ELEMENTS OF DEDUCTIVE LOGIC

Elements of Deductive Logic

AN INTRODUCTION TO FORMAL LOGIC
AND ELEMENTARY METATHEORY

WITH EXERCISES

Antony Eagle

University of Adelaide

This book is open source. X$_\exists$L^{A}T$_E$X source code for the book is available from the book's GitHub repository, github.com/antonyeagle/edl, or by contacting me: antony.eagle@adelaide.edu.au.

Last updated 2016-01-20

Preface

This is a textbook covering the basics of formal logic and elementary metatheory. Its distinguishing feature is that it has more emphasis on metatheory than comparable introductory textbooks. It was originally written to accompany lectures in the introductory-intermediate logic course at the University of Oxford, *Elements of Deductive Logic*. It can be used independently of the Oxford lectures, of course.

This book is made available under a Creative Commons license (CC BY-SA 4.0), which allows users to share the work as they like, and to create derivative and modified works for any purpose, as long as the original creator is appropriately credited, and derivative works are licensed in the same way. I hope this will encourage any users of the book to adapt it to their local environments.

Contents

List of Tables

List of Figures

Chapter 0

Introduction

Introductory logic courses serve generally to tell you how to *use* a new logical language. An introductory course will give you an introduction to the grammar, or *syntax*, of the new formal language, tell you how to interpret the grammatical sentences by giving a *semantics*, and most importantly from the perspective of applications, tell you how to translate from natural languages into the formal language. With a semantics and an translation scheme, you can now use a formal language to do various things more easily than you could otherwise: you can express claims precisely and without risk of ambiguity (No problem with 'everybody loves somebody' in formal languages!), and check the validity or otherwise of arguments of certain forms, which – if you've done your translation well – reflects on the validity or otherwise of certain informal arguments you might be interested in as a philosopher or mathematician. If you have a formal proof system, giving rules for formal derivations in your languages, you can even construct arguments directly in the formal language in a more or less mechanical fashion, thus enabling you to prove various claims directly without having to detour through semantics.

Thinking about this for a second, though, we need some reassurance that the formal languages will do what we want them to. Is it true that if

a formal proof is correct, then the argument it formalises is semantically valid? Is it the case that every valid argument has a corresponding formal derivation? Which things can and cannot be expressed in a given formal language? Is a formal language adequate to each argumentative use we might have for it in mathematics, or in philosophy? Are formal derivations really purely mechanical, or must there be human judgment involved too? These are questions *about* formal logical languages, questions in what we might call the *metatheory* of logic. Such questions are the focus of this book. Such questions aren't about how to use formal logics to do things, but about whether they are fit for purpose, whether they can in fact do the things we want them to do.[1]

Answering such questions involves drawing a distinction between the *object language*, the one we are talking about, and the *metalanguage*, the language we are using to talk about the object language. For this book, the object languages are various more-or-less orthodox languages of classical logic: a sentential language $\mathscr{L}_1$,[2] a predicate language $\mathscr{L}_2$, and a predicate language with identity $\mathscr{L}_=$. And the metalanguage is the mathematically enriched version of English we use to talk about formal languages.

The Structure of this Book   I begin, in chapters 1 and 2, by outlining some mathematical tools – specifically, some elementary *set theory*, including the theory of functions – that will help us study and understand the formal languages which are our subject. In chapters 3 and 4, we'll briefly review the syntax and semantics of classical sentential logic. Then we'll start proving some results in metatheory. We'll prove some small but still

---

[1]Such questions could easily be asked about natural languages too, though since the syntax and semantics of natural language are orders of magnitude harder to specify precisely, it is correspondingly harder to establish anything conclusively about the features of such languages.

[2]This is sometimes called the language of *propositional logic*, but as I take propositions to be the meanings of sentences, and logic to be the study of consequence relations between meaningful sentences, I wish to avoid the terminology of 'propositional logic' as potentially misleading.

interesting results about the truth-functions, and about the intimate connections between conjunction ('∧'), disjunction ('∨'), and negation ('¬'). In chapter 5, we'll prove our first 'big' result: the so-called *compactness* theorem for sentential logic. We'll take a break from metatheory in chapter 6, to introduce and talk about *formal derivations* in sentential logic, and introduce a particular kind of natural deduction derivation system. Then we'll prove two central results connecting derivability with validity: the *soundness* and *completeness* theorems for sentential logic. I also mention issues of computational implementation, in particular, the *decidability* of sentential logic. In chapter 7, we'll review *predicate* logic and (in chapter 8) its natural deduction derivation system, and we'll prove soundness for standard predicate logic. I sketch the proof of completeness, and briefly consider issues of decidability. In chapter 9, issues about predicate logic with identity are discussed, including the soundness of the natural deduction system we construct, and issues about numerical quantification and compactness are discussed. In the final chapter 10, we turn from metatheory to issues about the relation between logical languages and natural languages, particularly focussing on issues about conditionals, entailment, designation and the theory of relations.

Exercises and Further Reading   This book also includes exercises, contained in a section entitled 'Exercises' at the end of each chapter. These exercises are mostly review and consolidation of the material in the chapter, but some are intended to extend and stretch a student's knowledge of the material. No solutions are provided.

Many chapters also contain a section of 'Further Reading', which contains references to books and articles that may provide additional information on particular topics of interest. Details of the items there referred to can be found in the Bibliography at the end of the book.

Appendices   The text makes free use of Greek letters as variables of various kinds; a quick refresher of the English names of these letters can be found in Appendix B. There are a number of defined expressions; a list of definitions can be found in Appendix C.

Some Other Useful Readings   The languages and derivation systems used in this book are those developed in Volker Halbach's *Logic Manual* (Halbach, 2010). While the present book is self-contained, it may be helpful to review the material in Halbach's book as a useful adjunct. Another useful source of related material, that I drew on in a number of places in this book, is Bostock (1997).

Chapter 1

Mathematical Induction and The Notion of Proof

1.1 Inductive Proofs

Mathematical Proof   What is a proof? A proof of a theorem is just an argument, from assumptions that are known truths, that the theorem is true. In the mathematical case, the arguments involved are such that whenever the assumptions are true, the conclusions are true – they are supposed to be necessarily truth-preserving in English. And the assumptions in the mathematical case are the axioms of a particular mathematical theory: necessary mathematical truths that are collectively sufficient, when supplemented by appropriate definitions, to characterise a given body of mathematical truths. There are many issues surrounding the nature of proof that we will neglect; but basically, a proof is an absolutely *conclusive* argument for a given claim, based on assumptions which are themselves known.

Mathematical Induction   We shall often use a powerful method of reasoning known as proof by *mathematical induction*.

The basic idea is simple. Assume we have a numbered sequence of

things. Suppose (i) we can show that: the first member of the sequence has some property, and (ii) we can show that: if some member has the property, then then the subsequent member also has the property. Then we can conclude that *every* member of the sequence has the property.

The easiest case is where the sequence *is* the natural (counting) numbers $\mathbb{N}$, but the principle also works for sequences of other things that are ordered like the natural numbers.

The Weak Principle of Induction   There are a number of formulations of induction. This one follows our intuitions closely. It is a principle about the legitimacy of a certain pattern of reasoning.

Definition 1 (Weak Principle of Induction). Assume there is a sequence of items S, ordered by the natural numbers. If both of the following conditions hold,

1. P is true of the first member of S; and

2. if P is true of the n-th member of S, then P is true of $n+1$-th member of S;

then, P is true of every member of S.

To apply this principle, we first establish the *base case* for condition (i); and then the *induction case*, where we assume that (for some *arbitrary* n) P holds of the n-th member, and show it holds of the $n + 1$-th member, to show condition (ii). (Condition (ii) is a conditional claim: it can hold even if nothing is F, as long as, if any n is F, that suffices to establish that the subsequent thing is also F.) Having established those conditions, the principle of induction says, it is now legitimate to infer that every member of the sequence has the property in question. Here is an example.

Theorem 1

For every natural number n,

$$(1.1) \qquad 0 + 1 + 2 + \ldots + n = \frac{n(n + 1)}{2}.$$

Proof. We first prove the *base case*, that the property holds of 0 (the first member of $\mathbb{N}$). Easy: $0 = \frac{0 \cdot (0+1)}{2} = \frac{0}{2} = 0$. (Usually the base case is the easy part.)

Now we wish to prove the *induction step*. That is, we assume the property holds of n, and show it must hold of $n+1$. So we assume: $0+1+\ldots+n = (n(n + 1))/2$.

$$(1.2) \qquad (0 + 1 + \ldots + n) = \frac{n(n + 1)}{2}$$

$$(1.3) \qquad (0 + 1 + \ldots + n) + n + 1 = \left(\frac{n(n + 1)}{2}\right) + n + 1$$

$$(1.4) \qquad = \frac{n(n + 1)}{2} + \frac{2(n + 1)}{2}$$

$$(1.5) \qquad = \frac{(n + 2)(n + 1)}{2}$$

$$(1.6) \qquad = \frac{(n + 1)((n + 1) + 1)}{2}$$

Equation (1.6) is clearly just an instance of Equation (1.1) with '$n + 1$' in place of 'n', as required. This proves the induction step. $\qquad \square$

The Strong Principle of Induction

Definition 2 (Strong Principle of Induction). 'For all n, if for all $m < n$, $P(m)$ then $P(n)$' implies 'For all n, $P(n)$.'

The strong principle is just a conditional claim – there is no basis step. Why? Assume that for every n, if for all $m < n$ $F(m)$, then $F(n)$. Consider, in particular, $n = 0$. Is it true that for all $m < 0$, $F(m)$? Well, if it were false,

there would be a number less than 0 such that it isn't F. But there are no numbers less than zero, so in particular, there are none of them that are not F. So – in a degenerate or vacuous way – all numbers less than 0 are F. So the antecedent of the conditional is satisfied, and since the conditional is true, the consequent must be true. The consequent is $F(0)$ – which is the basis step. Still, sometimes its useful to set out the basis step even when we're using strong induction.

The Weak Principle of Induction is equivalent to the Strong Principle. I show the left-to-right direction of the equivalence.

Theorem 2 (Weak to Strong)

The weak principle of induction entails the strong principle of induction.

Proof. We assume the conditional premise of the strong principle: that for every n, if for all $m < n$ $F(m)$, then $F(n)$. We will also assume the weak principle of induction. And we'll prove from those two assumptions that every n is F. That is, assuming the validity of reasoning in accordance with the weak principle of induction, we can demonstrate the validity of reasoning in accordance with the strong principle.

Here's how. We define a new condition, G, as follows:

$$G(n) =_{\text{df}} \text{ for all } m < n, \ F(m).$$

(So a number has G if all its predecessors have F.) We can re-write our conditional assumption using G as follows:

(†)   For every n, if $G(n)$ then $F(n)$.

Base case: we show that $G(0)$. Trivial, by the argument just given.

Induction step: we want to show that, on the assumption that $G(n)$, it follows that $G(n+1)$. Assume $G(n)$. By (†), $F(n)$. By the definition of G, for all $m < n$, $F(m)$. So for all $m \leqslant n$, $F(n)$. But it is obvious that the collection of numbers less than or equal to n, is the same as the collection of numbers

less than $n + 1$. So we have shown that for all $m < n + 1$, $F(n)$. But by the definition of G again, we've just shown $G(n + 1)$.

Applying the weak principle of induction to this base case and this induction step, we conclude that for all n, $G(n)$. By (†) again, for all n, $F(n)$, as desired. □

The Least Number Principle

Definition 3 (Least Number Principle). If M is a subset of $\mathbb{N}$, and is non-empty, then M has a least member.

On the assumption that reasoning in accordance with the LNP is good, then reasoning in accordance with the weak principle is also good, as is shown by the following theorem.

Theorem 3 (LNP to Weak)

The Weak Principle of Induction follows from the Least Number Principle.

Proof. Assume that P is a property such that $P(0)$, and for every n, $P(n) \rightarrow P(n + 1)$. We prove from the LNP and these assumptions that for every n, $P(n)$ (which is the Weak Principle).

Let M be the set of numbers which do *not* satisfy P. By the LNP, M has a least member if it is not empty – call this member m. $P(0)$ holds, by assumption, so $m > 0$; and since m is the least member of M, $P(m - 1)$. But since if $P(n)$ then $P(n + 1)$, $P(m)$ follows from $P(m - 1)$. That contradicts our assumption that $m \in M$. So there can be no least member of M. A set only has no least member if it has no members at all; so there is no number n such that $\neg P(n)$, hence for every number n, $P(n)$.

To show that all three principles of induction we've considered are equivalent, we'd now need to show that the LNP follows from the Strong or Weak Principles of Induction; this is left for an exercise.

1.2   Proof by *Reductio*

Another important proof strategy that you will commonly use is proof by *reductio ad absurdum* (literally, 'reduction to the absurd'). In mathematics, this is commonly known as *proof by contradiction*. The idea of this proof is to show that some claim is true by showing that its negation cannot be true. In fact, we just used this proof strategy in the proof of Theorem 3.

Example of *Reductio* Proof   Consider, for example this elementary number theoretic result:

Theorem 4

There is no largest prime number.

Proof. Suppose – for the sake of argument – that k is the largest prime. Take all the prime numbers less than or equal to k. Multiply all these numbers together, and add 1. The resulting number cannot be divided by k, nor by any of the primes smaller than k – for if you were to divide by any of those numbers, the remainder would be one. Therefore it is either itself prime, or (by the fundamental theorem of arithmetic) it is divisible by a prime number greater than k. In either case, there is a prime greater than k, thus refuting our initial supposition that there is a greatest prime. So that initial supposition must be false, and theorem holds.   □

I have made the role of the supposition explicit – we show that an absurdity, or a flat out contradiction, follows from our assumption by impeccable deductive reasoning. But the hallmark of impeccable deductive reasoning is that, if you start with a truth, you will end up with a truth. Since we have ended up with something untrue – an impossibility – we must not have started with a truth.

1.3   Intuitionism

The foregoing line of reasoning is very persuasive, and its use is pervasive throughout mathematics and in this book. However, some mathematicians and philosophers, beginning with the influential topologist Brouwer, have wondered whether this proof strategy is legitimate. Everyone agrees that, in showing that our initial supposition of 'not S' is untrue, we have shown that its negation, 'not not S' is true. But these *intuitionists* say that 'not not S' is not equivalent to S. Why do they say this? It is because proof by contradiction permits us to show that there is an entity with a certain property (because to assume there is no such entity would lead to a contradiction), without telling us what that entity is. Such proofs are called *non-constructive*, for they do not construct the entity which has the property in question.[1]

Intuitionism on Mathematical Truth   What is the interpretation of mathematical truth that motivates intuitionists to reject the equivalence of 'not-not-P' and 'P'? It is this: *mathematical truth is mathematical proof* (Shapiro, 2000). Mathematical objects don't just exist, out there in some Platonic heaven, waiting for facts about them to be discovered. They need to be brought into existence by the activity of a mathematician. But bringing something into existence is a positive act – simply showing that it would be absurd if something didn't exist, is not yet to construct it. So intuitionists offer a reinterpretation of the language of proof – English, in our case – to reflect their philosophical commitments.

---

[1]Consider, for example, the classical proof that there are irrational numbers a and b such that a^b is rational. We can show this by considering $q = \sqrt{2}^{\sqrt{2}}$, and show that either $q^{\sqrt{2}}$ is rational if q is irrational, or that q itself is rational – either way, there is a pair of irrational numbers (either $a = b = \sqrt{2}$, or $a = q, b = \sqrt{2}$) such that a^b is rational. But we haven't shown *which* of these numbers is in fact rational – we haven't constructed an example, we've merely shown that there must be an example. (In fact, q is irrational, so that is an actual constructed example that verifies the theorem.)

- The absurd sentence $\perp$ is not constructible.

- A construction of 'P and Q' consists of a construction of P and a construction of Q.

- A construction of 'P or Q' consists of a construction of P or a construction of Q.

- A construction of 'If P then Q' is a technique which transforms any construction of P into a construction of Q.

They define 'not-P' to mean: 'If P then $\perp$', i.e., there is a technique which turns a construction of P into a construction of the absurdity. 'not not P', accordingly, says that there is a technique which turns a technique turning a construction of P into the construction of the absurdity, itself into a construction of an absurdity. But that is not at all obviously the same as a construction of P.

The Law of Excluded Middle   Characteristically, the rejection of the so-called *law of double negation elimination* leads to the rejection of the *law of excluded middle* (LEM): intuitionists reject the principle that for any P, either P or not P. For them, that is equivalent to: for any P, there is (already) either a construction of P, or a construction turning a construction of P into the construction of an absurdity. Why should we think, prior to any such construction being offered, that they antecedently exist? Of course we can prove something related to LEM.

Theorem 5 (LEM–)
For any P, not-not-(P or not-P)

Proof.   Assume (i) that for some P, it is the case neither that P, nor that not-P. Assume (ii) P: it follows that P or not-P, so our assumption (ii) must be false. If assumption (ii) were correct, in conjunction with assumption (i), we'd be able to construct a contradictory claim of the form Q and not-Q.

That is absurd, and not constructible. So, under assumption (i) alone, we can turn any construction of P into a construction of the absurd, which is of course a construction of not-P from assumption (i). But it follows from not-P, again trivially, that 'P or not-P'. So we can turn an assumption of (i) into a construction of the absurdity, which is a construction of not-not-(P or not-P). □

So we can show that it is absurd to deny LEM; but that is not yet tantamount to its being legitimate to assert it.

