation $\prec$ such that, for all $x, y, z \in S$:

(i). If $x \prec y$, then $y \nprec x$.

(ii). If $x \prec y$ and $y \prec z$, then $x \prec z$.

A *total order* on the set S is a partial order on S, which also satisfies the condition that, for all $x, y \in S$:

(iii). Either $x \prec y$, $y \prec x$ or $x = y$.

We say that the total order $\prec^*$ *extends* (or is an extension of) the partial order $\prec$, if $x \prec y$ then $x \prec^* y$, for all $x, y \in S$.

Theorem 2.22. *Let S be a finite set. Every partial ordering on S has a total extension.*

Proof. If S is the empty set then a partial ordering on S is also a total ordering. Thus, we may assume without loss of generality that S is nonempty.

Let $\prec$ be a partial order on S. We refer to $x \in S$ as a *maximal member of* S if $x \nprec y$, for all $y \in S$. First, suppose that S has no maximal member. Take $x_1, \ldots, x_n \in S$ to have been chosen such that $x_1 \prec x_2 \prec \cdots \prec x_n$. Since x_n is not a maximal member of S, there exists a $y \in S$, such that $x \prec y$. Let $x_{n+1} = y$. In this way we generate an infinite sequence of members of S, $x_1, x_2, \ldots$, such that:

$$x_1 \prec x_2 \prec x_3 \prec \cdots$$

From the definition of a partial order, it follows that $x_i \neq x_j$ for all $i \neq j$. This means that $x_1, x_2, \ldots$ are all distinct members of S, which contradicts the assumption that S is finite. Hence, S has at least one maximal member.

We will now prove the theorem by induction on the size (or cardinality) of S. Suppose that S contains n members, and let $x^* \in S$ be a maximal member of S. Let $S_0 = S - \{x^*\}$ be the set which results from removing x^* from S, and let $\prec_0$ be the partial order induced on S_0 by $\prec$. By the induction hypothesis, $\prec_0$ has a total extension $\prec_0^*$ on S_0. Let $\prec^*$ be a binary relation on S which satisfies the following properties:

(1) If $x, y \in S_0$, then $x \prec^* y$ iff $x \prec_0^* y$.

(2) For all $x \in S_0$, $x \prec^* x^*$.

It is easy to verify that $\prec^*$ is a total order on S which extends $\prec$. □

We can appeal to the compactness theorem to show that Theorem 2.22 applies to all enumerable sets, whether finite or infinite.

EXERCISE 2.16. Suppose that the atoms of $\mathscr{L}_S$ are doubly-indexed, so that instead of $\mathbf{P}_1, \mathbf{P}_2, \ldots$, the atomic sentences of the language are $\mathbf{P}_{1,1}, \mathbf{P}_{1,2}, \mathbf{P}_{2,1}, \ldots$.

Let S be an infinite enumerable set and let $x_1, x_2, \ldots$ be an enumeration of S. Let $\prec$ be a partial order on S and let $\prec^*$ be a binary relation on S.

(a). Suppose that we interpret the claim $val^{\sigma}(\mathbf{P}_{i,j}) = \mathbf{T}$ as meaning $x_i \prec^* x_j$. Construct a set of sentences which is satisfiable iff $\prec^*$ is a total extension of $\prec$.

(b). Use the compactness theorem to prove that $\prec$ has a total extension.

CHAPTER 3

First-Order Logic

1. The Syntax and Semantics of First-Order Logic

1.1. First-Order Languages. A *first-order language* is a formal language whose alphabet consists of:

- The two symbols:

$$\neg \qquad \rightarrow$$

  referred to collectively as *sentential connectives*, and, respectively, as the *negation operator* and the *conditional operator*.

- An infinite number of symbols:

$$\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots,$$

  referred to as *variables*.

- The symbol:

$$\forall$$

  referred to as the *universal quantifier*.

- Either none, a finite number, or an infinite number of symbols, referred to as *constant symbols*.

- For each $k \geq 1$, either none, a finite number, or an infinite number of symbols, referred to as *k-ary function symbols*. We refer to these symbols collectively as *function symbols*.

- For each $k \geq 1$, either none, a finite number, or an infinite number of symbols, referred to as *k-ary relation symbols*. We refer to these symbols collectively as *relation symbols*.

- The three punctuation marks:

$$(\qquad) \qquad ,$$

  referred to as the left and right parentheses, and the comma, respectively.

Note that the minimal alphabet for a first-order language consists of the sentential connectives, the universal quantifier, the variables, the equality symbol and the punctuation marks. The remaining symbols in the alphabet of a first-order language, are referred to as that language's *non-logical vocabulary*.

The formation rules for first-order languages will be defined in two stages. First, we will define the notion of a *formula* of the language. We then define the sentences of the language as a particular subclass of the formulas.

Definition 3.1. The *terms* of a first-order language are defined inductively as follows:

(1) Every variable and every constant symbol is a term. We refer to these terms as *atomic terms*.

(2) If $\mathbf{f}$ is a k-ary function symbol, and $t_1, \ldots, t_k$ are terms, then the expression $\mathbf{f}(t_1, \ldots, t_k)$ is a term.

A term is *open* if a variable occurs in it. It is *closed* otherwise. When we wish to emphasize that the only variables which occur in the term t are among $x_1, \dots, x_n$, we will denote t by $t(x_1, \dots, x_n)$.

Definition 3.2. The formulas of a first-order language are defined inductively as follows:

(1) if $\mathbf{R}$ is a k-ary relation symbol and $t_1, \dots, t_k$ are terms, then $\mathbf{R}(t_1, \dots, t_k)$ is a formula. We refer to these formulas as *atomic formulas*.

(2) If φ is a formula, then $(\neg\varphi)$ is a formula.

(3) If φ and ψ are formulas, then $(\varphi \to \psi)$ is a formula.

(4) If φ is a formula and x is a variable, then $(\forall x\varphi)$ is a formula.

The formula $(\forall x\varphi)$ is referred to as a *universal quantification over the variable* x, and, more specifically, as the universal quantification over x of φ.

Let $\mathscr{L}$ be a first-order language. If $\mathbf{f}$ is a 1-ary function symbol of $\mathscr{L}$; $\mathbf{g}$ a 2-ary function symbol; $\mathbf{c}$ a constant symbol; $\mathbf{R}$ a 1-ary relation symbol and $\mathbf{S}$ a 2-ary relation symbol, then the following are all formulas of $\mathscr{L}$:

(i) $\mathbf{R}(\mathbf{c})$

(ii) $\mathbf{R}(\mathbf{v}_1) \to \mathbf{S}(\mathbf{c}, \mathbf{v}_2)$

(iii) $\mathbf{S}(\mathbf{f}(\mathbf{c}), \mathbf{g}(\mathbf{v}_1, \mathbf{v}_2))$

(iv) $\neg\mathbf{S}(\mathbf{v}_1, \mathbf{g}(\mathbf{v}_2, \mathbf{c}))$

(v) $\mathbf{S}(\mathbf{c}, \mathbf{c}) \to \mathbf{R}(\mathbf{c})$

(vi) $\forall\mathbf{v}_1(\mathbf{S}(\mathbf{v}_1, \mathbf{c}) \to \mathbf{R}(\mathbf{v}_1))$

(vii) $\forall\mathbf{v}_1\neg(\mathbf{R}(\mathbf{f}(\mathbf{v}_2))$

(viii) $\forall\mathbf{v}_1(\mathbf{R}(\mathbf{g}(\mathbf{c}, \mathbf{v}_1)) \to \forall\mathbf{v}_2(\mathbf{S}(\mathbf{v}_1, \mathbf{f}(\mathbf{v}_2))))$

In accordance with the definition given above, first-order languages include only the two sentential connectives, $\neg$ and $\to$. As was the case in our study of sentential logic, the purpose of this restriction is to simplify the proof of claims expressing facts about such languages construed as syntactical entities. Nevetheless, for semantical purposes, it is often convenient to utilize all five connectives when expressing formulas. We may take advantage of these conveniences by interpreting the connectives $\wedge$, $\vee$ and $\leftrightarrow$, not as primitives of the language, but as shorthand, introduced in our own 'metalanguage' (i.e., the language in which we express reasoning about first-order languages). In particular, we may interpret $\varphi \wedge \psi$ as shorthand for the expression $\neg(\varphi \to \neg\psi)$; $\varphi \vee \psi$ as shorthand for $\neg\varphi \to \psi$; and $\varphi \leftrightarrow \psi$ as shorthand for $\neg((\varphi \to \psi) \to \neg(\psi \to \varphi))$.

In addition, we write

$$\exists x\varphi$$

for

$$\neg\forall x\neg\varphi.$$

and we refer to the latter formula as the *existential quantification over the variable* x of φ (hence, strictly speaking, an existential quantification is a particular sort of negation). We write $\forall x_1 \dots x_k\varphi$ for the formula $\forall x_1 \dots \forall x_k\varphi$, and we write $\exists x_1 \dots x_k\varphi$ for the formula $\exists x_1 \dots \exists x_k\varphi$.

Definition 3.3. Let x be any variable in a first-order language. The formulas of that language in which x *occurs freely* are defined inductively as follows:

(1) If $\mathbf{R}$ is a k-ary relation symbol and x occurs in the term t_n, where $1 \leq n \leq k$, then x occurs freely in the formula $\mathbf{R}(t_1, \dots, t_n, \dots, t_k)$.

(2) If x occurs freely in the formula φ, then x occurs freely in $\neg\varphi$.

(3) If x occurs freely in either of the formulas φ or ψ, then x occurs freely in $\varphi \rightarrow \psi$.

(4) If x occurs freely in the formula φ, and y is any variable other than x, then x occurs freely in $\forall y\varphi$.

If the variable x occurs in φ, but does not occur freely, we say that *x occurs bound* in φ. When we wish to emphasize that the only variables which occur freely in the formula φ are among $x_1, \ldots, x_n$, we will denote φ by $\varphi(x_1, \ldots, x_n)$.[1]

The intended meaning of a freely occurring variable will become clearer once we introduce the semantics for first-order languages. For now, we will simply use the notion to define which of the formulas are to count as sentences.

Definition 3.4. A *sentence* of a first-order language is a formula in which no variable occurs freely.

Since the alphabet of a first-order language comprises an enumerable set, the set of all expressions of the language is enumerable (see Example 7, p. 42), and so *a fortiori* the set of all formulas and the set of all sentences of a first-order language is enumerable.

Exercise 3.1. Of the eight formulas listed above, which are sentences?

Exercise 3.2. For a given first-order language, let Σ be the function which assigns to each formula φ of the language, a set of formulas $\Sigma(\varphi)$, in accordance with the following conditions:

(1) If φ is an atomic formula, then:

$$\Sigma(\varphi) = \{\varphi\}.$$

(2) If $\varphi = \neg\psi$, then:

$$\Sigma(\varphi) = \{\varphi\} \cup \Sigma(\psi).$$

(3) If $\varphi = \psi \rightarrow \rho$, then:

$$\Sigma(\varphi) = \{\varphi\} \cup \Sigma(\psi) \cup \Sigma(\rho).$$

(4) If $\varphi = \forall x\psi$, then:

$$\Sigma(\varphi) = \{\varphi\} \cup \Sigma(\psi).$$

Let x be any variable. We say that x is *quantified over in* φ if for every $\psi \in \Sigma(\varphi)$ in which x occurs, there exists some $\rho \in \Sigma(\varphi)$, such that (i) $\psi \in \Sigma(\rho)$ and ρ is a universal quantification over x.

Show that x occurs freely in φ iff it is *not* quantified over in φ.

1.2. Models, Satisfaction and Truth. The non-logical elements of the syntax of a first-order language are the symbols in its non-logical vocabulary. An interpretation of the language assigns to these symbols a particular semantic value according to their syntactic type. Such an assignment takes place in the context of a given set of objects, referred to as the interpretation's domain of discourse (or simply its *domain*, for short). We refer to an interpretation of a first-order language taken together with its associated domain as a *model* of the language.

[1]In what follows, the context will always clearly indicate whether we intend this usage of $\varphi(x_1, \ldots, x_k)$, or whether $\varphi(x_1, \ldots, x_k)$ names a certain expression in the language.

DEFINITION 3.5. Let $\mathscr{L}$ be a first-order language. A *model for* $\mathscr{L}$ is a structure $\mathscr{M}$ consisting of:

(1) A nonempty set $|\mathscr{M}|$, referred to as the *domain* of the model.

(2) A function which assigns to each constant symbol $\mathbf{c}$ of $\mathscr{L}$, an object $\mathbf{c}^{\mathscr{M}} \in |\mathscr{M}|$.

(3) A function which assigns to each k-ary function symbol $\mathbf{f}$, a function $\mathbf{f}^{\mathscr{M}} : |\mathscr{M}|^k \to |\mathscr{M}|$.

(4) A function which assigns to each k-ary relation symbol $\mathbf{R}$, a k-ary relation $\mathbf{R}^{\mathscr{M}}$ on $|\mathscr{M}|$.

The semantical rules for first-order languages must associate with each model of the language, a unique valuation, i.e., a function which assigns to each of the sentences of the language a value in the set $\{\mathbf{T}, \mathbf{F}\}$. In order to state these rules, we first define the more general notion of satisfaction in a model for a given assignment of values, which is a property that applies to all formulas of the language. (NB: throughout the remainder of this section, we take $\mathscr{L}$ to be an arbitrary first-order language and $\mathscr{M}$ to be an arbitrary model for $\mathscr{L}$).