Resisting Intuitionism   To insist on constructive proofs, and to eschew the use of proof by contradiction, is to cripple mathematics. This would be acceptable – perhaps – if there were a powerful philosophical rationale for thinking that, in general 'not not S' is not equivalent to S. But no persuasive examples have really been given, despite the best efforts of intuitionists to defend their view. So in this book we will adopt the standard highly plausible classical position, that proof by contradiction is perfectly acceptable.

Further Reading

On mathematical induction, see Machover (1996). Two philosophical attempts to motivate discussion of intuitionism are by Heyting (1956) and Dummett (1983). A useful discussion of the non-standard features of intuitionistic mathematics is by Iemhoff (2015).

Exercises

1. Prove that the Weak Principle of Induction entails the Least Number Principle.

2. *Bivalence* is the principle that every meaningful declarative sentence is either true or false. What is the relationship between Bivalence and the Law of Excluded Middle?

Chapter 2

Set Theory

2.1 Sets

A *set* is a collection of objects. Like a collection, it is a single thing, not a plurality. If X is a set, the objects in X are called its *elements* or *members*. We write $x \in X$ for 'x is an element of X' (and $y \notin X$ for 'y is not an element of X').

Sets are *individuated* by their members; or as we sometimes say, sets are individuated purely *extensionally*. There must be something to a set over and above its members – otherwise it would be simply a plurality identical to those members. But whatever that 'something' is, it is minimal enough that there cannot be two distinct sets which share the same members. Therefore, if X and Y name sets which have all and only the same members, $A = B$. So we can specify a set by listing the members, conventionally in curly braces: $X = \{x_1, ..., x_n, ...\}$. (But sets differ from *sequences*: two distinct sequences can have the same members as long as they occur in a different order in the two sequences.)

A set is not *composed* of its members in any straightforward sense. We can see this most easily in the case of a *singleton* set, a set with just one member. Clearly $x \in \{x\}$; but since x itself need not be a set, $x \notin x$. Since

sets are individuated by their members, and since $\{x\}$ has a member that x does not, $x \neq \{x\}$.

Axioms The standard approach to set theory is to take this informal conception of a set as a collection, and set down *axioms* that characterise the behaviour of sets. We've already seen one of the standard axioms of so-called Zermelo-Fraenkel set theory : the axiom of *extensionality*:

Axiom 1 (Extensionality). For any sets X and Y, if for all objects x, $x \in X$ iff $x \in Y$, then $X = Y$.

I will below mention various axioms, the basic starting points of modern set theory, but we will largely proceed in an informal way. In particular, I will not be offering proofs from the axioms of all the set theoretic principles I state below.

Sets in Sets One thing is worth being explicit about at the outset. *Sets can have other sets as members.* We might be used to thinking of collections as collections of material objects (rocks, china plates, etc.) But since a set is not identical to its members, and we can talk about and quantify over sets, they are also objects available to be the members of further sets. We might *start* with the non-sets, or *ur-elements* as they are sometimes known; but once we make sets of them, we can then make further sets which include sets as members, and then further sets involving sets of sets, etc. This is the so-called *iterative conception of a set.*

Defining Sets We've seen one way to define a set: list the members. But what if there are too many to list?

One thing we can do is define sets *inductively*. This means: given a base case, and a *generating relation*, we define a set that includes the base case, and anything related to a member of the set by the generating relation, and nothing else (the *closure condition*). Here's an example.

Definition 4 (Your Ancestors). *Base case:* Your mother and father are ancestors.
Generating Relation: Any parent of an ancestor is an ancestor.
Closure Condition: Nothing else is an ancestor of you.

Let A_{ae} be the set of my ancestors. Any member of A_{ae} is a parent of a parent of...a parent of me, with $n > 0$ occurences of 'a parent of'. The property 'is an ancestor of AE' is called the *ancestral* of the property 'is a parent of AE'. And more generally we can say that the relation R 'x is an ancestor of y' is the ancestral of the relation 'x is a parent of y'. Note that the set $A_{ae} = \{x : R(x, ae)\}$

Alternatively, we can specify a set by stating a condition F that the members all satisfy, typically in some formal mathematical language. Then we can specify the set of things x such that each x is F. We write this $\{x : F(x)\}$. If $Y = \{x : F(x)\}$, then $y \in Y$ iff $F(y)$. But in fact any collection of items forms a set: there needn't be a natural condition all the members satisfy.

Not just any condition will define a set. Consider the condition $F = x \notin x$. This defines a set $\mathbf{R} = \{x : x \notin x\}$, the set of all non-self-membered things.

Is $\mathbf{R}$ a member of $\mathbf{R}$? Assume $\mathbf{R} \in \mathbf{R}$. Then $\mathbf{R}$ must meet the defining condition F, so that $\mathbf{R} \notin \mathbf{R}$. But since $\mathbf{R}$ cannot both be and not be a member of itself, the assumption must be wrong. So $\mathbf{R} \notin \mathbf{R}$. But then $\mathbf{R}$ does meet F, and therefore $\mathbf{R} \in \mathbf{R}$. A contradiction follows. This is known as *Russell's paradox.*

The problem is assuming that *any* condition at all defines a set. The alternative is simple: to assume that, *if we are given a set X*, then any condition at all will *separate* the members of X into those which meet the condition and those which do not. That is: for any set X, and any condition F, there will exist a set Y which contains everything which is both a member of X and satisfies F. This is the axiom of *separation.*

Relations Between Sets

Definition 5 (Subset). A set X is a *subset* of Y, written $X \subseteq Y$ if every member of X is also a member of Y: for every x, if $x \in X$ then $x \in Y$. X is a *strict subset* (or *proper subset*) of Y if $X \subseteq Y$ and $Y \nsubseteq X$ (written '$X \subset Y$').

Trivially, for all sets X, $X \subseteq X$. It also follows that if $X \subseteq Y$ and $Y \subseteq X$ then $X = Y$ (easy). And also if $X \subseteq Y$ and $Y \subseteq Z$ then $X \subseteq Z$.

Definition 6 (Relative Complement). If X and Y are sets, then $X \setminus Y$ is the set of all members of X which are *not* members of Y, called the *relative complement* of Y with respect to X. If X and Y exist, then $X \setminus Y$ exists: $X \setminus Y = \{x : x \in X \wedge x \notin Y\}$.

Definition 7 (Union). If X and Y are sets, then $X \cup Y$ (read 'the *union* of X and Y') is the set which contains all the members of X and all the members of Y. If X and Y exist, $X \cup Y$ exists: $X \cup Y = \{x : x \in X \vee x \in Y\}$.

If $\mathbf{X} = X_1, \ldots, X_n, \ldots$ is a set of sets, $\bigcup \mathbf{X} = X_1 \cup \ldots \cup X_n \cup \ldots$.

The existence of unions is supported by this axiom:

Axiom 2 (Unions). Suppose $\mathbf{X}$ is any set of sets (i.e., a set with only other sets as its members). For any such $\mathbf{X}$, there is a set Y such that, for any z, $z \in Y$ iff z is a member of one of the members of $\mathbf{X}$.

That is, Y is the set containing all the things which are in any of the sets in $\mathbf{X}$. Since if A and B are sets, there is a set $\{A, B\}$, it follows from the axiom of Unions that there is a set which contains all the members of A and all the members of B. By extensionality, there is just one such set; we call it $A \cup B = \bigcup \{A, B\}$.

Definition 8 (Intersection). If X and Y are sets, the set $X \cap Y$ which contains only members of both is called the *intersection*. If X and Y exist, $X \cap Y$ exists: $X \cap Y = \{x : x \in X \wedge x \in Y\}$.

If $\mathbf{X} = X_1, \ldots, X_n, \ldots$ is a set of sets, $\bigcap \mathbf{X} = X_1 \cap \ldots \cap X_n \cap \ldots$.

Note the links in the definition between $\cap$ and 'and', and $\cup$ and 'or'.

The Empty Set   If X is a set, then by the definition of relative complement, $X \setminus X$ is a set. Suppose $X \setminus X$ has a member, z. Then $z \in X \setminus X$ iff $z \notin X$ and $z \in X$. Contradiction; so $X \setminus X$ has no members. This set is the *empty set*, written $\varnothing$. $Y \setminus Y$ is also empty; by extensionality, $X \setminus X = \varnothing = Y \setminus Y$; there is only one empty set.

The empty set is the collection which has nothing in it. It corresponds to the empty list. While the empty collection has no members, that doesn't mean that it itself is nothing: it is the unique thing which is a set and also has no members.

If X is any set, then for any x, if $x \in \varnothing$ then $x \in X$ – for the empty set has no members at all. By the definition of subset, therefore, $\varnothing \subseteq X$, for any X.

Absolute Complement and the Universal Set   The empty set is the smallest possible set. Is there a largest set? If there is a *universal set* Ω – a set which has everything (every individual *and every set*) as a member – we could define something a bit closer to negation than relative complement. We could define: $-X = \Omega \setminus X$; it would be the set of all things which are not in X.

This is *not* possible. Suppose there were a universal set. Then the axiom of separation, that if X is a set then there is a set Y which is a subset of X and such that for all $x \in X$ if $F(x)$ then $x \in Y$, will entail a contradiction.[1] So there is no universal set. (Puzzle: how do we quantify over all sets, as we have been doing, if there is no set big enough to contain them all?)

Power Set   So far you might think that sets are just collections of individuals (i.e., non-sets). But set theory is much more general than that: any-

---

[1]Hint for exercises: Apply the condition $x \notin x$ to the universal set Ω in the axiom of separation, and we will construct the Russell set **R** perfectly legitimately. Since the Russell set does not exist, something's gone wrong: the only candidate is the assumption that there is a set of everything.

thing at all, including other sets, can be gathered into a set. One useful set of sets is the *power set* of a set X, the set of all subsets of X:

Definition 9 (Power Set). Given a set X, the *power set* of X, denoted $\wp(X)$, is defined: $\wp(X) = \{Y : Y \subseteq X\}$. (It is an axiom of set theory that for any set, its power set exists.)

Notice that, for any X, $\varnothing \in \wp(X)$; and $X \in \wp(X)$. The name 'power set' comes from this fact:

Theorem 6

If X has n members, $\wp(X)$ has 2^n members.

Proof. A subset of X is a set containing perhaps some but not necessarily all the members of X. We can specify a subset by saying, of each member of X, whether it is in the subset. Suppose the members of X are $x_1, \ldots, x_n$. Then a finite binary sequence, consisting of 1s and 0s, of length n suffices to determine a subset of X, by the following principle: let x_i be a member of the subset associated with s iff the ith place in the sequence s is a 1. By extensionality, each subset of X corresponds to exactly one such sequence. So how many finite binary sequences of length n are there? Each place i in the sequence can be either 1 or 0, so there are 2^n such sequences. So there are 2^n distinct subsets of X, and hence 2^n members of $\wp(X)$. $\qquad\square$

Ordered Pairs and Sequences As mentioned above, sets are identical iff they have the same members. How those members happen to be listed is irrelevant: $\{x, y\} = \{y, x\} = \{x, x, y\}$. So if we want to capture the idea of an *ordered sequence*, it cannot be a set – order doesn't matter for sets, and one can't have repeated elements in a set. Yet we can model sequences just using sets.

Definition 10 (Ordered Pair). An *ordered pair* $\langle x, y \rangle$ is defined as the unordered set $\big\{\{x\}, \{x, y\}\big\}$.

We need to check whether this definition works. We do so by establishing the fundamental principle of the identity of ordered pairs follows from this definition, namely, that ordered pairs are identical iff they have the same first member, and the same second member.

Theorem 7 (Criteria of Identity for Ordered pairs)

$\langle x, y \rangle = \langle u, v \rangle$ *iff* $x = u$ *and* $y = v$.

Proof. $\langle x, y \rangle = \langle u, v \rangle$ iff $\{x, \{x, y\}\} = \{u, \{u, v\}\}$ iff (i) $x = u$ and $\{x, y\} = \{u, v\}$ or (ii) $x = \{u, v\}$ and $\{x, y\} = u$. Since (ii) is not possible – it would entail that x is a member of one of its own members – it must be that (i), which in turn holds iff $x = u$ and $y = v$. □

In the proof of this theorem, we had to appeal to the fact that no set is a member of a member of itself.[2] This obviously holds on the iterative conception of a set. It tacitly relies on this characteristic axiom of Zermelo-Fraenkel set theory:

Axiom 3 (Foundation). Every non-empty set X contains a member y such that $X \cap y = \emptyset$.

(This axiom is also known as *regularity*.)

Theorem 8

No set is a member of itself.

Proof. Let X be an arbitrary set. Clearly $X \in \{X\}$. By Foundation, since $\{X\}$ is non-empty, it has a member y such that $\{X\} \cap y = \emptyset$. Since the only member of $\{X\}$ is X, it must be that $\{X\} \cap X = \emptyset$. But if $X \in X$, then X and $\{X\}$ do have a member in common – namely, X – and their intersection isn't empty. So $X \notin X$. □

---

[2]Kuratowski's original definition of an ordered pair was this: $\langle x, y \rangle = \{\{x\}, \{x, y\}\}$. In proving the criteria of identity for ordered pairs for this alternative definition, we need not appeal to the above-cited fact. (Exercise.)

We can extend the idea of an ordered pair to an ordered sequence of any number of elements, as follows:

Definition 11 (Ordered n-tuple). Let the notation $\langle\!\langle x_1, \ldots, x_n \rangle\!\rangle$ denote an ordered n-tuple, for $n \geqslant 2$. We offer an inductive definition.

- *Base case.* $\langle\!\langle x_1, x_2 \rangle\!\rangle =_{\text{df}} \langle x_1, x_2 \rangle$.

- *Inductive step.* $\langle\!\langle x_1, \ldots, x_{n+1} \rangle\!\rangle =_{\text{df}} \langle \langle\!\langle x_1, \ldots, x_n \rangle\!\rangle, x_{n+1} \rangle$.

From now on, I will abuse notation, and use '$\langle$' and '$\rangle$' to delimit all ordered sequences, not only pairs. But, speaking strictly, when in the following I denote an ordered triple by $\langle x, y, z \rangle$, that expression should be interpreted as $\langle\!\langle x, y, z \rangle\!\rangle = \langle \langle x, y \rangle, z \rangle$. It doesn't really matter *how* we define ordered tuples; all that matters is that they are well-defined set-theoretic entities which behave as expected. (So, strictly speaking, we'd need to prove a theorem analogous to Theorem ?? showing that ordered n-tuples as just defined behave in the right way. We are not going to do that.)

One remaining case deserves attention. Our notation covers all ordered sequences of two or more members. What about the degenerate case of an ordered sequence with just one member? We shall adopt the following harmless convention (which will however be of use in §7.2):

Definition 12 (Ordered 1-tuple). An ordered 1-tuple $\langle x \rangle$ is simply x itself.

2.2 Relations and Functions

Intuitively, a relation R connects some things (its *relata*). We will model a relation set theoretically as a set of *pairs* of entities such that the first member of the pair bears the relation to the second. So the set

$$\{\langle \text{Melbourne}, \text{Victoria} \rangle, \langle \text{Brisbane}, \text{Queensland} \rangle, \ldots\}$$

is (part of) the relation of ' − is the capital city of − '.

Definition 13 (Binary Relation). X is a binary relation on a *domain D* iff it is a set that contains nothing except ordered pairs $\langle x, y \rangle$ where $x \in D$ and $y \in D$. ($\varnothing$ contains nothing at all, and so nothing other than ordered pairs, so is a binary relation).

Definition 14 (Cartesian Product). The *Cartesian product* of sets X and Y, written '$X \times Y$', is the set of all pairs $\langle x, y \rangle$ such that $x \in X$ and $y \in Y$.

A relation on D is a subset of $D \times D$.

   We discuss binary relations further, from a more metaphysical perspective, in Chapter 9.

Functions   Melbourne and Victoria are related by the 'is the capital city of' relation iff 'Melbourne' names the same thing as 'the capital of Victoria'. The term-forming expression 'the capital of $-$' denotes a *function*. We'll treat the notion of a function more precisely, as a special kind of relation.

Definition 15 (Function). f is an n-place function iff (i) it is relation on D (ii) and for any sequence $x_1, \ldots, x_n$ of objects from D, and any $y, z \in D$, if $\langle x_1, \ldots, x_n, y \rangle \in f$ and $\langle x_1, \ldots, x_n, z \rangle \in f$, then $y = z$ (i.e., each sequence of things stands in the relation to at most one thing).

A $n + 1$-place relation which meets this condition is an n-place function.

   If f is a unary function (i.e., a binary relation), we write $f(x) = y$ iff $\langle x, y \rangle \in f$. In this case we say x is the *argument* given to f, and y the *value* of f given that argument. The fact that, for a given argument, a function has a unique value, justifies our talking of '*the* value of the function for a given argument', and justifies the link to natural language definite descriptions like 'the capital of South Australia' as corresponding to a function with an argument – see the discussion of definite descriptions in chapter 8 for more on this issue.

Definition 16 (Domain). The *domain* (note: different to the domain of a relation) of a function f is the set

$$\mathscr{D}(f) = \{x : \text{there is a } y \text{ such that } \langle x, y \rangle \in f\}.$$

Definition 17 (Range). The *range* of f is the set

$$\mathscr{R}(f) = \{y : \text{there is a } x \text{ such that } \langle x, y \rangle \in f\}.$$

Definition 18 (Partial and Total). A function f is *partial* on X if $\mathscr{D}(f) \subset X$; it is *total* if $\mathscr{D}(f) = X$.