DEFINITION 3.6. A *assignment for* $\mathscr{M}$ is a function which assigns to each variable of $\mathscr{L}$ an object in $|\mathscr{M}|$.

Let θ be a assignment for $\mathscr{M}$. We asociate with each term t of $\mathscr{L}$, an object $t[\theta] \in |\mathscr{M}|$ as follows:

(1) If t is a constant symbol, then $t[\theta] = t^{\mathscr{M}}$.

(2) If t is a variable, then $t[\theta] = \theta(t)$.

(3) If $t = \mathbf{f}(t_1, \ldots, t_k)$, then

$$t[\theta] = \mathbf{f}^{\mathscr{M}}(t_1[\theta], \ldots, t_k[\theta]).$$

Let θ be an assignment for $\mathscr{M}$, x a variable of $\mathscr{L}$, and c an object in $|\mathscr{M}|$. We write θ^x_c for the assignment defined by:

$$\theta^x_c(v) = \begin{cases} c & \text{if } v = x \\ \theta(v) & \text{otherwise} \end{cases}$$

If $x_1, \ldots, x_n$ ($n > 1$) are any variables of $\mathscr{L}$ and $c_1, \ldots, c_n$ are any objects in $|\mathscr{M}|$, we write $\theta^{x_1 \cdots x_n}_{c_1 \cdots c_n}$ for the assignment:

$$(\theta^{x_1 \cdots x_{n-1}}_{c_1 \cdots c_{n-1}})^{x_n}_{c_n}$$

(Note that it follows from this definition that if $x_i = x_j$, for $i < j \leq n$, then $\theta^{x_1 \cdots x_n}_{c_1 \cdots c_n}(x_i) = c_j$).

PROPOSITION 3.1. *For any two assignment θ, ξ, and any objects $c_1, \ldots, c_n \in |\mathscr{M}|$:*

$$t(x_1, \ldots, x_n)[\theta^{x_1 \cdots x_n}_{c_1 \cdots c_n}] = t(x_1, \ldots, x_n)[\xi^{x_1 \cdots x_n}_{c_1 \cdots c_n}]$$

PROOF. The proof proceed by induction on t. If t is a constant symbol, then:

$$t[\theta^{x_1 \cdots x_n}_{c_1 \cdots c_n}] = t[\xi^{x_1 \cdots x_n}_{c_1 \cdots c_n}] = t^{\mathscr{M}}.$$

If t is the variable x_i ($1 \leq i \leq n$), then:

$$t[\theta^{x_1 \cdots x_n}_{c_1 \cdots c_n}] = t[\xi^{x_1 \cdots x_n}_{c_1 \cdots c_n}] = c_j,$$

where j is the smallest number for which $x_j = x_i$.

If t is of the form $\mathbf{f}(t_1, \ldots, t_k)$, then the only variables occurring in any of the t_i's are among $x_1, \ldots, x_n$. Suppose and that the claim holds for $t_1, \ldots, t_k$. Then

$$t[\theta^{x_1 \cdots x_n}_{c_1 \cdots c_n}] = \mathbf{f}^{\mathscr{M}}(t_1[\theta^{x_1 \cdots x_n}_{c_1 \cdots c_n}], \ldots, t_k[\theta^{x_1 \cdots x_n}_{c_1 \cdots c_n}]) = \mathbf{f}^{\mathscr{M}}(t_1[\xi^{x_1 \cdots x_n}_{c_1 \cdots c_n}], \ldots, t_k[\xi^{x_1 \cdots x_n}_{c_1 \cdots c_n}]) = t[\xi^{x_1 \cdots x_n}_{c_1 \cdots c_n}].$$

□

COROLLARY 3.1. *If t is a closed term, then $t[\theta] = t[\xi]$, for any two assignments θ, ξ.*

DEFINITION 3.7. The relation of a model $\mathscr{M}$ satisfying the formula φ under the assignment θ (written $\mathscr{M} \models \varphi[\theta]$) is defined inductively as follows:

(1) If φ is the atomic formula $\mathbf{R}(t_1, \ldots, t_k)$ then $\mathscr{M} \models \varphi[\theta]$ iff $\mathbf{R}^{\mathscr{M}}(t_1[\theta], \ldots, t_k[\theta])$.

(2) If φ is the formula $\neg\psi$, then $\mathscr{M} \models \varphi[\theta]$ iff $\mathscr{M} \not\models \psi[\theta]$.

(3) If φ is the formula $\psi \to \rho$, then $\mathscr{M} \models (\psi \to \rho)[\theta]$ iff either $\mathscr{M} \not\models \psi[\theta]$ or $\mathscr{M} \models \rho[\theta]$.

(4) If φ is the formula $\forall x\psi$, then $\mathscr{M} \models \varphi[\theta]$ iff, for every $c \in |\mathscr{M}|$, $\mathscr{M} \models \psi[\theta^x_c]$.

PROPOSITION 3.2. *For any two assignments θ, ξ, and any objects $c_1, \ldots, c_n \in |\mathscr{M}|$:*

$$\mathscr{M} \models \varphi(x_1, \ldots, x_n)[\theta^{x_1\cdots x_n}_{c_1\cdots c_n}] \Leftrightarrow \mathscr{M} \models \varphi(x_1, \ldots, x_n)[\xi^{x_1\cdots x_n}_{c_1\cdots c_n}].$$

PROOF. The proof proceeds by induction on φ. If φ is an atomic formula, then it is of the form

$$\mathbf{R}(t_1(x_1, \ldots, x_n), \ldots, t_k(x_1, \ldots, x_n)).$$

Thus, by Proposition 3.1 and clause (1) of Definition 3.7:

$$\mathscr{M} \models \varphi[\theta^{x_1\cdots x_n}_{c_1\cdots c_n}] \Leftrightarrow \mathbf{R}^{\mathscr{M}}(t_1[\theta^{x_1\cdots x_n}_{c_1\cdots c_n}], \ldots, t_k[\theta^{x_1\cdots x_n}_{c_1\cdots c_n}]) \Leftrightarrow \mathbf{R}^{\mathscr{M}}(t_1[\xi^{x_1\cdots x_n}_{c_1\cdots c_n}], \ldots, t_k[\xi^{x_1\cdots x_n}_{c_1\cdots c_n}]) \Leftrightarrow \mathscr{M} \models \varphi[\xi^{x_1\cdots x_n}_{c_1\cdots c_n}]$$

Now, suppose that φ is the formula $\neg\psi(x_1, \ldots, x_n)$. Then, by the induction hypothesis and clause (2) in Definition 3.7:

$$\mathscr{M} \models \varphi[\theta^{x_1\cdots x_n}_{c_1\cdots c_n}] \Leftrightarrow \mathscr{M} \not\models \psi[\theta^{x_1\cdots x_n}_{c_1\cdots c_n}] \Leftrightarrow \mathscr{M} \not\models \psi[\xi^{x_1\cdots x_n}_{c_1\cdots c_n}] \Leftrightarrow \mathscr{M} \models \varphi[\xi^{x_1\cdots x_n}_{c_1\cdots c_n}]$$

A similar argument applies to the case in which φ is the formula $\psi(x_1, \ldots, x_n) \to \rho(x_1, \ldots, x_n)$.

Now, suppose that φ is the formula $\forall x_{n+1}\psi(x_1, \ldots, x_n, x_{n+1})$. Then, by the induction hypothesis and clause (4) of Definition 3.7:

$$\begin{gathered} \mathscr{M} \models \varphi[\theta^{x_1\cdots x_n}_{c_1\cdots c_n}] \\ \Updownarrow \\ \text{for every } c \in |\mathscr{M}|, \mathscr{M} \models \psi[\theta^{x_1\cdots x_{n+1}}_{c_1\cdots c}] \\ \Updownarrow \\ \text{for every } c \in |\mathscr{M}|, \mathscr{M} \models \psi[\xi^{x_1\cdots x_{n+1}}_{c_1\cdots c}] \\ \Updownarrow \\ \mathscr{M} \models \varphi[\xi^{x_1\cdots x_n}_{c_1\cdots c_n}] \end{gathered}$$

This completes the proof. □

Proposition 3.2 affords us with a semantic interpretation of the notion of a freely occurring variable. It states that the freely occurring variables in φ are those variables whose values may be relevant for assessing whether or not φ is satisfied in a given model. Conversely, the variables which occur bound in φ are those variables whose values are irrelevant for assessing whether or not φ is satisfied in a given model. In light of this fact, for any formula $\varphi(x_1, \ldots, x_n)$, we may simply write:

$$\mathscr{M} \models \varphi[^{x_1\cdots x_n}_{c_1\cdots c_n}]$$

to express the fact that φ is satisfied in $\mathscr{M}$ under the assignment $\theta^{x_1\cdots x_n}_{c_1\cdots c_n}$, for some arbitrary assignment θ.

We have as a corollary of Proposition 3.2 that

COROLLARY 3.2. *If φ is a sentence of $\mathscr{L}$ then $\mathscr{M} \models \varphi[\theta]$ iff $\mathscr{M} \models \varphi[\xi]$, for any two assignments θ, ξ.*

Thus, if φ is a sentence of $\mathscr{L}$, we may simply speak of φ being satisfied in a given model $\mathscr{M}$, and assume that this statement is made with respect to some arbitrary assignment. In this case, we simply write $\mathscr{M} \models \varphi$. The single semantical rule for the language states that a sentence is true in a given model just in case it is satisified in that model:

Semantical Rule: Let φ be any sentence of $\mathscr{L}$, then: $val^{\mathscr{M}}(\varphi) = \mathbf{T}$ iff $\mathscr{M} \models \varphi$.

In general, it will be more convenient to speak of satisfaction in a model (under a given assignment), rather than truth under a model since the former notion possesses a wider scope of applicability. In particular we will extend the notion of logical consequence so as to apply to all formulas of $\mathscr{L}$:

DEFINITION 3.8. Let Γ be a set of formulas of $\mathscr{L}$ and let ψ be a formula. Then we write

$$\Gamma \models \psi$$

to express the fact that for any model $\mathscr{M}$ of $\mathscr{L}$ and any assignment θ for $\mathscr{M}$: If $\mathscr{M} \models \varphi[\theta]$ for all $\varphi \in \Gamma$, then $\mathscr{M} \models \psi[\theta]$.

Clearly, if Γ is a set of sentences and ψ is a sentence, then $\Gamma \models \psi$ iff, for any model $\mathscr{M}$: If $val^{\mathscr{M}}(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$, then $val^{\mathscr{M}}(\varphi) = \mathbf{T}$.

We take note of the following propositions which follow straightforwardly from our definition of satisfiability:

PROPOSITION 3.3. *$\mathscr{M} \models \forall x_1 \ldots x_k \varphi(x_1, \ldots, x_k)$ iff, for all $c_1, \ldots, c_k \in |\mathscr{M}|$, $\mathscr{M} \models \varphi[{}^{x_1 \cdots x_k}_{c_1 \cdots c_k}]$.*

PROOF. The proof proceed by induction on k. If $k = 1$, then the proposition follows directly from clause (4) of Definition 3.7. Suppose that $k > 1$, and let $\psi(x_1, \ldots, x_{k-1})$ be the formula $\forall x_k \varphi(x_1, \ldots, x_k)$. Then, by the induction hypothesis:

$$\mathscr{M} \models \forall x_1 \ldots x_k \varphi(x_1, \ldots, x_k)$$

iff, for all $c_1, \ldots, c_{k-1} \in |\mathscr{M}|$:

$$\mathscr{M} \models \forall x_k \varphi[{}^{x_1 \cdots x_{k-1}}_{c_1 \cdots c_{k-1}}].$$

But from clause (4) of 3.7, this is true iff for all $c_k \in |\mathscr{M}|$:

$$\mathscr{M} \models \varphi[{}^{x_1 \cdots x_k}_{c_1 \cdots c_k}].$$

□

PROPOSITION 3.4. *$\mathscr{M} \models \exists x_1 \ldots x_k \varphi(x_1, \ldots, x_k)$ iff, for some $c_1, \ldots, c_k \in |\mathscr{M}|$, $\mathscr{M} \models \varphi[{}^{x_1 \cdots x_k}_{c_1 \cdots c_k}]$.*

PROOF. $\exists x_1, \ldots, x_k \varphi(x_1, \ldots, x_k)$ is shorthand for the formula:

$$\neg \forall x_1, \ldots, x_k \neg \varphi(x_1, \ldots, x_k)$$

Hence $\mathscr{M} \models \exists x_1, \ldots, x_k \varphi(x_1, \ldots, x_k)$ iff $\mathscr{M} \not\models \forall x_1, \ldots, x_k \neg \varphi(x_1, \ldots, x_k)$. By Proposition 3.3, this is true iff there is some $c_1, \ldots, c_k \in |\mathscr{M}|$, such that $\mathscr{M} \not\models \neg \varphi[{}^{x_1 \cdots x_k}_{c_1 \cdots c_k}]$, which is true iff $\mathscr{M} \models \varphi[{}^{x_1 \cdots x_k}_{c_1 \cdots c_k}]$. □

EXAMPLE 9. In this example, we appeal to the notion of satisfaction to state what it means for a relation (or property) to be definable in a first-order language.

DEFINITION 3.9. Let R be a k-ary relation on $|\mathscr{M}|$. We say that R is *definable* in $\mathscr{L}$, if there exists a formula $\varphi(x_1, \ldots, x_k)$, such that:

$$\mathscr{M} \models \varphi[{}^{x_1 \cdots x_k}_{c_1 \cdots c_k}] \Leftrightarrow R(c_1, \ldots, c_k).$$

In this case, we say that φ is a *definition* of R in the variables $x_1, \ldots, x_k$.

As an example, let $\mathscr{L}_H$ be a first order language whose non-logical vocabulary consists of one 2-ary relation symbol $\mathbf{P}$ and two 1-ary relation symbols $\mathbf{M}$ and $\mathbf{F}$.

Consider a model $\mathscr{M}$ of $\mathscr{L}_H$ whose domain consists of all the human beings who have ever existed, past, present and future, and suppose:

- $\mathbf{P}^{\mathscr{M}}(x, y)$ iff x is a parent of y.
- $\mathbf{M}^{\mathscr{M}}(x)$ iff x is a male.
- $\mathbf{F}^{\mathscr{M}}(x)$ iff x is a female.

(for simplicity's sake, we assume that all objects in $|\mathscr{M}|$ satisfy exactly one of the two properties $\mathbf{M}^{\mathscr{M}}$ or $\mathbf{F}^{\mathscr{M}}$). We may now define other familial relations in $\mathscr{L}_H$. For example, the relation $G(x, y)$ which holds iff x is a grandparent of y can be defined (in $\mathbf{v}_1, \mathbf{v}_2$) by the formula:

$$\exists \mathbf{v}_3(\mathbf{P}(\mathbf{v}_1, \mathbf{v}_3) \wedge \mathbf{P}(\mathbf{v}_3, \mathbf{v}_2)).$$

Exercise 3.3. Provide definitions of the following properties and relations on $|\mathscr{M}|$:

(a). x is a brother of y.

(b). x is an aunt of y.

(c). x is a first cousin of y.

(d). x is a half sister of y.

(e). x and y have one son and one daughter.

(f). x has three children by two different men.

2. A Deductive System for First-order Logic

2.1. The Deductive System T_F. Given an arbitrary first-order language $\mathscr{L}$, we introduce a deductive system T_F for $\mathscr{L}$. As was the case in our discussion of sentential logic, the axioms and rules of inference for the system will be expressed schematically, however, in this case, we will allow the schematic variables to be instantiated not just by sentences of the language, but by any formulas of the language. Apart from the fact that non-sentential formulas may now appear in proofs in the system, the notions of proof, deductive consequence, etc., are defined just as before.

The first set of axiom schemas are the same as those of the deductive system for sentential logic:

(SL-1) $\varphi \to (\psi \to \varphi)$

(SL-2) $(\varphi \to (\psi \to \rho)) \to ((\varphi \to \psi) \to (\varphi \to \rho))$

(SL-3) $(\neg\varphi \to \neg\psi) \to (\psi \to \varphi)$

Here, φ, ψ and ρ are any formulas of $\mathscr{L}$. In addition T_F includes modus ponens as a rule of inference:

(MP) If $\Gamma \vdash \varphi$ and $\Gamma \vdash \varphi \to \psi$, then $\Gamma \vdash \psi$.

The fact that T_F contains all the axiom schemas of sentential logic and the rule MP, means that all deductively valid arguments in sentential logic, will likewise be deductively valid in first-order logic.

Definition 3.10. Let Φ be the set consisting of all atomic formulas of $\mathscr{L}$ as well as all universally quantified formulas of $\mathscr{L}$, and let $\varphi_1, \varphi_2, \ldots$ be an enumeration of Φ in which no term in the sequence is repeated. We assign to each formula φ of $\mathscr{L}$, a unique sentence $\mu(\varphi)$ of $\mathscr{L}_S$, as follows:

(1) If $\varphi = \varphi_i$, for some $i = 1, 2, \ldots$, then $\mu(\varphi) = \mathbf{P}_i$.

(2) $\mu(\neg\psi) = \neg\mu(\psi)$.

(3) $\mu(\psi \to \rho) = \mu(\psi) \to \mu(\rho)$.

We refer to $\mu(\varphi)$ as the *sentential form* of φ. If $\vdash_{T_S} \mu(\varphi)$, we refer to φ as a *tautology*.

Every formula of $\mathscr{L}$ has a unique sentential form, and conversely, every sentence of $\mathscr{L}_S$ is the sentential form of a unique formula of $\mathscr{L}$.

Theorem 3.3. *Let Γ be a set of formulas of $\mathscr{L}$, and let $\mu(\Gamma) = \{\mu(\varphi) : \varphi \in \Gamma\}$. If $\mu(\Gamma) \vdash_{T_S} \mu(\psi)$, then $\Gamma \vdash \psi$.*

Proof. Suppose that $\mu(\Gamma) \vdash_{T_S} \mu(\psi)$, and let $\mu(\varphi_1), \mu(\varphi_2) \ldots, \mu(\varphi_n)$ be a proof from $\mu(\Gamma)$ of $\mu(\varphi)$ in T_S. If $\mu(\varphi_i) \in \mu(\Gamma)$, then $\varphi \in \Gamma$. If $\mu(\varphi_i)$ is an instance of the axiom schema (SL-1) in T_S, then there exist formulas ψ, ρ of $\mathscr{L}$, such that $\mu(\varphi_i) = \mu(\psi) \to (\mu(\rho) \to \mu(\psi))$. But since $\mu(\psi) \to (\mu(\rho) \to \mu(\psi)) = \mu(\psi \to (\rho \to \psi))$, this means that φ_i is the formula $\psi \to (\rho \to \psi)$. Hence, φ_i is an instance of the axiom schema (SL-1) in T_F. More generally, if $\mu(\varphi_i)$ is an instance of any axiom schema of T_S, then φ_i is an instance of the corresponding axiom schema of T_F.

Now, suppose that $\mu(\varphi_i)$ is the result of applying MP to $\mu(\varphi_j)$ and $\mu(\varphi_k)$, for some $j, k < i$. Then, since $\mu(\varphi_k) = \mu(\varphi_j) \to \mu(\varphi_i)$, $\varphi_k = \varphi_j \to \varphi_i$, and so φ_i is the result of applying modus ponens to φ_j and φ_k. It thus follows that every formula in the sequence $\varphi_1, \varphi_2 \ldots, \varphi_n$ is either an axiom of T_F or else the result of applying modus ponens to two previous formulas in the sequence. Thus, $\varphi_1, \varphi_2, \ldots, \varphi_n$ is a proof Γ of φ in T_F. □

Corollary 3.4. *If φ is a tautology, then $\vdash \varphi$.*

In addition to the axioms of sentential logic and modus ponens, the deductive system T_F also includes axioms and rules which describe the operation of the universal quantifier. Among the latter is an axiom schema which allows for the 'instantiation' of a universal quantification over x by substituting for all free occurrences of x in the formula quanitified over a term in the language. In order to state this axiom schema, we must first make precise the notion of a (syntactic) substitution.

Definition 3.11. Let τ be any term, we define *the subtitution of t for x in τ* (written $\tau(x; t)$), inductively as follows:

(1) If τ is the variable x, then $\tau(x; t) = t$.

(2) If τ is an atomic term other than x, then $\tau(x; t) = \tau$.

(3) If τ is the term $\mathbf{f}(t_1, \ldots, t_k)$, then:

$$\tau(x; t) = \mathbf{f}(t_1(x; t), \ldots, t_k(x; t))$$

Let φ be any formula, we define *the substitution of t for x in φ* (written $\varphi(x; t)$), inductively as follows:

(1) If φ is the atomic formula $R(t_1, \ldots, t_k)$, then:

$$\varphi(x; t) = R(t_1(x; t), \ldots, t_k(x; t)).$$

(2) $(\neg\varphi)(x; t) = \neg(\varphi(x; t))$.

(3) $(\varphi \to \psi)(x; t) = \varphi(x; t) \to \psi(x; t)$.

(4) $(\forall x\varphi)(x; t) = \forall x\varphi$.

(5) If y is a variable other than x, then:

$$(\forall y\varphi)(x; t) = \forall y(\varphi(x; t)).$$

Proposition 3.5. *For any term τ and any assignment θ:*

$$\tau(x;t)[\theta] = \tau[\theta^x_{t[\theta]}]$$

Proof. The proof is by induction on τ. If τ is x, then since $\tau(x;t) = t$, the claim follows from the fact that:

$$x[\theta^x_{t[\theta]}] = t[\theta]$$

If t is an atomic term other than x, then since $\tau[\theta^x_{t[\theta]}] = \tau[\theta]$, the claim follows from the fact that:

$$\tau(x;t) = \tau$$

Suppose that τ is the term $\mathbf{f}(t_1,\ldots,t_k)$. Then, by the induction hypothesis:

$$\tau(x;t)[\theta] = \mathbf{f}(t_1(x;t)[\theta],\ldots,t_k(x;t)[\theta]) = \mathbf{f}(t_1[\theta^x_{t[\theta]}],\ldots,t_k[\theta^x_{t[\theta]}]) = \tau[\theta^x_{t[\theta]}].$$

□

It is often desirable to substitute multiple terms at once for the various variable occurring freely in a given formula. We might try to define the operation of multiple substitution, as follows: If $x_1,\ldots,x_n$ ($n > 1$) are distinct variables and $t_1,\ldots,t_n$ are terms, let $\mathrm{Sub}(\varphi, x_1,\ldots,x_n; t_1,\ldots,t_n)$ be the formula:

$$\mathrm{Sub}(\varphi, x_1,\ldots,x_{n-1}; t_1,\ldots,t_{n-1})(x_n; t_n).$$

This, however, will not do, for suppose that φ is the formula $\mathbf{R}(x,y)$, t_1 the term $\mathbf{f}(y)$ and t_2 the constant $\mathbf{c}$. Then while we would expect the formula $\mathrm{Sub}(\varphi, x, y; t_1, t_2)$ to be:

$$\mathbf{R}(\mathbf{f}(y),\mathbf{c}),$$

this substitution, in fact, yields:

$$\mathbf{R}(\mathbf{f}(\mathbf{c}),\mathbf{c}).$$

The reason for this is that, as defined above, the multiple substitutions of t_1 for x and t_2 for y occur *sequentially*. That is, t_1 is *first* substituted for x and *then* is t_2 substituted for y in the resulting formula. Consequently, since t_1 contains y, once t_1 has been substituted for x, the subsequent substitution of t_2 for y modifies this term. In order to obtain the desired effect of *simultaneous* substitution, we must first ensure that the substituted variables do not appear in any of the substituting terms.

Definition 3.12. Let $x_1,\ldots,x_n$ be distinct variables, and let $t_1,\ldots,t_n$ be terms. Choose any variables $v_1,\ldots,v_n$ such that, for all $i,j \in \{1,\ldots,n\}$, $v_i \neq x_j$ and v_i does not occur in t_j. We let $\varphi(x_1,\ldots,x_n; t_1,\ldots,t_n)$ be the formula:

$$\mathrm{Sub}(\mathrm{Sub}(\varphi, x_1,\ldots,x_n; v_1,\ldots,v_n), v_1,\ldots,v_n, t_1,\ldots,t_n)$$

(In other words, we first substitute (sequentially) for each variable x_i, the variable v_i, and then, in the resulting formula, we substitute (sequentially) for each variable v_i, the term t_i).

Henceforth, when we write $\varphi(x_1,\ldots,x_n)$ we will take this to imply that $x_1,\ldots,x_n$ are distinct variables. Once a formula has been introduced in this way, we write $\varphi(t_1,\ldots,t_n)$ for the formula $\varphi(x_1,\ldots,x_n; t_1,\ldots,t_n)$.

Definition 3.11 is meant to capture the intuitive idea that the formula $\varphi(x;t)$ should assert of t whatever it is that the formula φ asserts of x. To ensure that this is the case, however, we must take care, that in substituting t for x, we do not inadvertently introduce a variable which falls under the scope of some quantifier in φ.

Suppose, for example, that $\varphi(x)$ is the formula:

$$\forall y(\mathbf{R}(x,y) \to \mathbf{S}(y)).$$

In this case, the formula $\varphi(z)$ is

$$\forall y(\mathbf{R}(z,y) \to \mathbf{S}(y)),$$

which does indeed say of z just what $\varphi(x)$ says of x, namely, that any object in the domain which stands in the relation $\mathbf{R}$ to z, has the property $\mathbf{S}$. On the other hand, if we consider the formula $\varphi(y)$:

$$\forall y(\mathbf{R}(y,y) \to \mathbf{S}(y)),$$

we see that this does not assert the same of y, but rather says something altogether different, namely, that every object which stands in the relation **R** to *itself* has the property **S**. The problem here is that by substituting y for x in φ what was originally a free occurrence of x, becomes a bound occurrence of the variable y. To protect against the possibility of such inadvertent bindings, we introduce the notion of 'legitimate' substitution, or substitutability:

Definition 3.13. Let t be any term and let x be any variable. We define the class of formulas for which t is *substitutable for* x, inductively as follows:

(1) If φ is an atomic formula, then t is substitutable for x in φ.

(2) If t is substitutable for x in φ, then t is substitutable for x in $\neg\varphi$.

(3) If t is substitutable for x in both φ and ψ, then t is substitutable for x in $\varphi \to \psi$.

(4) t is subtitutable for x in the formula $\forall x\varphi$.