Operators

Definition 19 (Operator). If f is a function, and $\mathscr{R}(f) \subseteq \mathscr{D}(f)$, then f is an *operator*.

Consider the numerical function $\mathsf{sq} = \{\langle x, y \rangle : x^2 = y\}$, otherwise written $\mathsf{sq}(x) = x^2$. This generalises in the obvious way to binary functions: if the values of the function are appropriate arguments, it is an operator. An example is the addition function on the domain of integers, $\mathsf{plus} = \{\langle x, y, z \rangle : z = x + y\}$. This is an arithmetical operator.

Some properties of functions are conveniently illustrated by plus.

Definition 20 (Commutative). A binary function f is commutative iff

$$f(x, y) = f(y, x).$$

Definition 21 (Associative). A binary operator f is associative iff

$$f(x, f(y, z)) = f(f(x, y), z).$$

(We need f to be an operator to ensure that it is defined when its values are treated as arguments in this way.)

It is obvious that plus is both commutative and associative: $x + y = y + x$ and $x + (y + z) = (x + y) + z$. A commutative but not associative operator is $\mathsf{mean}(x, y) = \frac{x+y}{2}$. Clearly $\mathsf{mean}(x, y) = \mathsf{mean}(y, x)$. (This follows from the commutativity of plus.) But $\mathsf{mean}(x, \mathsf{mean}(y, z))$ is not in general equal to $\mathsf{mean}(\mathsf{mean}(x, y), z)$. An associative but non-commutative operator is the concatenation operator on strings $\frown$, which takes two strings and joins them into a longer string by appending the second to the first. Clearly $(x \frown y) \frown z = x \frown (y \frown z)$: 'te'$\frown$'a' = 'tea' = 't'$\frown$'ea'. But $x \frown y$ isn't necessarily $y \frown x$: 'tea' isn't 'ate'.

Properties of functions; enumerations Suppose f is a total function, and X and Y are some other sets.

Definition 22 (Into, Onto, etc.). **Into** f is a function from X *into* Y iff $\mathscr{D}(f) = X$ and $\mathscr{R}(f) \subseteq Y$.

Onto f is a function from X *onto* Y iff $\mathscr{D}(f) = X$ and $\mathscr{R}(f) = Y$. An *onto* function is also known as a *surjection*.

One-One f is a *one-one* function from X to Y iff f is into and whenever $x \neq y$, $f(x) \neq f(y)$. A one-one function is also known as an *injection*.

Bijection f is a *bijection* from X to Y (or, is a *one-one correspondence*) iff f is one-one and onto.

Some examples; these are functions from $\mathbb{N}$ to $\mathbb{N}$:

- The constant function $f(x) = 1$ is into.

- The function $g(x) = \begin{cases} x/2 & \text{if } x \text{ is even} \\ (x + 1)/2 & \text{if } x \text{ is odd} \end{cases}$ is onto. (Every natural number is the value of this function for two arguments.)

- The function $h(x) = x^3$ is one-one. (Every natural number has a unique cube, but not every natural number has a natural number cube root.)

- The identity function $i(x) = x$ is a bijection.

Definition 23 (Inverse). If g is a function, it is the *inverse* of f iff $\mathscr{D}(f) = \mathscr{R}(g)$, $\mathscr{R}(f) = \mathscr{D}(g)$, and for all x, $g(f(x)) = x$. In that case, we often write f^{-1} for g.

Theorem 9
If a function has an inverse, it is one-one.

Definition 24 (Partial function). f is a partial function from X to Y iff $\mathscr{D}(f) \subset X$. If $\mathscr{D}(f) = X$, f is a total function from X.

Theorem 10
If f is one-one but not a bijection, f^{-1} is a partial function.

2.3   Size

Definition 25 (Finite). A set X is finite iff there exists a set of natural numbers $N = \{1, \ldots, n\}$ and a bijection f from X to N.

A finite set can be put into one-one correspondence with some initial fragment of the sequence of natural numbers. A set is infinite iff it is not finite. Where there is a bijection between X and $\{1, \ldots, n\}$, X has n members – we say that X has cardinality n, $|X| = n$.

Now we will define some further notions useful for talking about *size*.

Definition 26 (Enumeration). f is an *enumeration* of X iff $\mathscr{D}(f) \subseteq \mathbb{N}$ (the natural numbers), $\mathscr{R}(f) = X$, and f is onto.

An enumeration associates each member of X with some number – it corresponds to the intuitive notion of counting a set. (Though since we only

require that an enumeration be onto, all sorts of weird 'counting' proced-
ures are included: e.g, the constant function from $\mathbb{N}$ onto $\{a\}$ is an enumer-
ation of $\{a\}$.) This gives us the following idea:

Definition 27 ([Un]Countable). X is *countable* iff there exists a function
which enumerates it; it is *denumerable* iff it is countable and infinite; it is
uncountable if it is not countable.

Obviously, all finite sets are enumerable, since if f is a bijection showing
X to be finite, f^{-1} exists and is an enumeration of X. But some infinite
sets are countable too: obviously, $\mathbb{N}$ itself (there is a no bijection between
$\mathbb{N}$ and some initial subset of $\mathbb{N}$, but the identity function enumerates it).

The notion of cardinality can be extended to infinite sets too (Machover,
1996: chs. 3, 6).

Definition 28 (Equinumerosity). X and Y are *equinumerous* iff there ex-
ists a bijection from X to Y.

Definition 29 (Cardinality). For each set, X – whether finite or infinite –
there exists its cardinality, $|X|$, meeting the following condition: for any X
and Y, $|X| = |Y|$ iff X and Y are equinumerous.

Certainly treating the natural numbers as cardinals of finite sets meets the
condition in Definition 29. But Definition 29 applies to any sets; it's just
that, for the infinite sets, we have no pre-theoretical grasp on what their
cardinalities might be. That's okay though: we can characterise them.

Definition 30 ($\leq$). If λ and μ are cardinals, such that $|X| = \lambda$ and $|Y| = \mu$,
then $\lambda \leq \mu$ iff there is an injection from X to Y.

Theorem 11
Let $|Y| = \mu$. Then $\lambda \leq \mu$ iff there exists an X such that $X \subseteq Y$ and $|X| = \lambda$.

Proof. Obvious: by Definition 30, $\lambda \leq \mu$ iff there is an injection from X to
Y iff there is a bijection from X to a subset of Y. $\qquad\square$

Theorem 12 (Schröder-Bernstein)

$\leq$ *partially orders the cardinals (Machover, 1996: 38–41).*

With the notion of a cardinal, we can prove that there are more sizes of set than finite sizes, because there are more cardinals than just the natural numbers, which are the cardinalities of finite sets. Since there is no bijection from any finite subset of natural numbers to the whole set of natural numbers, the cardinality of the set of natural numbers is not any finite natural number. We introduce $\aleph_0$ to name $|\mathbb{N}|$.

Theorem 13 (Cantor)

For any set X, $|X| < |\wp(X)|$.

Proof. Define a function f such that for all $x \in X$, $f(x) = \{x\}$. Since if $x \in X$, $\{x\} \subseteq X$, f is an injection from X into $\wp(X)$, and hence $|X| \leq |\wp(X)|$ by Definition 30.

To show that $|X| < |\wp(X)|$, we need to show that $|X| \neq |\wp(X)|$, i.e. (by Definition 29), that X and $\wp(X)$ are not equinumerous. Let g be any function from X to $\wp(X)$. The domain of g is the members of X, and the range are subsets of X, precisely the kinds of things members of X can be members of. But not every member of X needs to be a member of the subset of X picked out by $g(x)$. Let us define the set of such things, the $x \in X$ which aren't members of $g(x)$:

$$D = \{x \in X : x \notin g(x)\}.$$

Since D consists solely of members of X, $D \subseteq X$ and hence $D \in \wp(X)$. If g is a bijection, then there must be some $d \in X$ such that $g(d) = D$. Obviously, then $d \in D$ iff $d \in g(d)$. But by definition of D, $d \in D$ iff $d \notin g(d)$. So if g is a bijection, there is some d such that $d \in g(d)$ iff $d \notin g(d)$, and since there obviously is no such d, then g cannot be a bijection. Since g was arbitrary, *no* function from X to $\wp(X)$ is a bijection, and therefore X and $\wp(X)$ are not equinumerous. $\square$

Cantor's theorem shows us that the powerset of a set is always a strictly larger cardinality than the set. We already know that there is one countable infinite set, $\mathbb{N}$. Are there uncountable sets? Yes, obviously: $\wp(\mathbb{N})$: the set of all sets of natural numbers. This allows us to show that another set is uncountable: the real numbers, numbers with arbitrarily many decimal places. Indeed, we can show the following, from which it follows trivially that the set of reals is uncountable.

Theorem 14 (Uncountability of the Unit Interval)

The set of real numbers between 0 *and* 1 *inclusive,* [0, 1], *is uncountable.*

Proof. (Sketch.) Each subset of $\mathbb{N}$ can be associated with an infinite binary sequence (sequence of 0s and 1s): the sequence corresponding to $X \in \wp(\mathbb{N})$ is the one which has a 1 at position n iff $n \in X$. Each infinite binary sequence corresponds to exactly one real number in [0, 1]. So there is a bijection from $\wp(\mathbb{N})$ to [0, 1], so $|\wp(\mathbb{N})| = |[0, 1]|$, and since the former is uncountable, so is the latter. $\qquad\square$

A final theorem linking countability and set theory, useful – perhaps – for some exercise or other.

Theorem 15 (Countable Unions)

If $\mathbf{X} = \{X_1, \ldots, X_n, \ldots\}$ *is countable, and each* X_i *is countable, then* $\bigcup \mathbf{X}$ *is countable.*

I sketch the proof. Let P be the set of positive prime numbers. This set is in one-one correspondence with $\mathbb{N}$ (for every n, there is one and only one n-th prime). Since $\mathbf{X}$ is countable, there is a one-one function from $\mathbf{X}$ into the natural numbers, so there is a one-one function f from $\mathbf{X}$ into P.

Since each X_i is countable, there is a function g_i that is a one to one correspondence between X_i and a subset of $\mathbb{N}$. Define the function $h_i(x) = f(X_i)^{g_i(x)}$. Since every integer has a unique prime decomposition, if $h_i \neq h_j$ then $\mathscr{R}(h_i) \cap \mathscr{R}(h_i) = \varnothing$.

Now define: $h(x) = h_i(x)$ where $x \in X_i$ and x is not a member of any X_j where $j < i$. Clearly $\mathscr{D}(h) = \bigcup \mathbf{X}$. Obviously $\mathscr{R}(h) \subseteq \mathbb{N}$. And h is one-one, since each h_i is a function and their ranges are disjoint. So $\bigcup \mathbf{X}$ is in one-one correspondence with a subset of $\mathbb{N}$ so is countable.

Further Reading

Another presentation of the set theory needed for this book can be found in Beall and van Fraassen (2003). A presentation which goes well beyond what we need, but is philosophically nuanced, is Potter (2004). The iterative conception of a set is discussed by Boolos (1971).

Exercises

1. Prove:

    (a) If $X \subseteq Y$ and $Y \subseteq X$ then $X = Y$.

    (b) If $X \subseteq Y$ and $Y \subseteq Z$ then $X \subseteq Z$.

    (c) $A \not\subset A$.

    (d) If $A \subset B$ then $B \not\subset A$.

    (e) If $A \subset B$ then $A \subseteq B$.

2. Prove:

    (a) $A \cap A = A \cup A = A$.

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

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

    (d) $A \subseteq B$ iff $A \cap B = A$ iff $A \cup B = B$.

    (e) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

    (f) $A \setminus (A \cap B) = A \setminus B$.

    (g) $A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C)$.

3. Prove that if there is a universal set Ω, then the axiom of separation entails a contradiction.

4. Let us define a *Kuratowski ordered pair* $[\![x, y]\!]$ as the set $\{\{x\}, \{x, y\}\}$. Prove that

   (a) $[\![x, y]\!] = [\![u, v]\!]$ iff $x = u$ and $y = v$.

   (b) If $x = y$ then $[\![x, y]\!] = \{\{x\}\}$.

5. (a) Prove that if X has n members, $\wp(X)$ has 2^n members.

   (b) Prove that $X \subseteq Y$ iff $\wp(X) \subseteq \wp(Y)$.

   (c) Show, by providing a counterexample, that it is not always true that
   $\wp(X) \cup \wp(Y) = \wp(X \cup Y)$.

6. (a) Show that f has an inverse function iff it is one-one.

   (b) Give an example of a function that

      i. is into but is not onto;

      ii. is one-one but is not a bijection;

      iii. has no inverse;

      iv. is its own inverse.

   (c) Under exactly what conditions is the union of two functions itself a function? (I.e., state necessary and sufficient conditions.)

   (d) Two sets are said to be *equipollent* iff there is a one-one correspondence between them. Show that X and $\wp(X)$ are not equipollent.

   (e) Give an example of a set which is

      i. Countable;

      ii. Denumerable;

      iii. Uncountable. (Hint: use 6.(d) and 6.(e).ii.)

7. (a) Show that if X is countable, then if $Y \subseteq X$, Y is countable.

   (b) Show that if X and Y are both countable, $X \times Y$ is countable. (Hint: show that the set of ordered pairs of natural numbers is countable.)

Chapter 3

The Syntax of $\mathscr{L}_1$

3.1 Strings and Quotation

The Alphabet of $\mathscr{L}_1$ The language $\mathscr{L}_1$ contains an *alphabet*, consisting of the following three types of characters:

1. *Sentence letters $P, Q, R, P_1, Q_1, R_1, P_2, Q_2, R_2, \ldots.$*

2. *Logical connectives*: $\neg, \wedge, \vee, \rightarrow, \leftrightarrow.$[1]

3. *Parentheses*: $(,)$

No character falls into more than one category; that is, e.g., no sentence letter is also a connective. Also, no character in any category is identical to any other character in that category: e.g., '$\wedge$' is distinct from '$\neg$'.

Using this alphabet, we will now define the syntax, or grammar, of $\mathscr{L}_1$. That is, we will give rules which tell us which combinations of symbols form grammatical *sentences*. To do this, we will have to talk *about* sentence letters and sentences. Two devices of reference to strings will enable us to do this clearly.

---

[1]The notation I use for the logical connectives is standard, but there are a couple of other standards too. One very common variant is to use '&' for '$\wedge$'. Beall and van Fraassen (2003: 25) use '~' where we use '$\neg$', '$\supset$' where we use '$\rightarrow$', and '$\equiv$' where we use '$\leftrightarrow$'.

Definition 31 (String). A *string* in a language $\mathscr{L}$ is any finite ordered sequence of characters from the alphabet of $\mathscr{L}$.

I use lowercase Greek letters – φ, ψ, χ and so on – as *variables* for strings (see Appendix B).

There is a convenient way that our metalanguage – English – allows us to name strings. The convention is this. If φ is a string in $\mathscr{L}_1$, then the English expression consisting of left quotation mark, followed by the string φ, followed by the right quotation mark, is the name of φ (Richard, 1986: 397).[2]

Corner Quotes   A problem with this convention is that we are limited in our ability to quantify into quotation, which we will want to do in specifying our syntax. Here's an example to clarify what I mean by this. Suppose we wanted to be more precise about what a sentence letter is. We might say something like this:

Sentence Letters   A sentence letter is an instance of one of the following:
   'P_n', 'Q_n', 'R_n', for $n \geqslant 0$.

But this doesn't work: for, according to our convention, 'P_n' is the name of the string that contains a capital italic instance of the 16th letter of the alphabet, followed by a lowercase subscripted italic instance of the 14th letter of the alphabet. The 'n' that occurs in the string is being mentioned, not used, so is not a variable that can be governed by the condition '$n \geqslant 0$'. We can get around this, by introducing a device due to Quine (1940: §6): *corner quotes.*

Here's how they work. If φ is a string, the expression $\ulcorner \varphi \urcorner$ is the quotation name of that string – i.e., it consists of the left quote mark, the string

---

[2]Of course, 'φ' – the expression consisting of an instance of this lowercase letter of the Greek alphabet – is also being used, temporarily, to refer to that string, but it is a variable, not a name. Compare the pronoun 'him' with the name 'Antony', both of which can be used by you to refer to me, but the former only does so temporarily and under special circumstances.

itself, then the right quote mark. (So if φ is a string, $\ulcorner\varphi\urcorner$ is a name of the string, while 'φ' is the name of the variable that – temporarily – denotes the string.) Crucially, this construction allows variables to do their job: in forming $\ulcorner\varphi\urcorner$, we replace the variable by whatever it happens to denote, rather than talking about the variable itself. Now we can see how to fix up our definition of sentence letters:

Definition 32 (Sentence Letters). A symbol is a sentence letter iff it is an instance of one of the following: $\ulcorner P_n\urcorner$, $\ulcorner Q_n\urcorner$, or $\ulcorner R_n\urcorner$, for $n \geqslant 0$. For simplicity, we write 'P' for 'P_0'.

This can all seem a bit complicated, but we need to be precise if we are going to be sure that we can prove what we need to prove about our language. Pedantry is almost always a virtue in logic.

But one also needs to know when pedantry becomes the enemy of clarity. So in what follows, I'll often drop quotes and corner quotes when there is no chance of confusion. In particular, since the object language expressions of $\mathscr{L}_1$ are not ordinary English expressions, it won't be confusing in general if I write things like: 'P_{17} is a sentence letter' rather than the strictly correct '‘P_{17}’ is a sentence letter'.

3.2 Sentences

Making use of corner quotes and variables over strings of $\mathscr{L}_1$, we may now specify our *syntax*. We do it by specifying which things are the (grammatical) sentences of our language.