(5) If t is substitutable for x in φ and y does not occur in t, then t is substitutable for x in $\forall y\varphi$.

If t is a closed term, then it is substitutable for x in any formula φ. Also, if x does not occur freely in φ, then any term t is substitutable for x in φ (in this case, the substitution does not alter the original formula).

To show that the subtitution of subsitutable terms correctly captures the intended interpretation of syntactic substitution we observe the following fact:

Theorem 3.5. *If t is substitutable for x in φ, then*

$$\mathscr{M} \models \varphi(x;t)[\theta] \Leftrightarrow \mathscr{M} \models \varphi[\theta^{x}_{t[\theta]}]$$

Proof. The proof proceeds by induction on φ. For atomic formulas, t is always substitutable for x, and the result follows from Proposition 3.5. Suppose that t is substitutable for x in $\neg\varphi$. Then, t is substitutable for x in φ, and so by the induction hypothesis:

$$\mathscr{M} \models \neg\varphi(x;t)[\theta] \Leftrightarrow \mathscr{M} \not\models \varphi(x;t)[\theta] \Leftrightarrow \mathscr{M} \not\models \varphi[\theta^{x}_{t[\theta]}] \Leftrightarrow \mathscr{M} \models \neg\varphi[\theta^{x}_{t[\theta]}]$$

A similar argument can be given for conditional claims.

Now, suppose that φ is the formula $\forall x\psi$. Since x does not occur freely in ψ, for any term t, $\varphi(x;t)$ is the formula φ and $\mathscr{M} \models \varphi[\theta]$ iff $\mathscr{M} \models \varphi[\theta^{x}_{t[\theta]}]$.

Lastly, suppose that φ is the formula $\forall y\psi$, where y is some variable distinct from x, and that t is substitutable for x in φ. Then t is substitutable for x in ψ and y does not occur in t. Hence, for any assignment θ and any $c \in |\mathscr{M}|$, $t[\theta^{y}_{c}] = t[\theta]$, and so, by the induction hypothesis:

$$\begin{gathered}
\mathscr{M} \models \forall y\psi(x;t)[\theta] \\
\Updownarrow \\
\text{For every } c \in |\mathscr{M}|,\ \mathscr{M} \models \psi(x;t)[\theta^{y}_{c}] \\
\Updownarrow \\
\text{For every } c \in |\mathscr{M}|,\ \mathscr{M} \models \psi[\theta^{y\,x}_{c\,t[\theta^{y}_{c}]}] \\
\Updownarrow \\
\text{For every } c \in |\mathscr{M}|,\ \mathscr{M} \models \psi[\theta^{y\,x}_{c\,t[\theta]}] \\
\Updownarrow \\
\text{For every } c \in |\mathscr{M}|,\ \mathscr{M} \models \psi[\theta^{x\quad y}_{t[\theta]\,c}] \\
\Updownarrow \\
\mathscr{M} \models \forall y\psi[\theta^{x}_{t[\theta]}]
\end{gathered}$$

This completes the proof. □

We now are at last in a position to state the two axioms schemas in T_F that describe the behavior of the universal quantifier:

(Q) $(\varphi \to \psi) \to (\varphi \to \forall x\psi)$, where x does not occur freely in φ.

(UI) $\forall x\varphi \to \varphi(x;t)$, where t is substitutable for x in φ.

In addition to these two axiom schemas, the system T_F also includes a second rules of inference, which we refer to as universal generalization (UG):

(UG) If x does not occur freely in any formula in Γ, then if $\Gamma \vdash \varphi$, $\Gamma \vdash \forall x\varphi$.

2.2. The Soundness Theorem for T_F. The soundness theorem for T_F asserts that:

Theorem 3.6. *If $\Gamma \vdash \psi$, then $\Gamma \models \psi$.*

Proof. As before, we prove soundness by induction on the set of formulas which are provable from Γ. In particular, we must show that:

(1) If $\varphi \in \Gamma$, then $\Gamma \models \varphi$.

(2) If φ is an axiom of T_F, then $\Gamma \models \varphi$.

(3) If $\Gamma \models \varphi$ and $\Gamma \models \varphi \to \psi$, then $\Gamma \models \psi$.

(4) If x does not occur freely in any formula in Γ, and if $\Gamma \models \varphi$, then $\Gamma \models \forall x\varphi$.

Claim (1) follows immediately from the definition of $\models$.

For claim (3), suppose that $\mathscr{M} \models \varphi[\theta]$ and $\mathscr{M} \models \varphi \to \psi[\theta]$. Then, it follows from the definition of satisfaction that $\mathscr{M} \models \psi[\theta]$. Hence, $\{\varphi, \varphi \to \psi\} \models \psi$, which implies (3).

For claim (4), suppose that $\mathscr{M} \models \psi[\theta]$, for all $\psi \in \Gamma$. Since x does not occur freely in any $\psi \in \Gamma$, for any $c \in |\mathscr{M}|$, $\mathscr{M} \models \psi[\theta^x_c]$, for all $\psi \in \Gamma$. But, since $\Gamma \models \varphi$, it follows that $\mathscr{M} \models \varphi[\theta^x_c]$, for all $c \in |\mathscr{M}|$, i.e., $\mathscr{M} \models \forall x\varphi[\theta]$.

For claim (2), we will show that if φ is an axiom, then $\mathscr{M} \models \varphi[\theta]$, for any model $\mathscr{M}$ and assignment θ. This is obviously so for instances of the sentential axioms.

For UQ, suppose that $\mathscr{M} \models \forall x(\varphi \to \psi)[\theta]$ and $\mathscr{M} \models \varphi[\theta]$. The first implies that for any $c \in |\mathscr{M}|$, $\mathscr{M} \models (\varphi \to \psi)[\theta^x_c]$ and the second implies that $\mathscr{M} \models \varphi[\theta^x_c]$ since x does not occur freely in φ. It thus follows, that $\mathscr{M} \models \psi[\theta^x_c]$, for all $c \in |\mathscr{M}|$, i.e., $\mathscr{M} \models \forall x\psi[\theta]$.

For UI, suppose that $\mathscr{M} \models \forall x\varphi[\theta]$ and that t is substitutable for x in φ. Then, $\mathscr{M} \models \varphi[\theta^x_{t[\theta]}]$, and so $\mathscr{M} \models \varphi(x;t)[\theta]$. This completes the proof. □

2.3. Some Basic Facts about T_F. In this section, we prove various elementary facts about the system T_F which we will appeal to in our proof of the completeness of the system. We begin by proving that the deduction theorem (for sentences) holds for first-order logic.

Theorem 3.7 (Deduction Theorem). *Let φ be any sentence of $\mathscr{L}$. Then, $\Gamma \cup \{\varphi\} \vdash \psi$ if and only if $\Gamma \vdash \varphi \to \psi$.*

Proof. The proof proceeds just as in the case of sentential logic. The if part of the claim follows trivially from MP. To prove the only if part of the claim, suppose that $\psi_1, \ldots, \psi_k$ is a proof of ψ from $\Gamma \cup \{\varphi\}$. To show that $\Gamma \vdash \varphi \to \psi$ it will suffice to show that, for all $i = 1, \ldots, k$:

$$\Gamma \cup \{\varphi \to \psi_1, \ldots, \varphi \to \psi_{i-1}\} \vdash \varphi \to \psi_i \tag{2.1}$$

We prove this claim by induction on i. We first note that ψ_i is either (i) a member of Γ, (ii) equal to φ, (iii) an axiom of T_F, (iv) the result of applying MP to two previous lines in the proof, or (v) the result of applying UG to some previous line in the proof.

The first four cases are handled in exactly the same manner as was the case for the deduction theorem in sentential logic. For (v), suppose that ψ_i is the formula $\forall x\psi_j$ $(j < i)$, where x does not occur freely in any formula in Γ. Then, by the induction hypothesis:

$$\Gamma \cup \{\varphi \to \psi_1, \ldots, \varphi \to \psi_{i-1}\} \vdash \varphi \to \psi_j$$

Hence, it will suffice to prove (2.1) to show that

$$\varphi \to \psi_j \vdash \varphi \to \forall x \psi_j$$

Since φ is a sentence and x does not occur freely in ψ_j, x does not occur freely in $\varphi \to \psi_j$. We thus have the following proof:

(1) $\varphi \to \psi_j$ [Premise]

(2) $\forall x(\varphi \to \psi_j)$ [UG-1]

(3) $\forall x(\varphi \to \psi_j) \to (\varphi \to \forall x \psi_j)$ [UQ]

(4) $\varphi \to \forall x \psi_j$ [MP - 2,3]

This completes the proof. □

As an example of how the deduction theorem can be applied in first-order logic, we may note the following claim:

Proposition 3.6. $\vdash \forall x(\varphi(x) \to \psi(x)) \to (\forall x \varphi(x) \to \forall x \psi(x))$

Proof. We will first show that $\{\forall x(\varphi(x) \to \psi(x)), \forall x \varphi(x)\} \vdash \forall x \psi(x)$. The proof is as follows:

(1) $\forall x(\varphi(x) \to \psi(x))$ [Premise]

(2) $\forall x(\varphi(x) \to \psi(x)) \to (\varphi(x) \to \psi(x))$ [UI]

(3) $\varphi(x) \to \psi(x)$ [MP - 1,2]

(4) $\forall x \varphi(x)$ [Premise]

(5) $\forall x \varphi(x) \to \varphi(x)$ [UI]

(6) $\varphi(x)$ [MP - 4,5]

(7) $\psi(x)$ [MP - 3,6]

(8) $\forall x \psi(x)$ [UG - 7]

(note that the application of UG in step 7 of the proof relies on the fact that x does not occur freely in either of the premises). The desired result follows from two applications of the deduction theorem. □

Proposition 3.7. *Let Γ be a set of formulas,* $\mathbf{c}$ *a constant symbol, and let $\varphi(x)$ be a formula with one free variable. If* $\mathbf{c}$ *does not occur in any formula of Γ, then:*

$$\Gamma \vdash \varphi(\mathbf{c}) \Rightarrow \Gamma \vdash \forall x \varphi(x)$$

Proof. Let $\psi_1, \ldots, \psi_k$ be a proof of $\varphi(\mathbf{c})$ from Γ, and let x be a variable which does not occur in any formula Γ, or in any formula in this proof. Let ψ_i' be the formula which results from replacing every occurrence of $\mathbf{c}$ in ψ_i with x. Then, it easy to verify that $\psi_1', \ldots, \psi_k'$ is a proof of $\varphi(x)$ from Γ. Thus, from UG, $\Gamma \vdash \forall x \varphi(x)$. □

2.4. The Completeness Theorem for T_F. The completeness theorem for T_F asserts:

Theorem 3.8. *Let Γ be a set of sentences of $\mathscr{L}$, and let ψ be a sentence of $\mathscr{L}$. Then, if $\Gamma \models \psi$, then $\Gamma \vdash \psi$.*

(Note that while we we will prove completeness only for sentences of the language, this is not because the result does not hold for all formulas (it does). It is rather so as to avoid certain complications of a technical nature.)

As was the case in sentential logic, Theorem 3.8 can be reformulated in terms of consistency and satisfiability:

DEFINITION 3.14. A set of sentences Γ is *satisfiable* if there exists a model $\mathscr{M}$, such that $\mathscr{M} \models \varphi$, for all $\varphi \in \Gamma$.

DEFINITION 3.15. A set of sentences Γ is *consistent* if there exists a formula ψ, such that $\Gamma \vdash \psi$, and $\Gamma \vdash \neg\psi$.

The proofs of the following propositions are exactly the same as those which were given for sentential logic:

PROPOSITION 3.8. $\Gamma \not\models \psi$ *iff* $\Gamma \cup \{\neg\psi\}$ *is satisfiable.*

PROPOSITION 3.9. $\Gamma \nvdash \psi$ *iff* $\Gamma \cup \{\neg\psi\}$ *is consistent.*

PROPOSITION 3.10. *If* Γ *is a consistent set of sentences, then for any sentence* ψ*, either* $\Gamma \cup \{\psi\}$ *or* $\Gamma \cup \{\neg\psi\}$ *is consistent.*

The completeness theorem is thus equivalent to the following claim:

THEOREM 3.9 (Completeness Theorem). *Every consistent set of sentences is satisfiable.*

DEFINITION 3.16. Let $\mathscr{M}$ be a model for $\mathscr{L}$ and let $\mathscr{L}^+$ be any first-order language obtained from $\mathscr{L}$ by adding to the non-logical vocabulary of $\mathscr{L}$ additional constant symbols, function symbols or relation symbols. If $\mathscr{M}$ is a model for $\mathscr{L}^+$, then the *restriction of* $\mathscr{M}$ *to* $\mathscr{L}$ (written $\mathscr{M}_{\mathscr{L}}$), is the model of $\mathscr{L}$, defined by:

(i) $|\mathscr{M}_{\mathscr{L}}| = |\mathscr{M}|$

(ii) $\mathbf{c}^{\mathscr{M}_{\mathscr{L}}} = \mathbf{c}^{\mathscr{M}}$, for any constant symbol $\mathbf{c}$ in $\mathscr{L}$.

(iii) $\mathbf{f}^{\mathscr{M}_{\mathscr{L}}} = \mathbf{f}^{\mathscr{M}}$, for any function symbol $\mathbf{f}$ in $\mathscr{L}$.

(iv) $\mathbf{R}^{\mathscr{M}_{\mathscr{L}}} = \mathbf{R}^{\mathscr{M}}$, for any relation symbol $\mathbf{R}$ in $\mathscr{L}$.