Definition 33 (Sentences of $\mathscr{L}_1$). The *sentences* of $\mathscr{L}_1$ are those strings in the smallest collection including all strings satisfying these two conditions:

1. All sentence letters are sentences of $\mathscr{L}_1$.

2. If φ and ψ are sentences of $\mathscr{L}_1$, then so are:

- $\ulcorner \neg \varphi \urcorner$;
- $\ulcorner (\varphi \wedge \psi) \urcorner$;
- $\ulcorner (\varphi \vee \psi) \urcorner$;
- $\ulcorner (\varphi \rightarrow \psi) \urcorner$; and
- $\ulcorner (\varphi \leftrightarrow \psi) \urcorner$.

To say that the sentences are strings in the *smallest collection* including all strings satisfying the condition means this: every string satisfying the conditions is in the smallest collection of them, and no extra strings are in it. So, in effect, we are saying: the sentences include those strings which meet the conditions, and no other string that doesn't meet the conditions is a sentence.

Some Useful Terminology

Definition 34 (Arity). The *arity* of a connective is the number of constituent sentences the connective requires to form a grammatical sentence.

'$\neg$' is called a *unary* connective; it makes a sentence when it is supplied with one sentence. '$\wedge$', '$\vee$', etc., are called *binary* connectives, because they make a sentence when supplied with two sentences, which they connect together (hence the name). $\mathscr{L}_1$ does not involve ternary or higher arity connectives.

Definition 35 (Scope). The *scope* of an occurrence of a connective in a sentence φ is the occurrence of the smallest subsentence of φ that contains this occurrence of the connective.

So in the sentence $((P \wedge Q) \vee R)$, the scope of $\wedge$ is $(P \wedge Q)$.

Definition 36 (Main Connective). The *main connective* of a sentence is the connective (if there is any) which is in the scope of no other; alternatively, whose scope is the whole sentence.

Syntax   The foregoing suffices to specify the syntax of our language $\mathscr{L}_1$. The language contains an alphabet, which includes a set of sentence letters, and a set of connectives, and some other punctuation. And it contains a *syntax*, that is, a set of rules specifying which strings constructed from the alphabet are well-formed sentences. A language contains a syntax, but to complete our specification of the language we will also need to talk about the meanings of sentences, or its *semantics*; we look at the semantics of $\mathscr{L}_1$ in chapter 4.

Abbreviatory Conventions   When writing sentences of $\mathscr{L}_1$, we will usually from now on drop parentheses in accordance with the following conventions (Halbach, 2010: §2.3). Note these are only conventional abbreviations: they form no part of the official syntax, but are just for our own convenience.

1. The *outermost* pair of parentheses of a sentence may be dropped; that is, we may write '$P \lor Q$' in place of the official '$(P \lor Q)$'.

2. If $\oplus$ is one of the binary connectives '$\lor$', '$\land$' of $\mathscr{L}_1$ (that is, '$\oplus$' is a variable over binary connectives), then we may write any sentence of the form $\ulcorner((\varphi \oplus \psi) \oplus \chi)\urcorner$ as $\ulcorner(\varphi \oplus \psi \oplus \chi)\urcorner$.

We may combine these conventions, dropping the outer parentheses to yield $\ulcorner\varphi \oplus \psi \oplus \chi\urcorner$. Be careful: the conventions are designed to ensure that one can get back the original sentence without ambiguity. Since $P \lor Q \lor R$ can is a legitimate abbreviation of $((P \lor Q) \lor R)$ in accordance with the conventions, it *cannot* be used to abbreviate $(P \lor (Q \lor R))$. That latter sentence can only be abbreviated to $P \lor (Q \lor R)$.

3.3   Proofs About Syntax

We can be quite precise about the syntax of $\mathscr{L}_1$. For an example of what we can show, consider this result:

Theorem 16 (Parenthesis Matching)

If φ is a formula of $\mathcal{L}_1$, then if φ contains any left parenthesis '(', then it terminates with a right parenthesis ')', which is paired with the leftmost left parenthesis in φ (Shapiro, 2013: §2, Theorem 1).

Proof. By clause (3) of Definition 33, every sentence is built up from the sentence letters using the subclauses of clause (2) of Definition 33. Sentence letters have no parentheses (as no sentence letter is also a parenthesis). Parentheses are introduced only in subclauses involving '∧', '∨', '→', and '↔', and each time they are introduced as a matched set. So at any stage in the construction of a sentence, the parentheses are paired off. Moreover, each newly introduced pair of parentheses will be the leftmost and rightmost characters in a sentence. Since the only complex sentences not formed from these parenthesis-introducing rules are formed by negating existing sentences, and negation introduces a new character only at the left. So any complex formula which contains any parenthesis terminates in a right parenthesis, which is matched to the leftmost left parenthesis. □

Now, this isn't particularly interesting on its own; it's fairly obvious from the rules for the construction of sentences. However, it's important for two reasons. First, it shows how to construct a proof in the metalanguage: a systematic argument that a certain claim – the theorem – is correct. This is a fairly informal proof of this sort, and things will get a bit more precise later in semester, but it is illustrative of the kind of thing we'll do. Second, we can use this theorem in proving further theorems. For example:

Theorem 17 (Prefix-free)

Let φ, ψ be nonempty strings, such that $\ulcorner \varphi\psi \urcorner$ (i.e., φ followed by ψ) is a sentence. Then φ is not *a sentence (Shapiro, 2013: §2, Theorem 5).*

Proof. Either φ contains at least one parenthesis, or it does not.

By Theorem 16, if φ contains any parentheses, its leftmost parenthesis is matched to the terminal parenthesis in ψ. Since every sentence has the

same number of left and right parentheses (convince yourself that this is true), and $\ulcorner \varphi\psi \urcorner$ is a sentence, and ψ is non-empty, that means φ must contain more left parentheses than right parentheses, and is therefore not a sentence.

So now suppose φ does not contain any parentheses. If it is a sentence, it must either be a sentence letter, or a negated sentence letter. If it is either of these categories, and yet $\ulcorner \varphi\psi \urcorner$ is a sentence, ψ must be empty; but it's not. So it cannot be a sentence.

So, either way, φ is not a sentence of $\mathscr{L}_1$. □

Informally, what Theorem 17 shows is that no sentence of $\mathscr{L}_1$ is an initial substring of any other sentence. The set of grammatical sentences is thus *prefix-free*; in this way it is like the set of well-formed phone numbers (in which no phone number is the prefix of any other) or ISBNs.

3.4 Proof By Induction on Complexity

An important sort of inductive argument is *induction on the complexity of sentences*.

Definition 37 (Complexity). The complexity of a sentence of $\mathscr{L}_1$ is the sum of the arities of the connectives occurring in it. (E.g., $\neg P$ has complexity 1, $\neg\neg(P \to Q)$ has complexity 4, etc.)

The complexity of a sentence letter is 0. Since one cannot guarantee that a sentence has two constituents of equal complexity, we must use the strong principle of induction if one wants to induce on complexity. (Also because the complexity of $\varphi \oplus \psi$, where $\oplus$ is some binary connective, is the complexity of φ, plus the complexity of ψ, plus 2.)

In effect, we already used proof by induction on complexity in the proof of Theorem 16: we showed that the atomic sentences had matching parentheses (the base case), and showed that if some constituents had matching

parentheses, then the complex sentence formed by applying the further clauses of the syntax added parentheses in the appropriate way (the induction step).

Theorem 18

Each sentence consists of a string of zero or more negation symbols concatenated with either a sentence letter or a sentence with a binary connective as its main connective.

Proof. We prove this by strong induction on complexity. Assume that the theorem holds of all subsentences of a given sentence φ, and we'll show it to hold of the whole sentence. There are two possibilities:

1. φ was formed by applying the negation clause of the formation rules, so $\varphi = \neg\psi$ for some ψ. By the induction hypotheses, ψ consists of a string of zero or more negations concatenated with either a sentence letter or a sentence with a binary main connective. φ is formed by attaching a negation sign to the front, so φ is a sentence with *one* or more negations concatenated with either a sentence letter or a sentence with a binary main connective – clearly this satisfies the theorem.

2. φ was formed by applying one of the binary connective clauses of the formation rules. It is thus a sentence with a binary connective as its main connective, prefixed by zero negation symbols, and hence the theorem holds of it. □

3.5 The Size of $\mathscr{L}_1$

How big is the language $\mathscr{L}_1$? That is: what is the size of its set of sentences $S_{\mathscr{L}_1}$? There are infinitely many finite strings constructed from the alphabet of $\mathscr{L}_1$ – we can see this because, obviously, the set of sentence letters $\mathbf{P} =$

$\{P_n : n \in \mathbb{N}\}$ is in one-one correspondence with the set of natural numbers, and hence is countably infinite. But $\mathbf{P} \subset S_{\mathcal{L}_1}$, so $S_{\mathcal{L}_1}$ must be infinite too. The question is: is the size of $S_{\mathcal{L}_1}$ countably or uncountably infinite?

Each sentence is finitely long – we can prove this by induction on complexity, since at each complexity stage, we only form finitely long sentences from previously formed finite sentences, and there is never a stage at which an infinite sentence is formed. So each sentence of $\mathcal{L}_1$ is a finite sequence. Moreover, while our alphabet is infinite (we are treating each sentence letter as a distinct syntactically indivisible entity, at least at our level of analysis),[3] it is nevertheless countable. We can set up some standard enumeration of the alphabet: perhaps associate the parentheses with 0 and 1, the connectives with numbers 2–6, and the sentence letters with the subsequent numbers.

This means of associating numbers with items in our alphabet is in effect a *code*, which turns strings of $\mathcal{L}_1$ into corresponding sequences of natural numbers. And it turns out this entails the set of sentences of $\mathcal{L}_1$ is only countably infinite:

Lemma 19 (Key Lemma)

There is a one-one correspondence between finite sequences of natural numbers and the set of natural numbers $\mathbb{N}$.

Proof. For each l, there are countably many finite sequences of length l. For we can map any given sequence to a unique natural number, as follows: if p_k is the k-th prime number, we map the sequence $\langle n_1, \ldots, n_l \rangle$ to the product of the first l primes, each prime p_i raised to the power of n_i:

$$\mathsf{code}(\langle n_1, \ldots, n_l \rangle) = p_1^{n_1} \times \ldots \times p_l^{n_l}.$$

---

[3]We need not do this, of course: we could use the alphabet consisting of P, Q, R, and the digits $0, \ldots, 9$ to treat the sentence letters themselves as finite strings drawn from a finite alphabet. But as we'll see, this doesn't make any difference to the size of the set of sentences of $\mathcal{L}_1$.

Since the coding function code is one-one, by the fundamental theorem of arithmetic (Gowers, 2008: §V.14), it has an inverse, which is from a the natural numbers *onto* the set of finite sequences of length l, which is thus an enumeration, and hence the latter set is countable.

The set of all finite sequences of natural numbers is the union of every set of finite sequences of natural numbers of fixed length. (I.e., it is the union of: the set of such sequences of length 1, the set of such sequences of length 2, ...) There are countably many such sets (because finite lengths are natural numbers). So the set of all finite sequences is a countable union of countable sets; by Theorem 15, the set of all finite sequences of natural numbers is countable. □

By the Key Lemma, and the fact that we can model strings of $\mathscr{L}_1$ as finite sequences of natural numbers, there is a one-one correspondence between arbitrary finite strings of the alphabet of $\mathscr{L}_1$ and a countable set. Since the set of sentences of $\mathscr{L}_1$ $S_{\mathscr{L}_1}$ is a strict subset of the set of finite strings of the alphabet of $\mathscr{L}_1$ the cardinality of $S_{\mathscr{L}_1}$ is no greater than countable (by Theorem 9), and since it is infinite, it must therefore be countably infinite.

Exercises

1. Shapiro, 2013: §2:

    (a) Prove that if φ is a non-atomic $\mathscr{L}_1$ sentence, then there is exactly one formation clause from Definition 33 that could have been applied to existing sentences to produce φ.

    (b) Using this result, show that every sentence of $\mathscr{L}_1$ is *uniquely readable* – that is, each sentence can be produced from the sentence letters in accordance with the formation clauses in exactly one way. *(Hint: use the fact that no character is identical to any other. You'll need to show that it is not the case that, e.g., $(\varphi \vee \psi)$ is the same sentence as $(\chi \wedge \xi)$.)*

2. Sentences of $\mathscr{L}_1$ can be modelled as sequences of elements from $\mathscr{SL} \cup \{(,), \neg, \wedge, \vee, \rightarrow, \leftrightarrow\}$. Using Theorem 15 and Exercise 7 from chapter 2, show that the set $\mathscr{S}$ of sentences of $\mathscr{L}_1$ is countable.

Chapter 4

The Semantics of $\mathscr{L}_1$

4.1 Semantics for $\mathscr{L}_1$

To complete our specification of $\mathscr{L}_1$, we need to provide some account of the intended meanings of sentences of $\mathscr{L}_1$. In natural languages each expression has a specific meaning, and makes a specific contribution to the sentences in which it appears. Our language will differ from natural language in this respect. For we will not insist on any one fixed interpretation of the sentence letters of the language.[1] However, the meanings we assign to the logical connectives will resemble more closely the kinds of meanings in natural languages: the meanings of connectives will make a fixed contribution, alongside the syntax, to determining the meanings of complex sentences.

Truth Values   The meanings we assign to sentences in $\mathscr{L}_1$ are very minimal: they are *truth values.* An interpretation of the language will associate

---

[1] We resist the assignment of any fixed meaning to the sentence letters for practical reasons. If we wish to use our logical languages to model natural language arguments, it would be tedious indeed to have to figure out which sentence letter translates the natural language sentences we are considering. (Is it P_{371}?) So we allow the meaning of the sentence letters to be fixed anew in each application.

with every sentence of $\mathscr{L}_1$ a truth value such that the truth-value of a sentence depends on its semantically relevant structure and the truth-values of its constituents. Such an interpretation is called a *compositional semantics*: it identifies the meaning of a sentence with the conditions under which the sentence is true, and specifies those conditions by showing how the semantic values (truth values) of a complex sentence are determined by its structure and the semantic values of its parts.

Classical Valuations   In classical logic, there are two truth values: T and F (sometimes written 1 and 0, sometimes written $\top$ and $\bot$). We may say, then, that our semantics will provide a classical valuation of all the sentences in $\mathscr{L}_1$.

Definition 38 (Valuation).   A *valuation* of a syntax is an assignment of values (of some sort of other) to the sentences of that syntax. A *classical valuation* is a valuation in which the values are either T or F, and in which each sentence has exactly one value assigned to it.

We can express what we said about classical valuations more concisely using the language of functions: a classical valuation is a total function from the set of sentences of a language into the set of truth values. We may also use it to define a new notion:

Definition 39 ($\mathscr{L}_1$-Structure).   An $\mathscr{L}_1$-*structure* is a (total) function from the set of sentence letters of $\mathscr{L}_1$ into the set of classical truth values $\{T, F\}$.

A valuation need not be compositional, in the sense that there need not be any rules governing which sentences get which truth values. But ours will be, and we can use the rules governing the construction of complex sentences to extend an $\mathscr{L}_1$ structure to a full classical valuation for $\mathscr{L}_1$.

Definition 40 ($\mathscr{L}_1$-valuation).   When $\mathscr{A}$ is any $\mathscr{L}_1$-structure, $|\cdot|_{\mathscr{A}}$ is a valuation function from the set of $\mathscr{L}_1$ sentences to the set of truth values $\{T, F\}$ iff it meets these conditions:

1. If φ is a sentence letter, $|\varphi|_{\mathcal{A}} = \mathcal{A}(\varphi)$.

2. $|\neg\varphi|_{\mathcal{A}} = T$ if and only if (iff) $|\varphi|_{\mathcal{A}} = F$.

3. $|\varphi \wedge \psi|_{\mathcal{A}} = T$ iff $|\varphi|_{\mathcal{A}} = T$ and $|\psi|_{\mathcal{A}} = T$.

4. $|\varphi \vee \psi|_{\mathcal{A}} = T$ iff $|\varphi|_{\mathcal{A}} = T$ or $|\psi|_{\mathcal{A}} = T$ (or both).

5. $|\varphi \rightarrow \psi|_{\mathcal{A}} = T$ iff $|\varphi|_{\mathcal{A}} = F$ or $|\psi|_{\mathcal{A}} = T$ (or both).

6. $|\varphi \leftrightarrow \psi|_{\mathcal{A}} = T$ iff $|\varphi|_{\mathcal{A}} = |\psi|_{\mathcal{A}}$.

Definition 41 (Agreement of Structures). When S is a set of $\mathcal{L}_1$ sentence letters, let us say that two $\mathcal{L}_1$ structures $\mathcal{A}$ and $\mathcal{B}$ *agree on S* iff for each φ in S, $\mathcal{A}(\varphi) = \mathcal{B}(\varphi)$.

Since the value of a sentence is determined by the values of its constituents, it is obvious that if ψ is any $\mathcal{L}_1$ sentence, if two $\mathcal{L}_1$ structures $\mathcal{A}$ and $\mathcal{B}$ agree on the sentence letters in ψ, then $|\psi|_{\mathcal{A}} = |\psi|_{\mathcal{B}}$.

Falsity clauses   Here is another example of proof by induction on complexity of sentences. Recall the definition of a valuation. It told us under what circumstances a sentence of a given form was true in a valuation. Why didn't we also need to state when a sentence was false? Because a sentence is false iff it is not true:

Theorem 20 (Falsity is Untruth)

$|\varphi|_{\mathcal{A}} = 0$ *iff* $|\varphi|_{\mathcal{A}} \neq 1$.