If φ is a sentence of $\mathscr{L}$, $\mathscr{L}^+$ an extension of $\mathscr{L}$, and $\mathscr{M}$ a model for $\mathscr{L}^+$, then:

$$\mathscr{M} \models \varphi \Leftrightarrow \mathscr{M}_{\mathscr{L}} \models \varphi$$

Thus, in proving Theorem 3.9, it is enough to show that every consistent set of sentences $\mathscr{L}$ has a model for some extension of the language.

In outline, our proof of Theorem 3.9 will proceed along much the same lines as did our proof of the completeness of sentential logic: we will first show that every maximally consistent set of sentences (satisfying a given property) has a model and we will then show that every consistent set of sentences can be extended to a maximally consistent set (that satisfies this property). The definition of maximally consistent sets, and the basic properties of such sets are the same as before. We restate these facts here for the sake of completeness:

DEFINITION 3.17. A set of sentences Γ is *maximally consistent*, if the following two conditions hold:

(1) Γ is consistent.

(2) Every set Γ' of which Γ is a proper subset is inconsistent.

PROPOSITION 3.11. *If* Γ *is a maximally consistent set of sentences, then for every sentence* φ*, exactly one of the two sentences* φ*,* $\neg\varphi$ *is a member of* Γ.

The advantage of considering maximally consistent sets is that these sets are 'sufficiently informative' to allow us to read off from the sentences in the set an interpretation of the language, under which all of the sentences in the set are true. In particular, in the context of first-order logic, consideration of a maximally consistent set should allow for the construction of a model of the language which satisfies every sentence in this set.

What objects should be included in the domain of this model? A particularly convenient way of answering this qustion is to allow the terms in the language to themselves comprise the model's domain. By adopting this approach, we can consruct the model under the assumption that it interprets the terms of the language *autonomically*, i.e., as names of themselves.

Definition 3.18. Let Γ be a maximally consistent set of sentences in a first-order language $\mathscr{L}$ which contains at least one constant symbol (this assumption is needed in order to ensure that the domain of the model is nonempty). The *standard model* for Γ is the model $\mathscr{M}$ for $\mathscr{L}$ defined by:

(1) $|\mathscr{M}| = \{t : t \text{ is a closed term of } \mathscr{L}\}$

(2) For every constant symbol $\mathbf{c}$, $\mathbf{c}^{\mathscr{M}} = \mathbf{c}$.

(3) For every k-ary function symbol $\mathbf{f}$:

$$\mathbf{f}^{\mathscr{M}}(t_1, \ldots, t_k) = \mathbf{f}(t_1, \ldots, t_k),$$

where $t_1, \ldots, t_k \in |\mathscr{M}|$.

(4) For every k-ary relation symbol $\mathbf{R}$:

$$\mathbf{R}^{\mathscr{M}}(t_1, \ldots, t_k) \Leftrightarrow \mathbf{R}(t_1, \ldots, t_k) \in \Gamma.$$

At this point, it would be natural to attempt to prove the following claim:

Proposition 3.12. *Let Γ be a maximally consistent set of sentences in a first-order language $\mathscr{L}$ which contains at least one constant symbol, and let $\mathscr{M}$ be the standard model for Γ. Then for every sentence φ of $\mathscr{L}$:*

$$\varphi \in \Gamma \Leftrightarrow \mathscr{M} \models \varphi$$

Unfortunately, this claim does not hold in general. To see the difficulty note that the above claim implies that for any formula $\varphi(x)$:

$$\exists x\varphi(x) \in \Gamma \Leftrightarrow \mathscr{M} \not\models \forall x \neg\varphi(x)$$

But since the domain of $\mathscr{M}$ consists of the closed terms of the language, the right-hand side of this conditional is true iff for some closed term t of $\mathscr{L}$, $\mathscr{M} \not\models \neg\varphi(t)$, or equivalently, that $\mathscr{M} \models \varphi(t)$. Hence the above claim implies that for any maximally consistent set of sentences Γ:

$$\exists x\varphi(x) \in \Gamma$$
$$\Updownarrow$$
$$\varphi(t) \in \Gamma, \text{ for some closed term } t.$$

This, however, need not be the case. It is possible for a maximally consistent set to include both the sentence $\exists x\varphi(x)$, and all sentences of the form $\neg\varphi(t)$ since the object whose existence is asserted by the existential claim need not named by any closed term of the language. By considering the standard model of Γ we exclude this possibility since the objects in the domain are all names of themselves.

In order to overcome this difficulty, we must limit our attention to maximally consistent sets which imply that for every true existential claim $\exists x\varphi(x)$, there is a term t which can serve as a 'witness' to this claim, in the sense that the object referred to by t has the property φ.

Definition 3.19. Let $\mathscr{L}$ be a first-order language which contains at least one constant symbol. A set of sentences Γ is *witnessed* if, for every formula $\varphi(x)$ with one free variable, there is a closed term t, such that:

$$\exists x\varphi(x) \to \varphi(t) \in \Gamma$$

Theorem 3.10. *Let Γ be a witnessed, maximally consistent set of sentences in a first-order language $\mathscr{L}$ which contains at least one constant symbol, and let $\mathscr{M}$ be the standard model for Γ. Then for every sentence φ of $\mathscr{L}$:*

$$\varphi \in \Gamma \Leftrightarrow \mathscr{M} \models \varphi$$

Proof. The proof proceeds by induction on φ. If φ is an atomic sentence, then this follows directly from the definition of the standard model $\mathscr{M}$.

Suppose that $\varphi = \neg\psi$. If $\varphi \in \Gamma$, then $\psi \notin \Gamma$, and so, by hypothesis, $\mathscr{M} \not\models \psi$. Hence, $\mathscr{M} \models \varphi$. Now suppose that $\varphi \notin \Gamma$. Then $\psi \in \Gamma$, and so by hypothesis $\mathscr{M} \models \psi$. Hence $\mathscr{M} \not\models \varphi$.

Now, suppose that $\varphi = \psi \to \rho$ and that $\varphi \in \Gamma$. Then either $\psi \notin \Gamma$ or $\rho \in \Gamma$. If $\psi \notin \Gamma$, then by hypothesis, $\mathscr{M} \not\models \psi$. Hence, $\mathscr{M} \models \varphi$. If, on the other hand, $\rho \in \Gamma$, then, by hypothesis, $\mathscr{M} \models \rho$. Hence, $\mathscr{M} \models \varphi$. Thus, if $\varphi \in \Gamma$, $\mathscr{M} \models \varphi$. Now suppose that $\varphi \notin \Gamma$, then both $\psi \in \Gamma$ and $\rho \notin \Gamma$. But then, by hypothesis, $\mathscr{M} \models \psi$ and $\mathscr{M} \not\models \rho$. Hence $\mathscr{M} \not\models \varphi$.

Now, suppose that $\varphi = \forall x\psi(x)$. It will suffice to show that if the claim holds for all sentences of the form $\psi(t)$, where t is a closed term of $\mathscr{L}$, then it holds for φ. Suppose that $\varphi \in \Gamma$ and let t be a closed term of $\mathscr{L}$. Then, since every sentence which is an axiom of T_F is in Γ, $\varphi \to \psi(t) \in \Gamma$, and so $\psi(t) \in \Gamma$. Thus, by assumption $\mathscr{M} \models \psi(t)$, for all closed terms of $\mathscr{L}$, i.e., $\mathscr{M} \models \varphi$.

Now suppose that $\varphi \notin \Gamma$. Then $\exists x\neg\psi(x) \in \Gamma$. Since Γ is witnessed, for some closed term t, the sentence:

$$\exists x\neg\psi(x) \to \neg\psi(t)$$

and so $\neg\psi(t) \in \Gamma$. Thus, by hypothesis, $\mathscr{M} \models \neg\psi(t)$, from which it follows that $\mathscr{M} \not\models \forall x\psi(x)$, i.e., $\mathscr{M} \not\models \varphi$. □

To prove completeness, all that remains to be shown is that every consistent set of sentences can be extended to a witnessed, maximally consistent set. The construction of a maximally consistent extension of a consistent set of sentences, can proceed in exactly the same manner as it did in the context of sentential logic, *viz.*, the sentences of the language are enumerated, and at each stage of the construction, we add to the set either the sentence or its negation, in a manner which preserves consistency.

All that we must show then is that every consistent set can be extended to a consistent, witnessed set of sentences.

Theorem 3.11. *Let Γ be a consistent set of sentences of a first-order language $\mathscr{L}$. Then there exists a consistent, witnessed set of sentences Γ' of a first-order language $\mathscr{L}'$, which is an extension of $\mathscr{L}$, such that $\Gamma' \supset \Gamma$.*

Proof. Let $\mathscr{L}'$ be the first-order language which results from adding to $\mathscr{L}$ an infinite sequence of new constant symbols:

$$\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \ldots,$$

and let $\varphi_1, \varphi_2, \ldots,$ be an enumeration of all formulas of $\mathscr{L}'$, in which exactly one variable occurs freely. We construct an increasing sequence of sets of sentences of $\mathscr{L}'$, $\Gamma_0, \Gamma_1, \ldots$, as follows: Let $\Gamma_0 = \Gamma$, and let:

$$\Gamma_1 = \Gamma_0 \cup \{\exists x\varphi_1(x) \to \varphi_1(\mathbf{c}_1)\}$$

Suppose that Γ_n ($n \geq 1$) has been constructed, and that it consists of the union of Γ_0 and finitely many sentences in Γ_n. Then, since no constant of the form $\mathbf{c}_i$ appears in Γ_0, only finitely many constants $\mathbf{c}_i$ appear in Γ_n. Let j_{n+1} be the smallest j, such that $\mathbf{c}_j$ does not occur in any sentence in Γ_n, and put:

$$\Gamma_{n+1} = \Gamma_n \cup \{\exists x\varphi_{n+1}(x) \to \varphi_{n+1}(\mathbf{c}_{j_{n+1}})\}$$

If Γ_n is the union of Γ_0 and finitely many sentences in $\mathscr{L}'$, then so is Γ_{n+1}. Thus the construction can be continued ad infinitum.

Let Γ' be the union of the sets $\Gamma_0, \Gamma_1, \Gamma_2, \ldots$, i.e.:

$$\Gamma' = \bigcup_{n}^{\infty} \Gamma_n$$

Clearly Γ' is a witnessed set of sentences of $\mathscr{L}'$. It remains to show that Γ' is consistent. For this, it suffices to show that Γ_n is consistent, for all n. For $n = 0$, this follows from the fact that $\Gamma_0 = \Gamma$. Now suppose that Γ_{n-1} is consistent ($n \geq 1$), but that Γ_n is inconsistent. Since $\Gamma_n = \Gamma_{n-1} \cup \{\exists x \varphi_n(x) \rightarrow \varphi_n(\mathbf{c}_{j_n})\}$, this means that:

$$\Gamma_{n-1} \vdash \neg(\exists x \varphi_n(x) \rightarrow \varphi_n(\mathbf{c}_{j_n}))$$

By sentential logic, it follows that

$$\Gamma_{n-1} \vdash \exists x \varphi_n(x)$$

and

$$\Gamma_{n-1} \vdash \neg \varphi_n(\mathbf{c}_{j_n})$$

By Proposition 3.7, this means that $\Gamma_{n-1} \vdash \forall x \neg \varphi_n(x)$. But since $\exists x \varphi_n(x)$ is the formula $\neg \forall x \neg \varphi_n(x)$, Γ_{n-1} is inconsistent, which contradicts our assumption. □

The proof of Theorem 3.9 can now be stated as follows: Let Γ be a consistent set of sentences of $\mathscr{L}$. Then, by Theorem 3.11, there exists a set of sentences Γ' of some extended language $\mathscr{L}'$, such that $\Gamma \subset \Gamma'$ and Γ' is a witnessed, consistent set of sentences of $\mathscr{L}'$. We may then extend Γ' to a witnessed, maximally consistent set of sentences Γ^* of $\mathscr{L}'$, which, by Theorem 3.10 is satisfied by the standard model for Γ^*, $\mathscr{M}$. It follows that $\mathscr{M}$ is a model of Γ in the language $\mathscr{L}'$, and so $\mathscr{M}|\mathscr{L}$ is a model of Γ for $\mathscr{L}$.

Appendix A: General Notation

In this course, mathematical theorems and their proofs will be expressed in ordinary (if somewhat stilted) English supplemented by certain terminological devices, the precise meanings of which will be specified in definitions that are highlighted and numbered in the course notes for ease of reference. The terminological devices that are given explicit definitions in the notes will typically be introduced either for the purpose of formulating particular theorems clearly and concisely, or else as an aid in articulating the reasoning involved in the proofs of such theorems. As such, while any piece of notation may be freely utilized once it has been explicitly defined, the usefulness of such notations will often be limited to a specific subject matter, so that once that subject matter has been done with, the precise meaning of the notation may be, more or less, forgotten.

There is, however, a certain class of terminology which appears so often in mathematical writing that it is best regarded as a fundamental part of the language in which mathematics is done. We take this opportunity now to introduce this terminology explicitly for those who may not yet have had occassion to encounter it. Readers who fall into the latter category would do well to master its usage now since it will henceforth be utilized, without further comment, in nearly every theorem or proof that we write.