Proof. Base case: φ is a sentence letter. Because $\mathcal{A}$ is a function, if $|\varphi|_{\mathcal{A}} = 0$ then $|\varphi|_{\mathcal{A}} \neq 1$. Because $\mathcal{A}$ is into and total, if $|\varphi|_{\mathcal{A}} \neq 1$ then $|\varphi|_{\mathcal{A}} = 0$.

Induction step: Suppose φ is a sentence, but is complex, and that the theorem holds for the constituents of φ. We show two illustrative cases:

1. Suppose $\varphi = \neg\psi$. Then $|\varphi|_{\mathcal{A}} = 0$ iff $|\neg\psi|_{\mathcal{A}} = 0$, iff $|\psi|_{\mathcal{A}} = 1$, iff $|\psi|_{\mathcal{A}} \neq 0$ iff $|\neg\psi|_{\mathcal{A}} \neq 1$ iff $|\varphi|_{\mathcal{A}} \neq 1$.

2. Suppose $\varphi = (\psi \vee \chi)$. Then $|\varphi|_{\mathscr{A}} = 0$ iff $|\psi \vee \chi|_{\mathscr{A}} = 0$ iff $|\psi|_{\mathscr{A}} = 0$ and $|\chi|_{\mathscr{A}} = 0$ iff $|\psi|_{\mathscr{A}} \neq 1$ and $|\chi|_{\mathscr{A}} \neq 1$ iff $|\psi \vee \chi|_{\mathscr{A}} \neq 1$ iff $|\varphi|_{\mathscr{A}} \neq 1$. $\square$

4.2 Truth Tables

One way of representing $\mathscr{L}_1$ structures – or at least, representing as much of them as matters – is using a truth table. Consider any finite set of $\mathscr{L}_1$ sentences, Γ. Since each sentence in Γ is also only finitely long, there are at most finitely many sentence letters occurring in Γ. The values assigned to each of these sentence letters, in accordance with Definition 40, determines the values assigned to every sentence in Γ.

Since there are only finitely many sentence letters in Γ, each of which is assigned either T or F by our structures, it is clear there are only finitely many ways of assigning truth values to the sentence letters. (In fact, if there are n sentence letters, there are 2^n ways of assigning them truth values.) So we can write down – in principle – each of those ways of assigning truth values in a *truth table*. This table summarises the truth values of the sentences in Γ across all possible $\mathscr{L}_1$-structures.

Constructing Truth Tables   For each sentence letter that appears in some sentence within Γ, inscribe a column. For each way of assigning truth values to sentence letters, inscribe a row, by putting a truth value under the appropriate column. It is obvious that each row of the truth table corresponds to a class of structures: namely, a class of structures that all agree on the sentence letters in Γ. (They will disagree over other sentence letters, but since they are not relevant to determining the truth value of any sentence in Γ, who cares?) For any way of assigning truth-values to sentence letters in Γ, there is some row of the truth table that agrees with that assignment. Now add a column for each sentence in Γ, and for each row, inscribe under

P	Q	$\neg P$	$(P \wedge Q)$	$(P \vee Q)$	$(P \rightarrow Q)$	$(P \leftrightarrow Q)$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

Table 4.1: Truth Table for the Standard Connectives

the sentence the value assigned to that sentence by any structure corresponding to that row. (Since all such structures agree on the sentence letters in the sentence, they will all agree on what they assign to the sentence too, so it doesn't matter which one we pick.)

So, for instance, suppose Γ is this interesting set of $\mathscr{L}_1$ sentences:

$$\{\neg P, (P \wedge Q), (P \vee Q), (P \rightarrow Q), (P \leftrightarrow Q)\}.$$

The truth table for this set of sentences is pictured in Table 4.1.

4.3 Satisfaction, Entailment, and other Semantic Notions

Many conceptions of logic have it that logic is fundamentally about *consequence*: what it is for some sentences to follow from one another, in virtue of the logical form of the sentences involved (Beall and Restall, 2013). But what is consequence? We may characterise it semantically as follows: some sentences of $\mathscr{L}_1$ have another sentence as a consequence if the former sentences *entail* the latter sentence. And we may characterise entailment, and a number of other semantic relations between sentences, using the notions we've already introduced.

Definition 42 (Satisfaction). Suppose Γ is a (possibly empty, possibly infinite) set of sentences of $\mathscr{L}_1$, and $\mathscr{A}$ is a $\mathscr{L}_1$-structure, such that $|\gamma|_{\mathscr{A}} = T$

for every sentence $\gamma \in \Gamma$. In that case, $\mathscr{A}$ *satisfies* Γ, or $\mathscr{A}$ is a *model* of Γ.

Definition 43 (Entailment). A set of sentences Γ *(semantically) entails* a sentence φ iff every $\mathscr{L}_1$-structure which satisfies Γ also satisfies φ. Notation: $\Gamma \vDash \varphi$.

Definition 44 (Tautology). φ is a *tautology* iff every $\mathscr{L}_1$ structure satisfies $\{\varphi\}$. A tautology is sometimes called a *logical truth*.

Theorem 21
If φ is a tautology, then for any set of sentences Γ, $\Gamma \vDash \varphi$ – even if Γ is the empty set, which contains no sentences at all.

This theorem explains the fact that we often write '$\vDash \varphi$' to mean that φ is a tautology.

If no $\mathscr{L}_1$-structure satisfies Γ, Γ is *unsatisfiable*, or *semantically inconsistent*, which we write $\Gamma \vDash$. (Accordingly, a set of sentences is *semantically consistent* iff it is satisfiable.) If $\Gamma = \{\varphi\}$ and $\Gamma \vDash$, φ is a *contradiction*.

Theorem 22
φ is a tautology iff $\neg\varphi$ is a contradiction.

Proof. φ is a tautology iff $|\varphi|_{\mathscr{A}} = T$ in every $\mathscr{L}_1$ structure $\mathscr{A}$. By Definition 40, this is the case iff $|\neg\varphi|_{\mathscr{A}} = F$ in every $\mathscr{L}_1$ structure $\mathscr{A}$; iff no $\mathscr{L}_1$ structure satisfies $\{\neg\varphi\}$, i.e., $\neg\varphi$ is a contradiction. $\qquad\square$

Validity and Necessity An *argument* involves a set of sentences Γ, its *premises*, and a single sentences φ as its conclusion. An argument is *(logically) valid* iff Γ entails φ.

Sometimes one sees a purported definition of validity along these lines: that an argument is valid iff it is *impossible* for its premises all to be true while its conclusion is false (Henle, Garfield, and Tymoczko, 2011: 19). That is, if there is no possible situation in which the premises are true and the conclusion false. While there are some similarities between $\mathscr{L}_1$-structures

and possible situations, this sort of definition of validity should be resisted. We can think of a possible situation as specified by a description of a *way things could have been*. Importantly, the description will be in a certain actual language, which we need to hold fixed in order for it to do its descriptive job. But an $\mathscr{L}_1$-structure is, in some ways, precisely the opposite of this. It is a way of interpreting the sentence letters of $\mathscr{L}_1$, holding fixed the world at which those reinterpreted sentences are to be evaluated. In many cases it makes no difference whether we consider a sentence P to be evaluated at a possible situation in which what P says obtains, or to be evaluated at actuality under an interpretation which makes it true. But there are cases in which it matters. Consider this argument:

> Sylvester is a child;
> Therefore, Sylvester is not an adult.

The premise *necessitates* the conclusion, for there is no possible world in which a child is an adult. But this argument isn't logically valid; for under a reinterpretation of the non-logical vocabulary 'adult' on which it means 'child', the premise is actually true and the (reinterpreted) conclusion actually false. The premise guarantees the truth of the conclusion, in part because of what the non-logical expressions 'adult' and 'child' actually mean. But in logic, we are interested in arguments where the premises guarantee the truth of the conclusion in virtue only of the meanings of the logical expressions involved (Tarski, 1936).

4.4 Entailment and the Connectives

Theorem 23 (Conjunction and Satisfaction)
If $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$, then for every $\mathscr{L}_1$ structure $\mathscr{A}$, $\mathscr{A}$ satisfies Γ iff $\left| (\gamma_1 \wedge \ldots \wedge \gamma_n) \right|_{\mathscr{A}} = T$.

Definition 45 (Equivalence). If $\varphi \vDash \psi$ and $\psi \vDash \varphi$, φ and ψ are *logically equivalent*.

Theorem 24 (Equivalence and Biconditional)
φ and ψ are logically equivalent iff $\vDash \varphi \leftrightarrow \psi$.

Proof. $\varphi \vDash \psi$ iff every structure in which φ has the value T is one in which ψ also has the value T, and *vice versa*. That is, iff in every structure $\mathscr{A}$, $|\varphi|_{\mathscr{A}} = |\psi|_{\mathscr{A}}$. That holds in turn, by Definition 40, iff $|\varphi \leftrightarrow \psi|_{\mathscr{A}} = T$ in every structure, and $\varphi \leftrightarrow \psi$ is a tautology. □

Deduction Theorem   Let the notation '$\Gamma, \varphi \vDash \psi$' abbreviate $\Gamma \cup \{\varphi\} \vDash \psi$.

Theorem 25 (Deduction)
$\Gamma \vDash (\varphi \rightarrow \psi)$.

Proof. $\Gamma, \varphi \vDash \psi$ iff every $\mathscr{L}_1$-structure which satisfies $\Gamma \cup \{\varphi\}$ also satisfies $\{\psi\}$. That holds iff there is no structure $\mathscr{A}$ in which Γ is satisfied and in which $|\varphi|_{\mathscr{A}} = T$ and $|\psi|_{\mathscr{A}} = F$. By Definition 40, there is no structure $\mathscr{A}$ in which Γ is satisfied while $|\varphi \rightarrow \psi|_{\mathscr{A}} = F$, i.e., $\Gamma \vDash \varphi \rightarrow \psi$. □

This permits us to say that an argument $\gamma_1, \ldots, \gamma_n \vDash \varphi$ is valid iff $\vDash ((\gamma_1 \wedge \ldots \wedge \gamma_n) \rightarrow \varphi)$ – i.e., that an argument with finitely many premises is valid iff the corresponding conditional is a tautology.

Corollary 26
An argument from premises Γ to conclusion φ is valid iff the conditional with the conjunction of the premises in Γ as antecedent, and φ as consequent, is a tautology. That is $\Gamma \vDash \varphi$ iff $\vDash (\gamma_1 \wedge \ldots \wedge \gamma_n) \rightarrow \varphi$.

Proof. Obvious from Theorem 23 and Theorem 25. □

4.5   Structural Rules

The expression '$\Gamma \vDash \varphi$' is known as a *sequent*. This sequent is correct if Γ does entail φ. There are a set of rules, each of which can be justified from the definitions above, that governs correctness-preserving transformations

of one sequent into another. These rules are known as *structural rules*. Theorems justifying three standard structural rules follow (where Γ, Δ are sets of premises):

Theorem 27 (Permutation)

$\Gamma, \psi, \chi, \Delta \vDash \varphi$ *iff* $\Gamma, \chi, \psi, \Delta \vDash \varphi$ *(premise order doesn't matter).*

Theorem 28 (Contraction)

$\Gamma, \psi, \psi, \Delta \vDash \varphi$ *iff* $\Gamma, \psi, \Delta \vDash \varphi$ *(duplicate premises don't matter).*

Theorem 29 (Weakening)

If $\Gamma \vDash \varphi$*, then* $\Gamma, \psi \vDash \varphi$*.*

Proof. Note that every structure which satisfies all of Γ and also $\{\psi\}$ also satisfies Γ; as every structure satisfying Γ satisfies φ, every structure satisfying Γ, ψ satisfies φ. □

The structural rules are so-called because they don't depend on the meaning of any connectives, but only on the definition of entailment and the underlying set theory governing the behaviour of sets of premises. Interesting non-classical logics can arise when one varies the structural rules, which may even be done while retaining classical valuations (Restall, 2000).

Cut, Transitivity and Contraposition

Theorem 30 (Cut)

If $\Gamma, \psi \vDash \varphi$ *and* $\Gamma \vDash \psi$*, then* $\Gamma \vDash \varphi$*.*

Proof. Assume that $\Gamma, \psi \vDash \varphi$. If every structure in which Γ is satisfied is one in which $\{\psi\}$ is satisfied, then it is clear that the set of structures which make Γ, ψ true is just the set of structures which makes Γ true. Given that all the former satisfy φ, so must all the latter: $\Gamma \vDash \varphi$. □

Theorem 31 (Transitivity)

If $\Gamma \vDash \psi$ *and* $\psi \vDash \varphi$*, then* $\Gamma \vDash \varphi$*.*

Theorem 32 (Contraposition)

$\varphi \vDash \psi$ *iff* $\neg\psi \vDash \neg\varphi$*.*

4.6 Substitution

Formality of Logic   The notions of logical validity and entailment in $\mathscr{L}_1$ are *formal*: it is in virtue of the form of the premises that they entail the conclusion of a valid argument. But what does it mean to say that logic is formal? Here's one thing it could mean: take a sentence of the language, and replace some of its constituents with others of the same category, while preserving the logical structure of the sentence. These two sentences have the same form. If we can prove some sort of equivalence between the two sentences, we will have shown that what matters in the language is form, rather than the particular constituents of a given sentence.

This is something you will have relied on implicitly in earlier logic courses: how could it be that for example P and $P \rightarrow Q$ entail Q, yet R and $R \rightarrow P_1$ fail to entail P_1? But we should prove it explicitly, to make sure that our implicit assumptions aren't leading us astray.

In the case of $\mathscr{L}_1$, the substituable constituents are sentence letters, and the logical structure is given by the logical connectives.

Definition 46 (Uniform Substitution). If Γ is a set of sentences, θ a sentence, and s a sentence letter, then write $\Gamma[\theta/s]$ for the set of sentences that results by *uniformly substituting* θ for every occurrence of s in every sentence in Γ. (If s doesn't occur in Γ, then $\Gamma[\varphi/s] = \Gamma$.) Likewise, let $\varphi[\theta/s]$ be the sentence that results from replacing every occurrence of s in φ by θ.

Definition 47 (Substitution Instance). If $\varphi = \psi[\theta/s]$, for some θ, s, then φ is called a *substitution instance* of ψ.

We now show a useful lemma.

Lemma 33 (Substitution)

Suppose φ and θ are sentences, and s a sentence letter. For any structure $\mathscr{A}$,

define a structure:

$$\mathscr{A}^\star(\alpha) = |\alpha[\theta/s]|_\mathscr{A} = \begin{cases} \mathscr{A}(\alpha) & \text{iff } \alpha \neq s \\ |\theta|_\mathscr{A} & \text{iff } \alpha = s. \end{cases} \boxtimes$$

Then $|\varphi|_{\mathscr{A}^\star} = |\varphi[\theta/s]|_\mathscr{A}$.

Proof. The new structure $\mathscr{A}^\star$ is just like $\mathscr{A}$, with the possible exception that it assigns to s what $\mathscr{A}$ assigns to θ.

Base case: If φ is a sentence letter, then the result follows by construction of $\mathscr{A}^\star$.

Induction step: If φ is complex, and the theorem holds for less complex claims, then the result follows. I show one case: where φ is a conjunction of two simpler constituents. The induction hypothesis is that $|\varphi_i|_{\mathscr{A}^\star} = |\varphi_i[\theta/s]|_\mathscr{A}$ where φ_i are constituents of φ. By the semantic rules for the valuation function, and the induction hypothesis,

$$
\begin{aligned}
|\varphi|_{\mathscr{A}^\star} = T \text{ iff } & \left|(\varphi_1 \wedge \varphi_2)\right|_{\mathscr{A}^\star} = T \\
\text{iff } & |\varphi_1|_{\mathscr{A}^\star} = |\varphi_2|_{\mathscr{A}^\star} = T \\
\text{iff } & \left|\varphi_1[\theta/s]\right|_\mathscr{A} = \left|\varphi_2[\theta/s]\right|_\mathscr{A} = T \\
\text{iff } & \left|\varphi_1[\theta/s] \wedge \varphi_2[\theta/s]\right|_\mathscr{A} = T \\
\text{iff } & \left|(\varphi_1 \wedge \varphi_2)[\theta/s]\right|_\mathscr{A} = T.
\end{aligned}
$$

The desired result follows: $|\varphi|_{\mathscr{A}^\star} = |\varphi[\theta/s]|_\mathscr{A}$. *Mutatis mutandis* for the other connectives. □

With this lemma in hand, we can establish further results about substitution.

Theorem 34 (Substitution of Material Equivalents)

If $|\theta|_\mathscr{A} = \mathscr{A}(s)$, *then* $|\varphi|_\mathscr{A} = |\varphi[\theta/s]|_\mathscr{A}$.

Proof. If $|\theta|_\mathscr{A} = \mathscr{A}(s)$, then $\mathscr{A} = \mathscr{A}^\star$; applying Theorem 33, the result follows immediately. □

Theorem 35 (Substitution of Sentence Letters)

If Γ entails φ, then so too $\Gamma[\theta/s] \vDash \varphi[\theta/s]$, for any θ, s.

Proof. Now suppose for *reductio* that $\Gamma \vDash \varphi$, but $\Gamma[\theta/s] \nvDash \varphi[\theta/s]$. In that case, there is a structure $\mathscr{B}$ such that for each $\gamma \in \Gamma$, $|\gamma[\theta/s]|_{\mathscr{B}} = T$, but $|\varphi[\theta/s]|_{\mathscr{B}} = F$.