If and only if. The phrase "iff" is shorthand for the locution "if and only if". It expresses an equivalence between the claim which precedes it and the claim which follows it. To prove that φ iff ψ, one must prove both that φ is true only if ψ is true (i.e., that φ implies ψ) and also that φ is true if ψ is true (i.e., that ψ implies φ).[2]

The Ellipsis. The ellipsis ('$\cdots$') is used to to indicate an implied interpolation of the pattern indicated by that which precedes and succeeds it. Thus, for example, the expression

$$1, 2, \ldots, 100$$

refers to the list of numbers beginning with 1 and ending with 100, where each number in the list (except for the first) is 1 greater than that which immediately precedes it. If the ellipsis has no succeeding term, then the pattern implied by the terms preceding the ellipsis is understood to be continued *ad infinitum*. Thus, the expression:

$$\varphi_1, \varphi_2, \ldots$$

refers to an infinite list, the nth term of which is φ_n.

While it is most commonly used in expressions denoting lists of objects, an ellipsis may be used in expressions denoting any type of entity whatsoever, provided the denoting phrase involves some interpolatible pattern. Thus, for example, if $\varphi_1, \varphi_2, \ldots, \varphi_k$ are all sentences in some formal language, then we may express their conjunction (which we will assume is also a sentence in the language) as follows:

$$\varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_k.$$

Unless otherwise indicated, the pattern implied by the ellipsis should always be assumed to be the incrementation (i.e., addition by 1) of the most obvious index of the terms preceding the ellipsis (which may be the terms themselves). This is so even when there is some other pattern that may seem equally obvious. Thus, for example:

$$x_3, \ldots, x_{11}$$

---

[2] In the context of a definition, in spite of the typical usage of an "if" locution to specify the defining condition, it will always be assumed that the stated condition is both a necessary and sufficient for the defined term to apply. Thus, for example, the "if" appearing in the following definition:

DEFINITION. A natural number is *prime* if it is both greater than 1 and divisible by only 1 and itself.

is to be read as "iff".

refers to the list:

$$x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}$$

and not to the list:

$$x_3, x_5, x_7, x_9, x_{11}.$$

If the pattern to be interpolated does not simply involve the incrementation of an index, then it may be explicated by placing between two ellipses a functional expression indicating how any given elliptical term is to be inferred from its position in the list (as represented by some variable, e.g., i or n). Thus, for example, to express the list above, one may write:

$$x_3, \ldots, x_{2n+1}, \ldots, x_{11}$$

It will be notationally convenient to stipulate that an expression such as

$$\varphi_1, \varphi_2, \ldots, \varphi_0,$$

in which the term succeeding the ellipsis does not appear in the continuation implied by the terms preceding it, refers to the empty set (see below).

Basic Set Terminology. A set of objects can be represented by a comma-separated list of those objects placed between curly braces (i.e., '{' and '}'). Thus, for example, the set of all odd numbers between 1 and 10 may be written:

$$\{1, 3, 5, 7, 9\}$$

To say that an object x belongs to a set X, we write $x \in X$ (read: "x is in X" or "x is a member of X"). To say that x does not belong to X, we write $x \notin X$. Hence,

$$5 \in \{1, 3, 5, 7, 9\},$$

but

$$4 \notin \{1, 3, 5, 7, 9\}.$$

to express the fact that all of the objects appearing in the list $x_1, \ldots, x_k$ are members of X, we write $x_1, \ldots, x_k \in X$, and to express the fact that none of these objects are members of X, we write $x_1, \ldots, x_k \notin X$.

If the members of a set can be easily listed, say by the list $x_1, \ldots, x_k$, then we often substitute the expression "$= x_1, \ldots, x_k$" for the expression "$\in \{x_1, \ldots, x_k\}$." Thus, for example, it is more standard usage to write "for all $n = 1, 2, \ldots$, $2n + 1$ is odd," than it is to write "for all $n \in \{1, 2, \ldots\}$, $2n + 1$ is odd."

We say that one set X is a subset of another set Y (or that X is included in Y), if every member of X is a member of Y, i.e., if, for any x, $x \in X$ implies $x \in Y$. To express this fact, we write $X \subset Y$ (or, less commonly, $Y \supset X$). So, for example:

$$\{1, 5, 7\} \subset \{1, 3, 5, 7, 9\}.$$

Note that, according to the above definition, every set is a subset of itself, i.e., for any set X, $X \subset X$.

One will often encounter multiple statements of set inclusion "chained" together, as in:

$$X_1 \subset X_2 \subset \cdots \subset X_k.$$

This means that $X_1 \subset X_2$, $X_2 \subset X_3$, etc.. Since, however, the subset relation is a transitive relation (i.e., for any sets X, Y, Z, if $X \subset Y$ and $Y \subset Z$, then $X \subset Z$), these statements of inclusion are together equivalent to the claim that $X_i \subset X_j$, whenever $1 \leq i \leq j \leq k$.

We say that the two sets X and Y are equal (written: $X = Y$) if X and Y have exactly the same members. Since two sets are equal just in case each is a subset of the other, one often proves the equality of two sets X and Y, by first showing that $X \subset Y$ and then showing that $Y \subset X$. If $X \subset Y$, but $X \neq Y$, we say that X is a proper subset of Y.

The Empty Set. It will be convenient to assume that there exists a set containing no members, which we denote by $\emptyset$, and refer to as the empty set. The two defining properties of the empty set are as follows:

(1) For every property P, it is (vacuously) true that, for all $x \in \emptyset$, x has the property P; and

(2) For every property P, it is (vacuously) false that, for some $x \in \emptyset$, x has the property P.

Note that from (b), it follows that no object is a member of $\emptyset$ (for any object a, let P be the property possessed by all and only those x, such that $x = a$), and from (a) it follows that, for any set X, $\emptyset \subset X$ (let P be the property possessed by all and only those x, such that $x \in X$).

Representing Sets. In mathematics, nearly every set that we will have any occassion to reference will be most clearly defined, not by means of a simple listing of its members, but rather in terms of a property possessed by all and only the members of this set.[3] Thus, for instance, we may have occassion to refer to the set of all prime numbers, or the set of all expressions (in a formal language) of length greater than 12.

To refer directly to the set of all and only those objects which possess the property P, we write:

$$\{x : P(x)\},$$

where $P(x)$ stands for any statement expressing the fact that x has the property P. This description of the set is read: “the set of all x, such that $P(x)$.” So, for example, if we wish to refer to the set of all odd, natural numbers between 1 and 10, we could write:

$$\{n : n \text{ is an odd natural number and } 1 \leq n \leq 10\}$$

When the defining property of a set implies that the set in question is a subset of some other set Y, this fact can be indicated by writing $x \in Y$ on the left-hand side of the colon. The condition then appearing on the right-hand side of the colon is that which must hold of any member of Y in order for it to be included in the set. So, for example, if N is the set of all natural numbers, then the above set may be written:

$$\{n \in N : n \text{ is odd and } 1 \leq n \leq 10\}$$

If it is clear from the context, from what set of objects the members of a defined set are to be drawn, then we may omit any explicit reference to the containing set. Thus, for example, if it is clear from the context that we are defining a set of natural numbers, then, for the above set, we may simply write:

$$\{n : n \text{ is odd and } 1 \leq n \leq 10\}$$

Sometimes it is convenient to describe a set as that which results from applying a certain function to all only the members of another set. In particular, to refer to the set which results from applying the function f to the members of the set consisting of all and only the objects possessing the property P, we write:

$$\{f(x) : P(x)\}.$$

Hence, the above set may be written:

$$\{2n + 1 : 0 \leq n \leq 4\}.$$

Set Operations. The union of the two sets X and Y is defined by:

$$X \cup Y = \{x : x \in X \text{ or } x \in Y\}.$$

The intersection of X and Y is defined by:

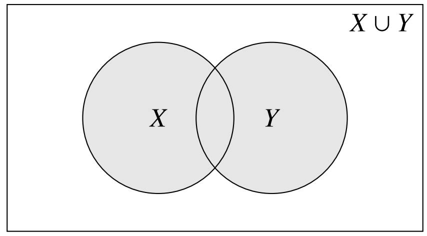
$$X \cap Y = \{x : x \in X \text{ and } x \in Y\}.$$

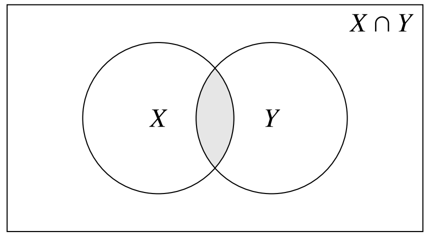
And the complement of Y (relative to X) is defined by:

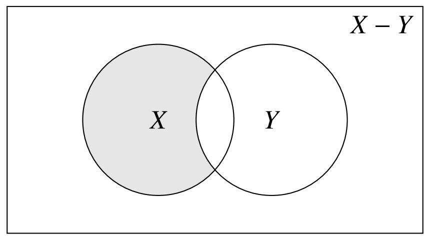
$$X - Y = \{x \in X : x \notin Y\}$$

The following Venn diagrams can help to visualize the effect of performing these set-operations:

[3] Indeed, as we shall later see, there are some sets whose members are too numerous to be listed.







Since the result of forming the union of multiple sets does not depend on the order in which the union operations are performed, we may simply write:

$$X_1 \cup \cdots \cup X_k,$$

in place of any expression of the form

$$(\cdots((X_{n_1} \cup X_{n_2}) \cdots X_{n_k}),$$

where $n_i \in \{1, \ldots, k\}$, for $i = 1, \ldots, k$, and $n_i \neq n_j$, for $i \neq j$ (and similarly for the intersection of multiple sets). More generally, if $\mathscr{X}$ is a set whose members are all sets, we define:

$$\bigcup \mathscr{X} = \{x : x \in X, \text{ for some } X \in \mathscr{X}\},$$

and

$$\bigcap \mathscr{X} = \{x : x \in X, \text{ for all } X \in \mathscr{X}\}.$$

If $\mathscr{X} = \{X_1, \ldots, X_k\}$, we will often write:

$$\bigcup_{i=1}^{k} X_i$$

in place of $\bigcup \mathscr{X}$, and similarly for intersections (this notation can be extended in the obvious way to any set $\mathscr{X}$ whose members comprise an indexed list of sets).

Solutions to Selected Exercises

Chapter 1

1.1 (a) For any expression φ, let $l(\varphi)$ be the length of φ. Let ψ be any sentence. Then ψ is of the form $\alpha = \beta$, where α and β are numerical expressions, and so

$$l(\psi) = l(\alpha) + l(=) + l(\beta) = 1 + l(\alpha) + l(\beta).$$

If both $l(\alpha)$ and $l(\beta)$ are odd, $1 + l(\alpha) + l(\beta)$ is odd. Hence, it will suffice to show that $l(\psi)$ is odd, to show that the length of every numerical expression is odd.

For any numerical expression α, let $c(\alpha)$ be the number of times the symbol $\times$ occurs in α. We will prove that the length of α is odd by induction on $c(\alpha)$. If $c(\alpha) = 0$, then α is a numeral, and so $l(\alpha) = 1$, which is odd.

Now suppose that $c(\alpha) \geq 1$ and that, for all numerical expressions β, such that $c(\beta) < c(\alpha)$, $l(\beta)$ is odd. Since $c(\alpha) \geq 1$, α is of the form:

$$n \times \alpha',$$

where n is a numeral and α' is a numerical expression, such that $c(\alpha') < c(\alpha)$. Thus, by the induction hypothesis $l(\alpha')$ is odd, and so

$$l(\alpha) = 2 + l(\alpha')$$

is odd. We have thus shown that the length of every numerical expression is odd, and, consequently, that the length of every sentence is odd.

(b)* Since any arbitrary string of symbols is an expression, the number of expressions of length k is simply the number of symbols in the alphabet of $\mathscr{L}_\times$ raised to the power k, i.e., 5^k.

To compute the number of sentences of length k, note that a sentence is any expression of the form:

$$n_1 s_1 n_2 s_2 \cdots s_{\frac{k-1}{2}} n_{\frac{k+1}{2}},$$

where each n_i is a numeral, and all but one of the s_i's is the symbol $\times$ (the remaining s_i being the symbol $=$). Since there are $3^{\frac{k+1}{2}}$ different ways of choosing the numerals $n_1, \ldots, n_{\frac{k+1}{2}}$, and $\frac{k-1}{2}$ different ways of choosing which of the s_i's is the symbol $=$, the number of distinct sentences of length k is

$$3^{\frac{k+1}{2}} \left(\frac{k-1}{2} \right).$$

The probability of choosing a sentence at random from among all the expressions of length k, is thus:

$$\frac{3^{\frac{k+1}{2}} \left(\frac{k-1}{2} \right)}{5^k} = \left(\frac{\sqrt{3}}{5} \right)^k \left(\frac{\sqrt{3}}{2} \right) (k-1).$$

Letting $\beta = \sqrt{3}/5$, we have:

$$\lim_{k\to\infty} \left[\beta^k \left(\frac{\sqrt{3}}{2} \right) (k-1) \right] \propto \lim_{k\to\infty} k\beta^k - \lim_{k\to\infty} \beta^k$$

Since $0 < \beta < 1$, $\lim_{k\to\infty} \beta^k = 0$, and by l'Hôpital's Rule, $\lim_{k\to\infty} k\beta^k = 0$. Hence, the probability of choosing a sentence at random from among all the expressions of length k goes to 0 as $k \to \infty$.

1.3 Let $V = \{val_1, val_2\}$, where:

$$val_1(\varphi_1) = \mathbf{T}$$
$$val_1(\varphi_2) = \mathbf{F}$$
$$val_2(\varphi_1) = \mathbf{F}$$
$$val_2(\varphi_2) = \mathbf{T}$$

and $val_i(\psi_j) = \mathbf{F}$, for all $i, j \in \{1, 2\}$.

Since there is no valuation in V which assigns $\mathbf{T}$ to all the sentences in Γ, every sentence is a logical consequence of Γ. Hence, condition (i) is satisfied. Moreover, $\varphi_i \not\models \psi_j$, for $i, j \in \{1, 2\}$, since $val_i(\varphi_i) = \mathbf{T}$ and $val_i(\psi_j) = \mathbf{F}$. Hence, condition (ii) is satisfied.

1.4 Let σ be an interpretation such that $val^{\sigma}(\varphi) = \mathbf{T}$, for all sentences $\varphi \in \Gamma'$. Then, since $\Gamma' \models \varphi$, for all $\varphi \in \Gamma$, $val^{\sigma}(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$. But since, by assumption, $\Gamma \models \psi$, $val^{\sigma}(\psi) = \mathbf{T}$. Hence, any interpretation under which all of the sentences in Γ' are true, is one in which ψ is true as well, i.e., $\Gamma' \models \psi$.

For $\Gamma = \emptyset$, the claim implies that if ψ is a logical truth, then ψ is a logical consequence of any set of sentences.

1.5 For any sentence φ, let

$$\Sigma(\varphi) = \{i : val^{\sigma_i}(\varphi) = \mathbf{T}\}.$$

Then,

$$\begin{aligned} \Sigma(0 = 0) &= \{0, 1\} \\ \Sigma(x = 0) &= \{0\} \\ \Sigma(x = 1) &= \{1\} \\ \Sigma(0 = 1) &= \emptyset \end{aligned}$$

Since, for any sentence φ, $\Sigma(\varphi) \subset \{0, 1\}$ and since $\varphi \equiv \psi$ iff $\Sigma(\varphi) = \Sigma(\psi)$, it follows that every sentence φ is equivalent to one and only one of the above four sentences.

1.6 Let $\varphi_1, \ldots, \varphi_k$ be a proof of ψ from Γ. We will prove, by induction on i, that $\Gamma' \vdash_T \varphi_i$, for $i = 1, \ldots, k$. There are three cases to consider: (i) φ_i is an axiom of T, (ii) $\varphi_i \in \Gamma$, (iii), φ_i is the result of applying some rule of T to sentences $\varphi_{n_1}, \ldots, \varphi_{n_j}$, where $n_m < k$, for $m = 1, \ldots, j$. In case (i), it is obvious that $\Gamma' \vdash_T \varphi_i$. In case (ii), $\Gamma' \vdash_T \varphi_i$ follows from the fact that $\Gamma' \vdash_T \varphi$, for all $\varphi \in \Gamma$. In case (iii), by the induction hypothesis, $\Gamma' \vdash_T \varphi_{n_m}$, for $m = 1, \ldots, j$, and so $\Gamma' \vdash_T \varphi_i$. Thus, $\Gamma' \vdash_T \varphi_k$, i.e., $\Gamma' \vdash \psi$.

1.7 (a) The proof is as follows:

1. $0 = 0$ [axiom]
2. $0 \times \beta = 0$ [0+,1]
3. $0 = 0 \times \beta$ [S,2]
4. $0 \times \alpha = 0 \times \beta$ [0+,3]

(b) Let $c(\alpha)$ be the number of times the symbol $\times$ occurs in α. We will prove the claim by induction on $c(\alpha)$. If $c(\alpha) = 0$, then α is a numeral, and so $\alpha = \alpha$ is an axiom. Hence $\vdash \alpha = \alpha$. Now, suppose that $c(\alpha) \geq 1$, and that $\vdash \beta = \beta$, for all β, such that $c(\beta) < c(\alpha)$. Since $c(\alpha) \geq 1$, $\alpha = n \times \beta$, where n is a numeral and β is a numerical expression such that $c(\beta) < c(\alpha)$.

There are four cases to consider:

(i) $n = 0$: In this case, $\vdash \alpha = \alpha$ follows from part (a), above.

(ii) $n = 1$: The proof of $\alpha = \alpha$ is as follows:

1. $\beta = \beta$ [theorem]

2. $1 \times \beta = \beta$ [1+,1]

3. $\beta = 1 \times \beta$ [S,2]

4. $1 \times \beta = 1 \times \beta$ [1+,3]

(iii) $n = x$ and x occurs in β: Since x occurs in β, there exists some (possibly empty) expression γ such that β is the result of applying the rule ($\times$) to the expression $x \times \gamma$. We thus have the following proof of $\alpha = \alpha$:

1. $\beta = \beta$ [theorem]

2. $x \times \gamma = \beta$ [$\times$,1]

3. $x \times x \times \gamma = \beta$ [x+,2]

4. $\alpha = \beta$ [$\times$,3]

5. $\beta = \alpha$ [S,4]

6. $x \times \gamma = \alpha$ [$\times$,5]

7. $x \times x \times \gamma = \alpha$ [x+,6]

8. $\alpha = \alpha$ [S,7]

(iv) $n = x$ and x does not occur in β: Since x does not occur in β, there exists some numeral $m \in \{0, 1\}$ and numerical expression γ, such that α is the result of applying the rule ($\times$) to $m \times \gamma$. Since $c(\gamma) < c(\alpha)$, following the proofs given in either (i) or (ii) above, we have that $\vdash m \times \gamma = m \times \gamma$. We thus have the following proof of $\alpha = \alpha$:

1. $m \times \gamma = m \times \gamma$ [theorem]

2. $\alpha = m \times \gamma$ [$\times$,1]

3. $m \times \gamma = \alpha$ [S,2]

4. $\alpha = \alpha$ [$\times$,3]

This completes the induction.

(c) Given (b), the proof is as follows:

1. $\alpha = \alpha$ [theorem]

2. $1 \times \alpha = \alpha$ [1+,1]

1.8 (a) Suppose that $\Gamma \vdash_T \psi$. Since T is sound with respect to V', $\Gamma \models_{V'} \psi$. Now, choose $val \in V$ such that $val(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$. Since $V \subset V'$, $val \in V'$, and so, since $\Gamma \models_{V'} \psi$, $val(\psi) = T$. Hence, for all $val \in V$, if $val(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$, $val(\psi) = \mathbf{T}$, i.e., $\Gamma \models_V \psi$. Thus, if $\Gamma \vdash_T \psi$, then $\Gamma \models_V \psi$, i.e., T is sound with respect to V.

(b) Suppose that $\Gamma \models_{V'} \psi$. Then for all $val \in V'$, if $val(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$, $val(\psi) = \mathbf{T}$. But since $V \subset V'$, this is also true for all $val \in V$, and so $\Gamma \models_V \psi$. But since T is complete with respect to V, it follows that $\Gamma \vdash_T \psi$. Hence, if $\Gamma \models_{V'} \psi$, then $\Gamma \vdash_T \psi$, i.e., T is complete with respect to V'.

(c) Let V be the class of all valuations, and suppose that T is sound with respect to V. Then, if $\Gamma \vdash_T \psi$, it follows that $\Gamma \models_V \psi$. But since V is the class of all valuations, it must be that $\psi \in \Gamma$, for otherwise, there would exist a valuation such that $val(\varphi) = \mathbf{T}$, for all $\varphi \in \Gamma$, but $val(\psi) = \mathbf{F}$. Hence, if $\Gamma \vdash_T \psi$, then $\psi \in \Gamma$.

Now, suppose, for contradiction, that T is a deductive system such that, either (i) T has an axiom; or (ii) T has a rule of inference of the form:

$$\frac{\Gamma}{\psi},$$

where $\psi \notin \Gamma$. In case (i), if φ is an axiom of T, then $\emptyset \vdash_T \varphi$, but $\varphi \notin \emptyset$. In case (ii), $\Gamma \vdash_T \psi$, but $\psi \notin \Gamma$. Hence, in either case, we contradict the assumption that if $\Gamma \vdash_T \psi$, then $\psi \in \Gamma$. Consequently, if T is sound with respect to V, then T has no axioms or non-trivial rules of inference.

(d) Let T be any deductive system. If V is the set of all possible valuations, then if $\Gamma \models_V \varphi$, $\varphi \in \Gamma$. But if $\varphi \in \Gamma$, $\Gamma \vdash_T \varphi$. Hence T is complete with respect to V.

1.9 (a) Suppose that $\{x = 0, x = 1\} \vdash_{T_0} 0 = 1$. Let $\varphi_1, \ldots, \varphi_k$ be a proof of $0 = 1$ from $\{x = 0, x = 1\}$ in T_0 that is no longer than any other such proof. We will show that $\varphi_1, \ldots, \varphi_k$ is a direct proof of $0 = 1$ from either $x = 0$, $x = 1$, or some axiom of the system. To do this, we must show (1) φ_1 is either $x = 0$, $x = 1$, or an axiom of the system; (2) that for each $i = 2, \ldots, k$, φ_i is the result of applying some rule of T_0 to φ_{i-1}; and (3) that the rule ($\neq$) is not applied in the proof.

Claim 1 follows from the fact that $\varphi_1, \ldots, \varphi_k$ is a proof from $\{x = 0, x = 1\}$.

We will prove claim 2 by backwards induction on i. Choose $i \geq 2$, and suppose that φ_j follows from φ_{j-1} by some rule of T_0, for all $j > i$, but that φ_i does not follows from φ_{i-1} by any rule of T_0. Then either (i) φ_i is an axiom of T_0 or a member of $\{x = 0, x = 1\}$, or (ii) φ_i is the result of applying some rule of T_0 to φ_n, for some $n < i - 1$. In case (i), $\varphi_i, \ldots \varphi_k$ is a proof of length $k - i + 1 < k$ of $0 = 1$ from $\{x = 0, x = 1\}$, in case (ii), $\varphi_1, \ldots, \varphi_n, \varphi_i, \ldots, \varphi_k$ is a proof of length $k + n - (i - 1) < k$ of $0 = 1$ from $\{x = 0, x = 1\}$. Hence, both cases contradict the assumption that there is no proof of $0 = 1$ from $\{x = 0, x = 1\}$ in T_0 of length less than k.

To prove claim 3, we observe that the rule ($\neq$) can only be applied to the sentence $0 = 1$, hence, if the rule is applied at some point in the proof $\varphi_1, \ldots, \varphi_k$, it follows that φ_j is $0 = 1$, for some $j < k$. But then $\varphi_1, \ldots, \varphi_j$ is a proof of $0 = 1$, from $\{x = 0, x = 1\}$ in T_0, but this contradicts the assumption that there is no proof of $0 = 1$ from $\{x = 0, x = 1\}$ in T_0 of length less than k.

(b) There cannot be a direct proof of $0 = 1$, from either $x = 0$, $x = 1$ or an axiom, since this would imply that $0 = 1$ is logically equivalent to one of these sentences, which it is not. Thus, from part (a), it follows that $\{x = 0, x = 1\} \not\vdash_{T_0} 0 = 1$, but since $\{x = 0, x = 1\} \models 0 = 1$, T_0 is incomplete.

Chapter 2

2.1 (a) Since there are 2^k distinct ways of determining the arguments that are given to a k-ary truth function, and since a truth-function is determined by the truth-values it returns when it is given any one of these sequences of arguments, there are 2^{2^k} k-ary truth functions.

(b) We will prove this claim by induction on k. Since g^2 is just the function g, the claim obviously holds for $k = 2$.

Choose $n_1, n_2, n_3 \in \{1, 2, 3\}$, such that $n_i \neq n_j$ for all $i \neq j$. If $n_3 = 3$, then by the symmetry of g:

$$\begin{aligned} g^3(\alpha_{n_1}, \alpha_{n_2}, \alpha_{n_3}) &= g(g(\alpha_{n_1}, \alpha_{n_2}), \alpha_{n_3}) \\ &= g(g(\alpha_{n_1}, \alpha_{n_2}), \alpha_3) \\ &= g(g(\alpha_1, \alpha_2), \alpha_3) \\ &= g^3(\alpha_1, \alpha_2, \alpha_3) \end{aligned}$$

Suppose $n_3 \neq 3$ and let i be the unique member of $\{1,2\}$, such that $n_i \neq 3$. Then, by the symmetry and associativity of g, we have:

$$\begin{aligned}
g^3(\alpha_{n_1}, \alpha_{n_2}, \alpha_{n_3}) &= g(g(\alpha_{n_1}, \alpha_{n_2}), \alpha_{n_3}) \\
&= g(g(\alpha_{n_i}, \alpha_3), \alpha_{n_3}) \\
&= g(\alpha_{n_i}, g(\alpha_3, \alpha_{n_3})) \\
&= g(\alpha_{n_i}, g(\alpha_{n_3}, \alpha_3)) \\
&= g(g(\alpha_{n_i}, \alpha_{n_3}), \alpha_{n_3}) \\
&= g(g(\alpha_1, \alpha_2), \alpha_3) \\
&= g^3(\alpha_1, \alpha_2, \alpha_3)
\end{aligned}$$

Thus, the claim holds for $k = 3$.