By the Substitution theorem 33, there is therefore a structure $\mathscr{B}^{\star}$ such that for each $\gamma \in \Gamma$, $|\gamma|_{\mathscr{B}^{\star}} = T$ and $|\varphi|_{\mathscr{B}^{\star}} = F$. But in that case, $\Gamma \nvDash \varphi$ after all, since there is an $\mathscr{L}_1$ structure where the premises are true and the conclusion false, contradicting our *reductio* supposition, which must have been wrong. □

Generalised Substitition and Formal Logic   A generalisation of Theorem 35 is also provable. The notation '$\varphi[\beta/\alpha][\gamma/\beta]$' is to represent the uniform substitution, first, of β for α throughout φ, and *then* the substitution of γ for β throughout the resulting formula. It is clear that $\varphi[\beta/\alpha][\gamma/\beta] = \varphi[\gamma/\alpha]$, as long as β didn't already occur as a subsentence of φ.

Theorem 36 (General Substitution)

When Γ is finite, if $\Gamma \vDash \varphi$, then also $\Gamma[\theta/\chi] \vDash \varphi[\theta/\chi]$, for any θ, χ.

Proof. Suppose for *reductio* that $\Gamma \vDash \varphi$, but there is a sentence letter s *not occurring in Γ or φ* such that $\Gamma[s/\chi] \nvDash \varphi[s/\chi]$. There must be a structure $\mathscr{B}^{\star}$ such that for each $\gamma \in \Gamma$, $|\gamma[s/\chi]|_{\mathscr{B}^{\star}} = T$, but $|\varphi[s/\chi]|_{\mathscr{B}^{\star}} = F$. By theorem 33, there is therefore a structure $\mathscr{B}$ such that for each $\gamma \in \Gamma$, $|\gamma|_{\mathscr{B}} = T$ and $|\varphi|_{\mathscr{B}} = F$.[2] But in that case, $\Gamma \nvDash \varphi$ after all, contradiction: so our supposition must have been wrong. So in fact for any sentence letter s not occurring in Γ or φ, when $\Gamma \vDash \varphi$, $\Gamma[s/\chi] \vDash \varphi[s/\chi]$. But since there are infinitely many sentence letters and each of the finitely many members of Γ and φ contains only finitely many sentence letters, we may always find such a 'new' s.

---

[2]This is because for any sentence γ in which s does not occur, $\gamma = \gamma[s/\chi][\chi/s]$.

Now we may apply Theorem 35 to show that $\Gamma[s/\chi][\theta/s] \vDash \varphi[s/\chi][\theta/s]$. But this latter sequent just is $\Gamma[\theta/\chi] \vDash \varphi[\theta/\chi]$. □

Theorem 36 establishes the formality of logic.[3] We propose that φ shares a logical form with ψ if there exist χ, ξ such that $\psi = \varphi[\chi, \xi]$.[4] Any valid argument, when we uniformly and systematically substitute one constituent sentence for another throughout, will transform into another argument of the same logical form; and that second argument will also be valid. Since $P, P \rightarrow Q \vDash Q$, accordingly so too $R, (R \rightarrow \neg R) \vDash \neg R$, which is a substitution instance of the former argument. But as this example shows, while we retain validity, we need not retain *soundness*. Soundness is not a matter of form, but the particular interpretation of a sentence of a given form.

Equivalence Revisited   There is one case where soundness is *guaranteed* to be preserved: if what is substituted for one another are logically equivalent sentences.

Theorem 37 (Equivalents)

If $\Gamma \vDash \varphi$ is sound, then so is any sequent $\Gamma[\theta/\chi] \vDash \varphi[\theta/\chi]$ where θ and χ are logically equivalent.

Proof. Left for exercise. □

Define a *subsentence* of φ as any sentence which occurs within φ (i.e., if one decomposes the sentence successively in accordance with the syntax, every subsentence will appear at some stage).

---

[3]The reasoning is actually a little subtle: what happens when Γ is infinite? Then we cannot be assured that there is some new sentence letter occurring in neither φ nor Γ. The Compactness Theorem 48, which we will prove in chapter 5, assures us that: whenever $\Gamma \vDash \varphi$, then there is some finite subset Γ^- such that $\Gamma^- \vDash \varphi$, and then we can apply Theorem 36. But don't worry about this wrinkle until then; we rarely come across arguments with infinitely many premises in everyday life....

[4]So $P \wedge Q$ shares a logical form with $P \wedge P$ – but not *vice versa*, since uniform substitution into $P \wedge P$ cannot produce $P \wedge Q$. We could get sharing of logical form to be a symmetric relation if we liberalised the notion of substitution to include substitutions of some but not all instances of a particular subsentence.

Theorem 38 (Equivalence)

If φ is a subsentence of χ, and $\chi' = \chi[\psi/\varphi]$. Then $\varphi \leftrightarrow \psi \vDash \chi \leftrightarrow \chi'$.

Proof. Base case: Suppose $\chi = \varphi$. Then $\chi' = \psi$, and obviously $\varphi \leftrightarrow \psi \vDash \varphi \leftrightarrow \psi$.

Induction step: Suppose χ is complex, e.g., $\chi = \neg\gamma$, and $\varphi \leftrightarrow \psi \vDash \gamma \leftrightarrow \gamma'$. For any structure, $|\alpha \leftrightarrow \beta|_{\mathscr{A}} = |\neg\alpha \leftrightarrow \neg\beta|_{\mathscr{A}}$, so $\varphi \leftrightarrow \psi \vDash \neg\gamma \leftrightarrow \neg(\gamma')$. But since $\neg\gamma = \chi$, and $\neg(\gamma') = (\neg\gamma)' = \chi'$, the result follows. (The other cases are left for exercises.) $\qquad\square$

4.7 Truth-functions

Definition 48 (Truth Function). A *truth function* f_n is any total function from ordered sequences of n truth values into the set of truth values $\{T, F\}$.

Suppose S is a set of n sentence letters, $\{s_1, \ldots, s_n\}$. Since every $\mathscr{L}_1$ structure $\mathscr{A}$ assigns truth values to these sentence letters, we can understand an n-place truth function f_n as yielding a truth value given a set of sentence letters in a given structure:

$$f_n\left(\left|s_1\right|_{\mathscr{A}}, \ldots, \left|s_n\right|_{\mathscr{A}}\right) \in \{T, F\}.$$

Since every structure agrees in assignment of truth values with one row of a truth table for sentence letters in S, an n-place truth function can be thought of as a function from rows of the truth table for S to truth values. But, of course, a *column* of a truth table precisely represents a function from rows to truth values!

Expression

Definition 49 (Expression). A sentence φ *expresses* a truth function f_n iff

at least the sentence letters $s_1, \ldots, s_n$ occur in φ,[5] and for every $\mathscr{L}_1$ structure $\mathscr{A}$,

$$f\left(\left|s_1\right|_{\mathscr{A}}, \ldots, \left|s_n\right|_{\mathscr{A}}\right) = f(\mathscr{A}(s_1), \ldots, \mathscr{A}(s_n)) = \left|\varphi\right|_{\mathscr{A}}.$$

Every $\mathscr{L}_1$ sentence expresses a truth-function. (The converse principle, that every truth-function is expressed by some sentence, will be proved in chapter 5 when we talk about *expressive adequacy*.) For the rules on valuations (Definition 40) assign every sentence a truth value in every structure, so the valuation rules determines a function f_φ from structures to truth values for each sentence φ. We've already seen some sentences which express interesting truth functions above: those sentences expressing the truth functions associated with the connectives of $\mathscr{L}_1$ (Table 4.1).

How Many Truth Functions?

Theorem 39

There are 2^{2^n} n-place truth functions.

Proof. There are 2^n distinct sequences of truth values of length n, each of which is an input to an n-place truth function. Each of these sequences can be given T or F as its value. If one function assigns a different value to a given input than another function, they are different functions. So there are at least 2^{2^n} different truth functions (one for each way of assigning Fs and Ts to sequences of truth values of length n). And if f and g have the same value for every input, then f and g are the same function. (By extensionality of the underlying sets.) So there are exactly 2^{2^n} different n-place truth functions. □

Consider the 1-place truth functions. By Theorem 39, there are 4 such functions, outlined in Table 4.2. They are the *constant* functions $\mathbf{t}$ and $\mathbf{f}$, which yield the same value for every argument; the *identity* function $\mathbf{i}$, which

---

[5]Why the 'at least' formulation? This is because $(P \wedge \neg P) \vee Q$ is false in every structure, so is false on every row, so expresses the constant one-place truth function $\mathbf{f}$ (Table 4.2).

	t	f	i	n
T	T	F	T	F
F	T	F	F	T

Table 4.2: The four 1-place truth functions

yields as value the argument; and the *negation* function **n**, which yields the opposite truth value.

Truth-Functional Connectives

Definition 50 (Truth-functional Connective). A *truth-functional connective* is an n-place connective $\oplus$ such that there exists a n-place truth function f such that, for any $\varphi_1, \ldots, \varphi_n$,

$$\left| \oplus(\varphi_1, \ldots, \varphi_n) \right|_{\mathscr{A}} = f\left(\left| \varphi_1 \right|_{\mathscr{A}}, \ldots, \left| \varphi_n \right|_{\mathscr{A}} \right),$$

i.e., if $\oplus(\varphi_1, \ldots, \varphi_n)$ expresses some truth function f.

It is evident from Table 4.1 that each of the connectives of $\mathscr{L}_1$ are truth-functional connectives, so $\mathscr{L}_1$ is a *truth-functional language*. We've already encountered the 1-place truth function corresponding to the negation connective: it is **n** from Table 4.2. For another example: let **k** be the 2-place truth function defined as follows:

$$\mathbf{k}(x, y) = \begin{cases} T & \text{if } x = y = T; \\ F & \text{otherwise.} \end{cases}$$

As can be readily verified, this truth-function is expressed by $P \wedge Q$ (and by $Q \wedge R$, and $P_{17} \wedge R_{101}$, etc., etc.), and so is the truth function corresponding

to *conjunction*. Here are three other two-place functions we may name:

$$\mathbf{a}(x, y) = \begin{cases} F & \text{if } x = y = F; \\ T & \text{otherwise.} \end{cases}$$

$$\mathbf{c}(x, y) = \begin{cases} F & \text{if } x = T \neq y; \\ T & \text{otherwise.} \end{cases}$$

$$\mathbf{e}(x, y) = \begin{cases} T & \text{if } x = y; \\ F & \text{otherwise.} \end{cases}$$

It is easy to verify which connectives in Table 4.1 express these truth-functions.

Non-Truth Functional Connectives Consider the natural language connective 'necessarily', i.e., the connective that, given a sentence φ, yields a sentence ⌜Necessarily φ⌝. If φ is false in a structure, that obviously guarantees that ⌜Necessarily φ⌝ is false in that structure too (assuming the standard meaning for 'Necessarily'). But if φ is true in a structure, that doesn't tell us whether ⌜Necessarily φ⌝ is true in that structure: for some truths are necessary, others merely contingent. So to deal with the 'Necessarily' operator, we will have to introduce some further technology than we have available in $\mathscr{L}_1$ structures and classical valuations. That is the topic taken up by *modal logic* – see the discussion at the end of chapter 9. We already can see, however, one expressive limitation of $\mathscr{L}_1$ as compared to English: for English includes non-truth-functional connectives like 'necessarily', 'probably', 'typically', etc., and $\mathscr{L}_1$ does not. Every sentence in $\mathscr{L}_1$ has its truth value determined by the valuation rules, and the valuation rules only make use of truth functions.

Revisiting the Semantics We could use our truth-functions to express the semantics for our language in another form, giving this alternative

definition of a valuation.

Definition 51 (Alternative $\mathscr{L}_1$-valuation). When $\mathscr{A}$ is any $\mathscr{L}_1$-structure, the valuation function generated by $\mathscr{A}$ is this function:

$$|\varphi|_{\mathscr{A}} = \begin{cases} \mathscr{A}(\varphi) & \text{if } \varphi \text{ is a sentence letter;} \\ \mathbf{n}\left(|\psi|_{\mathscr{A}}\right) & \text{if} \varphi = \neg\psi; \\ \mathbf{k}\left(|\psi|_{\mathscr{A}}, |\chi|_{\mathscr{A}}\right) & \text{if} \varphi = (\psi \wedge \chi); \\ \mathbf{a}\left(|\psi|_{\mathscr{A}}, |\chi|_{\mathscr{A}}\right) & \text{if} \varphi = (\psi \vee \chi); \\ \mathbf{c}\left(|\psi|_{\mathscr{A}}, |\chi|_{\mathscr{A}}\right) & \text{if} \varphi = (\psi \to \chi); \\ \mathbf{e}\left(|\psi|_{\mathscr{A}}, |\chi|_{\mathscr{A}}\right) & \text{if} \varphi = (\psi \leftrightarrow \chi). \end{cases} \qquad \boxtimes$$

This definition is obviously equivalent to that in Definition 40, in that both definitions determine the same valuation of all the sentences of the language given the same initial structure.

Definition 52 (Base). Let $|\cdot|_{\mathscr{A}}$ be a classical valuation. A *base* $\mathfrak{B}$ for $|\cdot|_{\mathscr{A}}$ is a triple $\langle V, \mathbb{O}, c \rangle$ where V is a set of elements, $\mathbb{O}$ a set of operators on V (functions such that their domain and range are both V – recall chapter 2), and c is a function from the set of connectives of $\mathscr{L}_1$, $\{\neg, \wedge, \vee, \to, \leftrightarrow\}$, into $\mathbb{O}$, such that

- $|\varphi|_{\mathscr{A}} \in V$, for any $\mathscr{L}_1$ sentence φ;

- for any sentences $\varphi_1, \ldots, \varphi_n$, and any n-ary connective $\oplus$,

$$\left|\oplus\left(\varphi_1, \ldots, \varphi_n\right)\right|_{\mathscr{A},\mathscr{L}} = c(\oplus)\left(\left|\varphi_1\right|_{\mathscr{A},\mathscr{L}}, \ldots, \left|\varphi_n\right|_{\mathscr{A},\mathscr{L}}\right).$$

(Beall and van Fraassen, 2003: 27)

Informally, a base for a valuation is a set of values V, a set of truth-functions $\mathbb{O}$, and a mapping that takes connectives of the language to the truth-functions they express (Definition 50). A base for $\mathscr{L}_1$ is this: let $V = \{0, 1\}$, $\mathbb{O} = \{\mathbf{n}, \mathbf{k}, \mathbf{a}, \mathbf{c}, \mathbf{e}\}$, and let c be the function that maps $\neg$ to $\mathbf{n}$, $\wedge$ to $\mathbf{k}$, *etc.*

Exercises

1. Show that $\Gamma \vDash \varphi$ iff $\Gamma \cup \{\neg\varphi\} \vDash$.

2. State what features the truth table for $\Gamma \cup \{\varphi, \psi\}$ will possess when

   (a) Γ entails φ;

   (b) Γ is consistent;

   (c) φ is a contradiction;

   (d) φ and ψ are logically equivalent.

3. (a) Prove Transitivity.

   (b) Prove Contraposition.

   (c) Prove this rule, related to Cut: if $\Gamma \vDash \varphi$, and $\varphi, \Delta \vDash \psi$, then $\Gamma, \Delta \vDash \psi$.

4. Show the remaining clauses of the induction step for the proof of the theorem 'No Need for Falsity Clauses' for the cases where the complex sentence is of the forms:

   (a) $(\varphi \wedge \psi)$;

   (b) $(\varphi \rightarrow \psi)$;

   (c) $(\varphi \leftrightarrow \psi)$.

5. Prove Equivalents (Theorem 37).

6. (a) Prove the remaining induction cases in the proof of Equivalence (Theorem 38).

   (b) As a corollary of the Equivalence Theorem, prove that if $\vDash \varphi \leftrightarrow \psi$ and φ is a subsentence of χ, then $\vDash \chi \leftrightarrow \chi[\psi/\varphi]$.

7. Show that, if $\varphi, \psi \vDash \chi$ is a false sequent, it has a substitution instance $\varphi', \psi' \vDash \chi'$, in which φ' and ψ' are tautologies and χ' is inconsistent. (Note: by 'substitution instance' in the question, you are to understand a sequent that results from one *or more* uniform substitutions of sentence letters. It follows from General Substitution [Theorem 36] that successive chains of uniform substitutions never allow one to transform a correct sequent to an incorrect one.)

8. Show that if there are n sentence letters in Γ, the truth table for Γ has 2^n lines.

9. Prove that two sentences are logically equivalent iff they express the same truth-function.

10. Consider the deprived language which has the same syntactic formation rules and semantics as $\mathscr{L}_1$, but only contains the connectives $\neg$ and $\leftrightarrow$. Show that, in this language, every sentence is logically equivalent to a sentence in which no occurrence of $\neg$ has $\leftrightarrow$ in its scope.

Chapter 5

Metatheory of $\mathscr{L}_1$

5.1 Disjunctive Normal Form

Definition 53 (Arbitrary Conjunction). An inductive definition:

- $\left[\bigwedge_{i=1}^{1} \varphi_i\right] =_{df} \varphi_1;$

- $\left[\bigwedge_{i=1}^{n+1} \varphi_i\right] =_{df} \left(\left[\bigwedge_{i=1}^{n} \varphi_i\right] \wedge \varphi_{n+1}\right).$

The definition of arbitrary disjunction, $\bigvee_{i=1}^{n}$, is wholly parallel to that for arbitrary conjunction. (Giving it explicitly is left as an exercise.)

Definition 54 (Literal). A *literal* is any sentence letter or negated sentence letter.

Definition 55 (Disjunctive Normal Form (DNF)). An $\mathscr{L}_1$ sentence of φ is in *disjunctive normal form* iff there exist $n, m_1, \ldots, m_n$ such that

$$\varphi = \left[\bigvee_{i=1}^{n} \left[\bigwedge_{j=1}^{m_i} \pm s_{i,j}\right]\right],$$

where each $\pm s_{i,j}$ is a literal.

That is: a sentence is in disjunctive normal form iff it is a (perhaps degenerate) disjunction of (perhaps degenerate) conjunctions of (perhaps negated) sentence letters. One clear example is $((P \wedge \neg Q) \vee (P \wedge Q))$. But since P is a literal, and it is a degenerate arbitrary conjunction, and a degenerate arbitrary disjunction, it is in DNFalso. Another way of characterising which $\mathscr{L}_1$ sentences are in DNFis as follows: a sentence is DNFiff it contains at most only connectives drawn from $\{\neg, \wedge, \vee\}$, and such that $\vee$ never occurs in the scope of $\wedge$ or $\neg$, and $\wedge$ never occurs in the scope of $\neg$.

Our interest in disjunctive normal form lies in the following theorem:

Theorem 40 (DNF)

Every truth function is expressed by a $\mathscr{L}_1$ sentence in DNF.

Proof. Suppose f is an n-place truth function. Let $s_1, \ldots, s_n$ be sentence letters, and let $\mathscr{A}$ be any structure. Define an f-agreeing structure $\mathscr{A}$ as one where $f(\left|s_1\right|_{\mathscr{A}}, \ldots, \left|s_n\right|_{\mathscr{A}}) = T$. Assume there is at least one f-agreeing structure. Then define $\mathfrak{l}_{\mathscr{A},i} = s_i$ if $\mathscr{A}(s_i) = T$; otherwise let $\mathfrak{l}_{\mathscr{A},i} = \neg s_i$. For any structure $\mathscr{B}$ that agrees with $\mathscr{A}$ on $s_1, \ldots, s_n$, let $\mathfrak{l}_{\mathscr{B},i} = \mathfrak{l}_{\mathscr{A},i}$. It is obvious that $\left|\mathfrak{l}_{\mathscr{A},i}\right|_{\mathscr{A}} = T$ for each $\mathscr{A}, i$.

For every structure $\mathscr{A}$, define $\mathfrak{c}_{\mathscr{A}} = \mathfrak{l}_{\mathscr{A},1} \wedge \ldots \wedge \mathfrak{l}_{\mathscr{A},n} = \left[\bigwedge_{i=1}^{n} \mathfrak{l}_{\mathscr{A},i}\right]$. Again, it is obvious that $\left|\mathfrak{c}_{\mathscr{A}}\right|_{\mathscr{A}} = T$. Since there are only finitely many structures that differ from one another in their assignments to $s_1, \ldots, s_n$, there are only finitely many distinct sentences $\mathfrak{c}_{\mathscr{A}}$ for various $\mathscr{A}$. Therefore, there are only finitely many $\mathfrak{c}_{\mathscr{A}}$ *such that $\mathscr{A}$ is an f-agreeing structure*; let them be enumerated $\mathfrak{c}_1, \ldots, \mathfrak{c}_m$. It is obvious that each $\mathfrak{c}_i$ is true in one and only one structure, and that each such structure is f-agreeing.

Define $\mathfrak{d} = \mathfrak{c}_1 \vee \ldots \vee \mathfrak{c}_m = \left[\bigvee_{j=1}^{m} \mathfrak{c}_j\right]$, *unless* there were no f-agreeing structures, in which case let $\mathfrak{d} = P_1 \wedge \neg P_1$. It is obvious that $\mathfrak{d}$ is true in any structure which is among those such that some $\mathfrak{c}_i$ is true, i.e., $\mathfrak{d}$ is true in any f-agreeing structure.

$\mathfrak{d}$ is in DNF, by construction; and $|\mathfrak{d}|_{\mathscr{A}} = f(\left|s_1\right|_{\mathscr{A}}, \ldots, \left|s_n\right|_{\mathscr{A}})$ for all $\mathscr{A}$, because it captures all and only the f-agreeing structures. So $\mathfrak{d}$ expresses

f. □

The truth-table rationale for this proof is clear: first, find the rows of the truth table on which the truth-function gets T (those are the f-agreeing rows). Specify a sentence true in exactly one row by conjoining the relevant literals (so if the row is $|P|_= T$, $|Q|_= F$, the relevant conjunction is $P \wedge \neg Q$). Then disjoin those conjunctions which correspond to the f-agreeing rows; the result is a DNF sentence true in exactly the f-agreeing rows.

We can obviously use this result to show that, for any sentence φ, there is a sentence φ' which is logically equivalent to φ and which is in DNF form. This follows immediately from the DNF theorem and the fact that every sentence expresses a truth function.

CNF and Positive Truth Functions A formula is said to be in *Conjunctive Normal Form* iff it is a (possibly degenerate) conjunction of (possibly degenerate) disjunctions of literals. This is obviously a dual notion to CNF. In the case of DNF, the DNF sentence expressing some truth function f is the disjunction of conjunctions, each of which specifies an f-agreeing structure. Each conjunct is therefore *sufficient* for f to have the value T in the corresponding structure; the DNF sentence is the disjunction of all of these sufficient conditions. For CNF, we want the dual notion: the CNF expression of f will be a sentence which is a conjunction of disjunctions, each of which specifies a necessary condition for a structure to be f-agreeing. And what is a necessary condition for a structure to be f-agreeing? It must avoid being f-*disagreeing* – so each disjunct will be a disjunction of sentence letters and negated sentence letters, the truth of any of which is sufficient to ensure that some f-disagreeing structure is avoided. (A problem asks you to make the foregoing remarks into a more precise proof of the CNF theorem.)

Call a truth function f *positive* iff $f(T, \ldots, T) = T$ (Bostock, 1997: exercise 2.9.3). (Equivalently, if the top row of its truth table is T.) We can

show that all truth-functions which can be expressed using only $\rightarrow$ and $\wedge$ are positive (left for problem 9). Of interest is the converse theorem:

Theorem 41

All positive truth functions can be expressed by $\rightarrow$ and $\wedge$.

Proof. I sketch the proof. Suppose the contrary, for *reductio*. Then there is a truth function f which is positive but can't be expressed by $\wedge, \rightarrow$. There is a CNF sentence φ_c which expresses f; this will be a conjunction of disjunctions of (negated) sentence letters. By the constructive proof of the CNF theorem, it can be shown (by induction on complexity of sentences, and using the result that $\varphi \vee \psi \dashv\vDash (\psi \rightarrow \varphi) \rightarrow \varphi$) that each conjunct of the resulting CNF sentence will be equivalent to either (i) an arrow sentence *without* negation, or (ii) a conjunct of the form $\neg\varphi \vee \neg\psi$ (problem 10). In the second case, the CNF sentence as a whole can't be positive. So the first case must obtain for φ_c. By substituting each conjunct of φ_c for the arrow-sentence equivalent, we obtain a sentence φ, which is a conjunction of arrow sentences without negation. But φ expresses f and contains only $\rightarrow$ and $\wedge$. □

5.2   Functional Completeness, Expressive Adequacy

Functional Completeness and Expressive Adequacy   A language is said to be *functionally complete* iff there is a formula in that language that, under the intended semantics, expresses every truth function. The DNF theorem shows that every truth function is expressed by a formula in $\mathscr{L}_1$ (as DNF claims clearly are), and so $\mathscr{L}_1$ is functionally complete.

A set of connectives is *expressively adequate* iff there is a sentence containing only those connectives which expresses any truth function. The DNF theorem shows that $\{\vee, \wedge, \neg\}$ is expressively adequate, and therefore that the set of connectives of $\mathscr{L}_1$ is expressively adequate.

φ	ψ	$\varphi \wedge \psi$	$\neg(\neg\varphi \vee \neg\psi)$	$\varphi \vee \psi$	$\neg(\neg\varphi \wedge \neg\psi)$
T	T	T	T	T	T
T	F	F	F	T	T
F	T	F	F	T	T
F	F	F	F	F	F

Table 5.1: De Morgan Equivalences

What this last result shows is that, in some sense, $\rightarrow$ and $\leftrightarrow$ are *dispensible* – every truth function expressible using them can be expressed by a sentence without them (i.e., a DNF one).

There is another way to show this same result. Since two sentences are logically equivalent iff they express the same truth-function, the DNF theorem, together with the fact that every $\mathscr{L}_1$ sentence expresses a truth-function, suffices to show that every $\mathscr{L}_1$ sentence is logically equivalent to one in DNF form. Hence sentences involving $\rightarrow$ and $\leftrightarrow$ are logically equivalent to sentences in DNF form, and therefore in a sense dispensible.

The De Morgan Equivalences   Let '$\varphi \dashv\vDash \psi$' abbreviate '$\varphi \vDash \psi$ and $\psi \vDash \varphi$'.

Theorem 42 (De Morgan Equivalences)

$$(\varphi \vee \psi) \dashv\vDash \neg(\neg\varphi \wedge \neg\psi),$$
$$(\varphi \wedge \psi) \dashv\vDash \neg(\neg\varphi \vee \neg\psi).$$

Proof. These equivalences can be easily demonstrated using a truth table: Table 5.2. □

These equivalences show that $\{\neg, \vee\}$ and $\{\neg, \wedge\}$ are expressively adequate, given the DNF theorem.

φ	ψ	$(\varphi \uparrow \psi)$	$(\varphi \downarrow \psi)$
T	T	F	F
T	F	T	F
F	T	T	F
F	F	T	T

Table 5.2: Sheffer Stroke and Peirce Arrow

Inadequacy and New Connectives   A set of connectives is expressively *inadequate* iff there is some truth function which cannot be expressed by sentences involving only those connective. Accordingly, $\{\neg\}$ is expressively inadequate: quite apart from whether it can express a 2- or more place truth function, consider the 1-place function f_T which is constantly true. Since $f_T(T) = f_T(F)$, we would need a sentence φ involving only negation such that in even one case $|\varphi|_{\mathscr{A}} = |\neg\varphi|_{\mathscr{A}}$; this clearly cannot be.

Given our definition of a connective, we can imagine extending our language $\mathscr{L}_1$ by adding new sentence-forming operators that express different connectives. We could add the 0-place connective $\top$ which expresses f_T. As we know from the DNF theorem, adding $\top$ doesn't add to the expressive power of the language: $\top$ is logically equivalent to $(P \vee \neg P)$, and so adding it would be redundant. Adding $\wedge$ to the language $\mathscr{L}_\neg$, in which the only connective is negation, is not redundant: $\mathscr{L}_{\neg,\wedge}$ is functionally complete and $\mathscr{L}_\neg$ is not.

Sheffer Stroke and Peirce Arrow   The connectives $\uparrow$ and $\downarrow$ can be added to a language, with the truth tables in Table 5.2: The $\uparrow$, also written |, is called the *Sheffer stroke*; the $\downarrow$ has no standard name, but is sometimes called the *Peirce arrow*.

We can show that both of these connectives are expressively adequate. The Sheffer stroke is, since $\neg\varphi$ is logically equivalent to $(\varphi \uparrow \varphi)$, and $(\varphi \wedge \psi)$ is logically equivalent to $((\varphi \uparrow \psi) \uparrow (\varphi \uparrow \psi))$.

φ	ψ	$\varphi \wedge \psi$	$\varphi \uparrow \psi$
T	T	T	F
T	F	F	F
F	T	F	F
F	F	F	T

$\Leftrightarrow$

φ	ψ	$\varphi \vee \psi$	$\varphi \downarrow \psi$
F	F	F	T
F	T	T	T
T	F	T	T
T	T	T	F

Table 5.3: Duality Illustrated Using Truth Tables

5.3 Duality

Duality $\wedge$ and $\vee$, as well as $\uparrow$ and $\downarrow$, are *duals* of each other.

Definition 56 (Duality). The *dual* of a truth-functional connective is that connective whose truth table results from that of the given connective by replacing *every* occurrence of T by F and every occurrence of F by T (see Table 5.3).

Other connectives have duals too: $\varphi \leftrightarrow \psi$ is dual to the operator $\not\leftrightarrow$ (where $\varphi \not\leftrightarrow \psi$ is true just in case φ and ψ have different truth values).

The Duality Lemma for $\wedge, \vee$ Suppose that φ is a sentence of $\mathcal{L}_1$ which involves only connectives in $\{\wedge, \vee, \neg\}$ (any sentence of $\mathcal{L}_1$ will be equivalent to some such φ). Now we define two operations on sentences of this form:

1. Let $\overline{\varphi}$ be the sentence that results from writing '$\neg$' directly in front of every sentence letter in φ (even those already negated).

2. Let $\varphi^{\star}$ be the sentence that results from substituting each connective in φ by its dual.

Lemma 43 (Duality for Conjunction and Disjunction)
For any φ, $\varphi^{\star} \dashv\vDash \neg\overline{\varphi}$.

Proof. *Base case:* φ is just a sentence letter: so that $\varphi^{\star} = p$, and $\neg\overline{\varphi} = \neg\neg p$, which are equivalent (by truth tables).

Induction step:

1. φ is $\neg\psi$, and the hypothesis holds for ψ. Then $\varphi^{\star} = (\neg\psi)^{\star} = \neg(\psi^{\star})$. Because the theorem holds of ψ, $\neg(\psi^{\star}) \dashv\vDash \neg\neg\overline{\psi}$; but $\neg\neg\overline{\psi} = \neg\overline{\neg\psi} = \neg\overline{\varphi}$, which proves the case.

2. φ is $(\psi \wedge \chi)$, and the hypothesis holds for ψ, χ. $\varphi^{\star} = (\psi \wedge \chi)^{\star} = \psi^{\star} \vee \chi^{\star}$. By the induction hypothesis, $\psi^{\star} \vee \chi^{\star} \dashv\vDash \neg\overline{\psi} \vee \neg\overline{\chi}$. By De Morgan, $\neg\overline{\psi} \vee \neg\overline{\chi} \dashv\vDash \neg(\overline{\psi} \wedge \overline{\chi})$, which is $\neg(\overline{\psi \wedge \chi}) = \neg(\overline{\varphi})$.

3. φ is $(\psi \vee \chi)$; much as for $\wedge$. $\qquad\qquad\qquad\qquad\square$

Duality Theorem

Theorem 44 (Duality for Conjunction and Disjunction)

If $\varphi^{\star}$ is dual to φ and $\psi^{\star}$ is dual to ψ, then $\varphi \vDash \psi$ iff $\psi^{\star} \vDash \varphi^{\star}$.

Proof. Suppose $\varphi \vDash \psi$. Then, by Substitution of each sentence letter by its negation, $\overline{\varphi} \vDash \overline{\psi}$. By Contraposition, $\neg\overline{\psi} \vDash \neg\overline{\varphi}$. By the Duality Lemma, $\psi^{\star} \vDash \varphi^{\star}$. $\qquad\qquad\qquad\qquad\square$

Corollary 45

If $\varphi \dashv\vDash \psi$, $\varphi^{\star} \dashv\vDash \psi^{\star}$.

Theorem 46 (Duality for Tautologies)

If $\vDash \varphi$, then $\vDash \neg(\varphi^{\star})$.

Proof. Suppose $\vDash \varphi$. Since φ is true in every structure, $\overline{\varphi}$ is also true in every structure (obvious by substitution). So $\vDash \overline{\varphi}$. By the Duality Lemma, $\overline{\varphi} \dashv\vDash \neg\varphi^{\star}$. So $\vDash \neg\varphi^{\star}$. $\qquad\qquad\qquad\qquad\square$

Duality Generalised Every truth functional connective c – in any truth-functional language – has a dual, which we denote $c^\star$ by extending our previous notation. A set of truth functors $\mathbb{C} = \{c_1, \dots, c_n\}$ is *self-dual* iff $\mathbb{C} = \{c_1^\star, \dots, c_n^\star\}$. It is obvious that $\{\neg\}$ is self-dual. Given the result above, the set $\{\wedge, \vee\}$ is self-dual, as $\vee = \wedge^\star$ and $\wedge = \vee^\star$.

Define a function $\star$ on truth-values such that $T^\star = F$ and $F^\star = T$. If $\mathscr{A}$ is a structure, let $\mathscr{A}^\star$ be the structure such that for all φ, $\mathscr{A}^\star(\varphi) = T^\star$ iff $\mathscr{A}(\varphi) = T$. For any language $\mathscr{L}$, let a connective c be in $\mathscr{L}$ iff $c^\star$ is in $\mathscr{L}$. The dual $\varphi^\star$ of a sentence φ of $\mathscr{L}$ results from replacing *every* connective in φ by its dual. Then we can show:

$$|\varphi|_{\mathscr{A}} = \left(\left|\varphi^\star\right|_{\mathscr{A}^\star}\right)^\star.$$

5.4   Interpolation

Craig Interpolation Theorem for Sentential Logic

Theorem 47 (Craig Interpolation)
If $\varphi \vDash \psi$, and there is a non-empty set B of sentence letters occurring in both φ and ψ, then there is a sentence χ (an interpolant*) such that both $\varphi \vDash \chi$ and $\chi \vDash \psi$, and each sentence letter in χ is in B.*

Proof. Let the members of B be enumerated $b_1, \dots, b_n$. Let $\mathscr{A}$ be a *B-variant* of $\mathscr{B}$ iff $\mathscr{A}(b) = \mathscr{B}(b)$ for all $b \in B$.