Now, choose $k > 3$, and suppose that g^m is symmetrical for all m such that $2 \leq m < k$. Choose $n_1, \ldots, n_k \in \{1, \ldots, k\}$, such that $n_i \neq n_j$ for all $i \neq j$. If $n_k = k$, then by the symmetry of g^{k-1}:

$$\begin{aligned}
g^k(\alpha_{n_1}, \alpha_{n_2}, \alpha_{n_k}) &= g(g^{k-1}(\alpha_{n_1}, \ldots, \alpha_{n_{k-1}}), \alpha_{n_k}) \\
&= g(g^{k-1}(\alpha_{n_1}, \ldots, \alpha_{n_{k-1}}), \alpha_k) \\
&= g(g^{k-1}(\alpha_1, \ldots, \alpha_{k-1}), \alpha_k) \\
&= g^k(\alpha_1, \ldots, \alpha_k)
\end{aligned}$$

Suppose $n_k \neq k$, and let $m_1, \ldots, m_{k-2}$ be the $k-2$ distinct members of the set $\{1, \ldots, k-1\}$, such that $n_{m_i} \neq k$. Then, by the symmetry of g^{k-1} and the symmetry and associativity of g, we have:

$$\begin{aligned}
g^k(\alpha_{n_1}, \ldots, \alpha_{n_k}) &= g(g^{k-1}(\alpha_{n_1}, \ldots, \alpha_{n_{k-1}}), \alpha_{n_k}) \\
&= g(g^{k-1}(\alpha_{m_1}, \ldots, \alpha_{m_{k-2}}, \alpha_k), \alpha_{n_k}) \\
&= g(g(g^{k-2}(\alpha_{m_1}, \ldots, \alpha_{m_{k-2}}), \alpha_k), \alpha_{n_k}) \\
&= g(g^{k-2}(\alpha_{m_1}, \ldots, \alpha_{m_{k-2}}), g(\alpha_k, \alpha_{n_k})) \\
&= g(g^{k-2}(\alpha_{m_1}, \ldots, \alpha_{m_{k-2}}), g(\alpha_{n_k}, \alpha_k)) \\
&= g(g(g^{k-2}(\alpha_{m_1}, \ldots, \alpha_{m_{k-2}}), \alpha_{n_k}), \alpha_k) \\
&= g(g^{k-1}(\alpha_{m_1}, \ldots, \alpha_{m_{k-2}}, \alpha_{n_k}), \alpha_k) \\
&= g(g^{k-1}(\alpha_1, \ldots, \alpha_{k-1}), \alpha_k) \\
&= g^k(\alpha_1, \ldots, \alpha_k)
\end{aligned}$$

Thus, g^k is symmetrical. This completes the induction.

2.7 (a) Let F be a set of symmetrical functions, and let f be a j-ary function definable in terms of F. Then either (1) $f \in F$, in which case f is symmetrical, or (2) there exist j-ary functions $g_1, \ldots, g_k \in F$, and a k-ary function $h \in F$, such that:

$$f = h(g_1, \ldots, g_k).$$

In the latter case, choose $n_1, \ldots, n_k \in \{1, \ldots, k\}$, such that $n_i \neq n_j$ for all $i \neq j$. Then, it follows from the symmetry of the g_i's that:

$$\begin{aligned}
f(\alpha_{n_1}, \ldots, \alpha_{n_j}) &= h(g_1(\alpha_{n_1}, \ldots, \alpha_{n_j}), \ldots, g_k(\alpha_{n_1}, \ldots, \alpha_{n_j})) \\
&= h(g_1(\alpha_1, \ldots, \alpha_j), \ldots, g_k(\alpha_1, \ldots, \alpha_j)) \\
&= f(\alpha_1, \ldots, \alpha_j)
\end{aligned}$$

Hence, f is symmetrical.

(b) It can easily be verified that each of the functions $f_\wedge$, $f_\vee$ and $f_\leftrightarrow$ are symmetrical. Hence, by part (a), any function definable in terms of $\{f_\wedge, f_\vee, f_\leftrightarrow\}$ is symmetrical. But $f_{\mathbf{P}_{1,2}}$ is *not* symmetrical since

$$f_{\mathbf{P}_{1,2}}(\mathbf{T}, \mathbf{F}) = \mathbf{T},$$

but

$$f_{\mathbf{P}_1,2}(\mathbf{F},\mathbf{T}) = \mathbf{F}.$$

Thus, $f_{\mathbf{P}_1,2}$ is not definable in terms of $\{f_\wedge, f_\vee, f_\leftrightarrow\}$.

2.9 (a) The truth-function $f_\uparrow$ can be summarized by the following truth-table:

φ	ψ	$\varphi \uparrow \psi$
T	**T**	**F**
T	**F**	**T**
F	**T**	**T**
F	**F**	**T**

By appeal to this truth table it is easy to verify that

$$\neg\varphi \equiv \varphi \uparrow \varphi$$

and

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

Thus, since $\{\neg, \wedge\}$ is truth-functionally complete, so is $\{\uparrow\}$.

The truth-function $f_\downarrow$ can be summarized by the following truth-table:

φ	ψ	$\varphi \downarrow \psi$
T	**T**	**F**
T	**F**	**F**
F	**T**	**F**
F	**F**	**T**

By appeal to this truth table it is easy to verify that

$$\neg\varphi \equiv \varphi \downarrow \varphi$$

and

$$\varphi \vee \psi = (\varphi \downarrow \psi) \downarrow (\varphi \downarrow \psi).$$

Thus, since $\{\neg, \vee\}$ is truth-functionally complete, so is $\{\downarrow\}$.

(b) Suppose that $f_*(\mathbf{T},\mathbf{T}) = \mathbf{T}$. We first prove that for every $\varphi \in \Gamma_{\{*\}}$ such that $\mathbf{P}_1$ is the only atom occuring in φ, $f_{\varphi,1}(\mathbf{T}) = \mathbf{T}$. If φ is the atomic sentence $\mathbf{P}_1$, then $f_{\mathbf{P}_1,1}(\mathbf{T}) = \mathbf{T}$. Suppose that $\varphi = (\psi * \rho)$, and that $f_{\psi,1}(\mathbf{T}) = \mathbf{T}$ and $f_{\rho,1}(\mathbf{T}) = \mathbf{T}$. Then

$$f_{\varphi,1}(\mathbf{T}) = f_*(f_{\psi,1}(\mathbf{T}), f_{\rho,1}(\mathbf{T})) = f_*(\mathbf{T},\mathbf{T}) = \mathbf{T}.$$

Since $f_\neg(\mathbf{T}) = \mathbf{F}$, it follows that $f_\neg \neq f_{\varphi,1}$, and so $\{*\}$ is not truth-functionally complete.

Thus, if $\{*\}$ is truth-functionally complete, $f_*(\mathbf{T},\mathbf{T}) = \mathbf{F}$. Moreover, by an exactly analogous argument it follows that if $\{*\}$ is truth-functionally complete, then $f_*(\mathbf{F},\mathbf{F}) = \mathbf{T}$.

There are four cases left to consider:

(i) $f_*(\mathbf{T},\mathbf{F}) = \mathbf{T}$ and $f_*(\mathbf{F},\mathbf{T}) = \mathbf{T}$:

In this case, $f_* = f_\downarrow$.

(ii) $f_*(\mathbf{T},\mathbf{F}) = \mathbf{F}$ and $f_*(\mathbf{F},\mathbf{T}) = \mathbf{F}$:

In this case, $f_* = f_\uparrow$.

(iii) $f_*(\mathbf{T},\mathbf{F}) = \mathbf{T}$ and $f_*(\mathbf{F},\mathbf{T}) = \mathbf{F}$:

In this case, $f_* = f_{\neg\mathbf{P}_2,2}$. We will prove that $\{*\}$ is not truth-functionally complete, by showing that for every $\varphi \in \Gamma_{\{*\}}$ such that $\mathbf{P}_1$ and $\mathbf{P}_2$ are the only atoms occurring in φ, $f_{\varphi,2}$ is neither equal to $f_\wedge$ nor $f_\uparrow$. If φ is atomic, this is obvious. Suppose $\varphi = \psi * \rho$ and that the claim holds for both ψ and ρ. Note that if $f_{\rho,2}$ is neither equal to $f_\wedge$ nor $f_\uparrow$, then the same is true of $f_{\neg\rho,2}$. But, then since

$$f_{\varphi,2} = f_*(f_{\psi,2}, f_{\rho,2}) = f_{\neg\mathbf{P}_2,2}(f_{\psi,2}, f_{\rho,2}) = f_\neg(f_{\rho,2}) = f_{\neg\rho,2},$$

$f_{\varphi,2}$ is neither equal to $f_\wedge$ nor $f_\uparrow$.

(iv) $f_*(\mathbf{T},\mathbf{F}) = \mathbf{F}$ and $f_*(\mathbf{F},\mathbf{T}) = \mathbf{T}$:

In this case, $f_* = f_{\neg\mathbf{P}_1,2}$, and so, by an analogous argument to that given in (iii), $\{*\}$ is not truth-functionally complete

It follows that if f_* is not equal to either $f_\uparrow$ or $f_\downarrow$, then $\{*\}$ is not truth-functionally complete.

2.10 (a) Let ψ be any axiom of T_S. From theorem 2.9, we have

$$\{\neg\varphi, \neg\neg\varphi\} \vdash \neg\psi.$$

It follows from the deduction theorem that

$$\neg\neg\varphi \vdash \neg\varphi \to \neg\psi,$$

and so by (SL-3) and (MP):

$$\neg\neg\varphi \vdash \psi \to \varphi.$$

But since ψ is an axiom, $\neg\neg\varphi \vdash \psi$, and so by (MP), $\neg\neg\varphi \vdash \varphi$.

(b) From (a), it follows that $\neg\neg\neg\varphi \vdash \neg\varphi$, and so, by the deduction theorem:

$$\vdash \neg\neg\neg\varphi \to \neg\varphi.$$

By (SL-3) and (MP), it follows that:

$$\vdash \varphi \to \neg\neg\varphi,$$

which, by the deduction theorem, implies that $\varphi \vdash \neg\neg\varphi$.

(c) By (SL-1), $\vdash \varphi \to (\psi \to \varphi)$, and so by the deduction theorem $\varphi \vdash \psi \to \varphi$.

(d) From theorem 2.9, it follows that $\{\neg\varphi, \varphi\} \vdash \psi$. Hence, by the deduction theorem, $\neg\varphi \vdash \varphi \to \psi$.

(e) From (a), $\neg\neg(\varphi \to \psi) \vdash \varphi \to \psi$, and, from (b), $\psi \vdash \neg\neg\psi$. Therefore, since $\{\varphi, \varphi \to \psi\} \vdash \psi$, we have

$$\{\varphi, \neg\neg(\varphi \to \psi)\} \vdash \neg\neg\psi,$$

and so, by the deduction theorem:

$$\varphi \vdash \neg\neg(\varphi \to \psi) \to \neg\neg\psi.$$

By (SL-3) and (MP), it follows that:

$$\varphi \vdash \neg\psi \to \neg(\varphi \to \psi),$$

which, by the deduction theorem, implies that $\{\varphi, \neg\psi\} \vdash \neg(\varphi \to \psi)$.

2.11 For a given interpretation σ, we define the function $ill^\sigma(\varphi)$, inductively as follows:

(a) $ill^\sigma(P_i) = \sigma(P_i)$

(b) $ill^\sigma(\neg\varphi) = \mathbf{F}$

(c) $ill^\sigma(\varphi \to \psi) = f_\to(ill^\sigma(\varphi), ill^\sigma(\psi))$

We say that a sentence φ is an *illogical truth* if $ill^\sigma(\varphi) = \mathbf{T}$, for all interpretations σ. It can easily be verified that if ρ is an instance of (SL-1), (SL-2) or the axiom schema:

$$(\varphi \to \psi) \to (\neg\varphi \to \neg\psi)$$

then ρ is an illogical truth. Moreover, if φ and $\varphi \to \psi$ are illogical truths, then so is ψ. Thus, if $\vdash_{T_S^*} \varphi$, then φ is an illogical truth. But the sentence

$$(\neg\mathbf{P}_1 \to \neg\mathbf{P}_2) \to (\mathbf{P}_2 \to \mathbf{P}_1)$$

is not an illogical truth, since if $\sigma(\mathbf{P}_1) = \mathbf{F}$ and $\sigma(\mathbf{P}_2) = \mathbf{T}$, then:

$$\begin{aligned} ill^\sigma((\neg\mathbf{P}_1 \to \neg\mathbf{P}_2) \to (\mathbf{P}_2 \to \mathbf{P}_1)) &= f_\to(ill^\sigma(\neg\mathbf{P}_1 \to \neg\mathbf{P}_2), ill^\sigma(\mathbf{P}_2 \to \mathbf{P}_1)) \\ &= f_\to(f_\to(ill^\sigma(\neg\mathbf{P}_1), ill^\sigma(\neg\mathbf{P}_2)), f_\to(\sigma(\mathbf{P}_2), \sigma(\mathbf{P}_1))) \\ &= f_\to(f_\to(\mathbf{F}, \mathbf{F}), f_\to(\mathbf{T}, \mathbf{F})) \\ &= f_\to(\mathbf{T}, \mathbf{F}) \\ &= \mathbf{F} \end{aligned}$$

Thus, $\nvdash_{T_S^*} (\neg\mathbf{P}_1 \to \neg\mathbf{P}_2) \to (\mathbf{P}_2 \to \mathbf{P}_1)$.