Let $\{\mathscr{B}_1, \dots, \mathscr{B}_{2^n}\}$ be a set of structures every pair of which differ on their assignment of a truth value to at least one $b \in B$. Define a n-place truth function f_χ, for every structure $\mathscr{B}_i$:

$$f_\chi\left(\left\langle |b_1|_{\mathscr{B}_i}, \dots, |b_n|_{\mathscr{B}_i} \right\rangle\right) = \begin{cases} F & \begin{array}{l} \text{if there is a } B\text{-variant } \mathscr{A} \text{ of } \mathscr{B}_i \\ \text{in which } |\psi|_{\mathscr{A}} = F; \end{array} \\ T & \text{otherwise.} \end{cases} \qquad ⊠$$

By construction, f_χ is a total truth function. The sentence χ that (by DNF)

expresses f_χ is our needed interpolant.

- χ clearly entails ψ, since by construction no structure that makes χ true has a B-variant that makes ψ false.

- Moreover, χ is clearly entailed by φ, since if there were a structure $\mathscr{C}$ which made φ true but χ false, $\mathscr{C}$ would be a B-variant of some $\mathscr{B}_i$ that makes ψ true; but by construction χ would be true in $\mathscr{B}_i$ and in every B-variant of it, including $\mathscr{C}$. There can therefore be no such $\mathscr{C}$ that both makes χ true and false. □

The Craig Interpolation Theorem actually names the version of this result proved for predicate logic (Craig, 1957).

5.5   Compactness

Compactness   If Γ is a inconsistent set of $\mathscr{L}_1$ sentences ($\Gamma \vDash$), then it needn't be that there is a proper subset $\Gamma' \subset \Gamma$ such that $\Gamma' \vDash -$ while $P, \neg P \vDash$, clearly $\{P\}$ and $\{\neg P\}$ are both consistent.

But, it turns out, when Γ is *infinite* and inconsistent, there is always some *finite* subset of Γ which is inconsistent. This property is known as *compactness*.

It has the immediate consequence that if $\Gamma \vDash \varphi$, then there is some finite set of premises Γ' such that $\Gamma' \vDash \varphi$; every valid argument in $\mathscr{L}_1$ can be captured by a finite sequent.

Compactness Theorem   A set Γ is *finitely satisfiable* iff every finite subset of Γ is satisfiable (i.e., has a model).

Theorem 48 (Compactness)
If Γ is finitely satisfiable then Γ is satisfiable. (The contrapositive of this is Compactness in the sense introduced just above.)

This theorem can be proved a number of ways. It is a fairly quick consequence of the Completeness theorem for the natural deduction formalism for sentential logic, which we prove in the next chapter. Here I will prove it directly. The proof comes in several stages, marked out below.

Note that the *converse* of Compactness – that satisfiability entails finite satisfiability – is trivial from the structural rules.

Proof of Compactness, Stage 1: Definition of Γ_is   Suppose that Γ is finitely satisfiable. The set of sentence letters occurring in Γ is enumerable, so let its members be enumerated $\varphi_1, \varphi_2, \dots$.

We define a sequence of supersets of Γ as follows:

$$\Gamma_0 = \Gamma$$
$$\vdots$$
$$\Gamma_{n+1} = \begin{cases} \Gamma_n \cup \{\varphi_{n+1}\} & \text{if } \Gamma_n \cup \{\varphi_{n+1}\} \text{ is finitely satisfiable;} \\ \Gamma_n \cup \{\neg\varphi_{n+1}\} & \text{otherwise.} \end{cases} \boxtimes$$

Proof of Compactness, Stage 2: Γ_is are finitely satisfiable

Lemma 49

Each Γ_i is finitely satisfiable.

Proof. Base case: $\Gamma_0 = \Gamma$, which is finitely satisfiable by hypothesis.

Induction step: Suppose that each Γ_i, $i \leqslant n$, is finitely satisfiable. Now suppose Γ_{n+1} is not. Then, by definition of Γ_{n+1}, neither $\Gamma_n \cup \{\varphi_{n+1}\}$ nor $\Gamma_n \cup \{\neg\varphi_{n+1}\}$ is finitely satisfiable.

So there must be finite subsets Δ and Θ of Γ_n such that both $\Delta, \varphi_{n+1} \vDash$ and $\Theta, \neg\varphi_{n+1} \vDash$. By weakening, both $\Delta, \Theta, \varphi_{n+1} \vDash$ and $\Delta, \Theta, \neg\varphi_{n+1} \vDash$.

But $\Delta \cup \Theta$ is a finite subset of a finitely satisfiable set, so is consistent. In any structure $\mathscr{A}$ in which all the members of $\Delta \cup \Theta$ are true, either $\left|\varphi_{n+1}\right|_{\mathscr{A}} = T$ or $\left|\neg\varphi_{n+1}\right|_{\mathscr{A}} = T$. So either $\Delta, \Theta, \varphi_{n+1} \nvDash$ or $\Delta, \Theta, \neg\varphi_{n+1} \nvDash$. Contradiction; so Γ_{n+1} is finitely satisfiable. $\square$

Proof of Compactness, Stage 3: Define a structure from the Γ_is For all φ_i, define

$$\mathscr{G}(\varphi_i) = \begin{cases} T & \text{if } \varphi_i \in \Gamma_i; \\ F & \text{otherwise (i.e., if } \neg\varphi_i \in \Gamma_i). \end{cases} \boxtimes$$

For any k, Γ_k is finitely satisfiable; since for all $i \leqslant k$, either $\varphi_i \in \Gamma_k$ or $\neg\varphi_i \in \Gamma_k$, the finite set of literals

$$\Phi_k = \{\varphi_i \in \Gamma_k : i \leqslant k\} \cup \{\neg\varphi_i \in \Gamma_k : i \leqslant k\}$$

is a subset of Γ_k and is consistent.

By construction, for any k, and for all $\psi \in \Phi_k$, $|\psi|_\mathscr{G} = T$. Moreover, *every* structure $\mathscr{G}'$ which agrees with $\mathscr{G}$ on the sentence letters in Φ_k also assigns T to all members of Φ_k. (And obviously, by construction, only those structures which agree on these sentence letters can satisfy Φ_k.)

Proof of Compactness, Stage 4: $\mathscr{G}$ satisfies Γ Suppose $\mathscr{G}$ does not satisfy Γ. Then for some $\gamma \in \Gamma$, $|\gamma|_\mathscr{G} = F$.

γ has finitely many sentence letters occurring in it; let φ_k be the highest numbered. Since every $\psi \in \Phi_k$ is true in every structure which agrees with $\mathscr{G}$ on φ_i, $i \leqslant k$, γ is false in every such structure.

So $\Phi_k \cup \{\gamma\}$ is unsatisfiable. But it is a finite subset of Γ_k, which is finitely satisfiable, so $\Phi_k \cup \{\gamma\}$ satisfiable. Contradiction; so $\mathscr{G}$ must satisfy Γ. Γ is satisfiable if all of its finite subsets are.

An alternative proof of compactness can be found in Appendix A.

5.6 Decidability

Decidability An *effective procedure* for determining some query is an automatic and mechanical algorithm that terminates in a *finite* time with

a correct 'yes' or 'no' answer.

A property is *decidable* iff there is an effective procedure which tests for it.

Theorem 50 (Decidability of Finite Validity)
If Γ is a finite set of sentences of $\mathscr{L}_1$, then it is decidable whether $\Gamma \vDash$.

This has the immediate consequence that it is decidable whether $\Gamma \vDash \varphi$ (check whether $\Gamma, \neg\varphi \vDash$), and thus that *validity is decidable*.

(The notion of an effective procedure is introduced informally, but it can be made precise by investigating *Turing machines* and *recursive functions* – see Boolos *et al.* (2007: chs. 1–8).)

Finite Decidability Using Truth Tables   Let Γ be a finite set of $\mathscr{L}_1$ sentences. Each $\gamma \in \Gamma$ is only finitely long, so there are n sentence letters occurring in Γ, for some finite n.

Construct a truth-table with 2^n lines such that each line corresponds to some structure assigning truth values to each sentence letter in Γ; this can be done mechanically in a finite time. For each $\gamma \in \Gamma$, determine its truth value in each row; this can be done mechanically in a finite time for each γ in each row, and there are finitely many rows, so the truth value of each γ in every structure can be determined mechanically in a finite time. Since Γ is finite, we can determine the truth value of each sentence in Γ in every structure in a finite time.

Check each row of the truth table to see if all sentences in Γ have been assigned T: If they all have, in some row, then stop, and answer 'No' to the query, 'Is $\Gamma \vDash$?'. If no row has all $\gamma \in \Gamma$ assigned T, then answer 'Yes'. There are only finitely many rows, so this final check can be done in a finite time. All steps can be finitely mechanised, and there are finitely many steps: inconsistency is decidable.

Positive Infinite Decidability   A query is *positively decidable* iff an effective procedure will terminate and correctly answer 'Yes' in a finite time. A

query is *negatively decidable* iff an effective procedure will terminate and correctly answer 'No' in a finite time. It is decidable iff it is both positively and negatively decidable.

Theorem 51 (Positive Decidability of Inconsistency)

If Γ is an infinite set of sentences of $\mathscr{L}_1$, then if $\Gamma \vDash$, there is an effective procedure that will demonstrate that; but there is no effective procedure that will decide if arbitrary Γ is consistent.

We first show a lemma.

An Effective Procedure Generates Finite Subsets of $\mathbb{N}$

Lemma 52

For every finite subset of the natural numbers $\mathbb{N}$, there is an effective procedure which produces all and only members of that set.

Proof. **Proof sketch**: Begin by announcing 1. Then proceed inductively: if n is the highest number announced so far, list every set which contains only numbers $\leqslant n$. Since at any point we will only have announced finitely many numbers, there are only finitely many such sets, so this is finitely achievable. Once this is complete, proceed to announce $n + 1$. It's clear that any finite set of natural numbers will be generated at some finite time from the beginning of this series – since each such set has a largest member n, it will be listed before $n + 1$ is announced. $\square$

Proof of Positive Infinite Decidability

Proof. The set of sentences of $\mathscr{L}_1$ is enumerable, so $\Gamma = \{\gamma_1, \dots, \gamma_n, \dots\}$ where each γ_i is assigned some natural number as an index. If $\Gamma \vDash$, then by the Compactness theorem, there is a finite subset $\Delta \subseteq \Gamma$ such that $\Delta \vDash$.

By the Lemma, we can effectively generate each finite subset Δ by generating the indices i on $\gamma_i \in \Delta$, and since finite consistency is decidable (by Finite Decidability), if Δ is inconsistent that will be decided at some finite

point. So if Γ is inconsistent, this test will indicate that some finite subset is inconsistent after some finite time.

Yet if Γ is consistent, no finite subset is inconsistent, and this procedure will continue indefinitely checking consistent sets and showing that they are consistent. It never halts after a finite time, as there remain infinitely many other consistent finite subsets of Γ not yet checked. So inconsistency is only positively decidable, and infinite consistency is undecidable. □

Exercises

1. Give, explicitly, the inductive definition of arbitrary disjunction.

2. (a) Show that if there are n sentence letters in S, there are 2^n sentences of the form $\mathfrak{c}_{\mathscr{A}}$ defined in the proof of the DNF theorem.

   (b) Explain clearly why, in the proof of the DNF theorem, the formula $\mathfrak{d}$ there constructed expresses the truth function f.

3. Prove, by an argument analogous to the DNF theorem, this claim:

   Theorem 53 (CNF)
   Every truth function is expressed by a formula in Conjunctive Normal Form.

4. Show that all truth-functions which can be expressed using only $\to$ and $\wedge$ are positive. (Prove by induction on complexity of formulae.)

5. The proof of theorem 41 was only sketched. Fill in these two crucial gaps.

   (a) Show that $\varphi \vee \psi$ is logically equivalent to $(\psi \to \varphi) \to \varphi$;

   (b) Show (by induction on complexity of sentences, and using the above result) that each conjunct of a CNF sentence will be equivalent to either (i) an arrow sentence *without* negation, or (ii) a conjunct of the form $\neg \varphi \vee \neg \psi$.

   (c) Show that a necessary and sufficient condition for a sentence φ to be logically equivalent to a sentence involving just the truth functional connectives $\to$ and $\wedge$ is that the sentence has the value T in any structure that assigns the value T to each sentence letter in φ.

6. (a) Show that $\{\vee, \neg\}$ is expressively adequate.

(b) Is $\{\rightarrow, \neg\}$ expressively adequate?

(c) Is $\{\leftrightarrow, \wedge, \rightarrow, \vee\}$ expressively adequate?

(d) Is any connective in $\mathcal{L}_1$ expressively adequate by itself?

(e) Prove that $\downarrow$ is expressively adequate.

(f) How many 2-place truth-functional connectives are expressively adequate by themselves?

(g) Consider the 0-place connective $\bot$ that expresses the constant 0-place function $f : \varnothing \mapsto F$. Is this expressively adequate, or can you add a connective to get an expressively adequate set (where the added connective is not itself expressively adequate)?

7. (a) Consider the self-dual set of connectives $\{\rightarrow, \rightarrow^\star\}$. Is this set of connectives expressively adequate? Is self-duality either necessary or sufficient for expressive adequacy of a set of connectives?

(b) Show that if a two-place truth functional connective $\oplus$ is self-dual, then the function that expresses it, $f_\oplus$, must be such that $f_\oplus(T, F) \neq f_\oplus(F, T)$ and $f_\oplus(F, F) \neq f_\oplus(T, T)$. Make use of this result, establish how many self-dual two-place truth functors there are.

(c) Let φ and ψ be sentences of $\mathcal{L}_{\neg, \wedge, \vee}$. We say that a sentence φ is *dualable* iff for all $\mathscr{A}$, $(|\varphi|_{\mathscr{A}^\star})^\star = |\varphi^\star|_{\mathscr{A}}$.

   i. Prove that if φ and ψ are dualable, so too are $\neg\varphi$, $(\varphi \wedge \psi)$, and $(\varphi \vee \psi)$.

   ii. Prove using this result that all sentences of $\mathcal{L}_{\neg, \wedge, \vee}$ are dualable.

(d) If φ is an $\mathcal{L}_1$ sentence, let $\varphi^\ddagger$ be the sentence that results from uniformly substituting in φ any subsentence of the form

   - $(\psi \wedge \chi)$ by $(\psi \vee \chi)$;
   - $(\psi \vee \chi)$ by $(\psi \wedge \chi)$;
   - $(\psi \rightarrow \chi)$ by $\neg(\chi \rightarrow \psi)$; and
   - $(\psi \leftrightarrow \chi)$ by $\neg(\psi \leftrightarrow \chi)$.

   i. Show that $\vDash \varphi \leftrightarrow \psi$ iff $\vDash \varphi^\ddagger \leftrightarrow \psi^\ddagger$.

ii. Letting the dual operator $\star$ apply to all $\mathcal{L}_1$ connectives in the natural way, show that $\varphi^{\ddagger} =\!\!\models \varphi^{\star}$.

8. Let $\mathcal{L}_{\uparrow,\downarrow}$ be the language which has $\uparrow$ and $\downarrow$ as its only connectives.

   (a) For any φ in $\mathcal{L}_{\uparrow,\downarrow}$, and any structure $\mathscr{A}$, show that $|\varphi|_{\mathscr{A}} = (|\varphi^{\star}|_{\mathscr{A}^{\star}})^{\star}$.

   (b) For any $\mathcal{L}_{\uparrow,\downarrow}$-sentence φ, let $\varphi^{\star}$ be the sentence which results from replacing every connective in φ by its dual, and let $\overline{\varphi}$ be the result of substituting $(P_i \uparrow P_i)$ for every sentence letter P_i occurring in φ.

      i. Prove that for every $\mathcal{L}_{\uparrow,\downarrow}$ sentence φ, $\varphi^{\star}$ is logically equivalent to $(\overline{\varphi} \uparrow \overline{\varphi})$.

      ii. Prove that if φ and ψ are $\mathcal{L}_{\uparrow,\downarrow}$ sentences, $\varphi \models \psi$ iff $\psi^{\star} \models \varphi^{\star}$.

9. Find an interpolant for the following sequents. Be sure in each case to give the simplest interpolant (i.e., find the most elegant sentence that is equivalent to your chosen interpolant).

   (a) $((Q \vee P) \rightarrow R) \models ((P_1 \wedge \neg R) \rightarrow \neg Q)$;

   (b) $(\neg(P \vee Q) \wedge (P \leftrightarrow R)) \models ((R \rightarrow P) \wedge \neg(P_1 \wedge R))$;

   (c) $((Q_2 \leftrightarrow Q) \wedge \neg((R \rightarrow \neg P_1) \vee \neg(P \rightarrow Q))) \models (R_1 \rightarrow (P_2 \rightarrow (\neg P \vee Q)))$;

10. (a) Let Γ be a possibly infinite set of sentences of $\mathcal{L}_1$ such that $\Gamma \models$. Show that there is a finite disjunction, δ, each disjunct of which is the negation of a sentence in Γ, and such that $\models \delta$.

    (b) Consider the following relation holding between sets of sentences:

    > Where Γ and Δ are any sets of sentences, $\Gamma \models^{\star} \Delta$ is correct iff every structure which satisfies *every* $\gamma \in \Gamma$ is also one which satisfies *at least one* $\delta \in \Delta$.

    Show that if $\Gamma \models^{\star} \Delta$, there is a finite conjunction of $\mathcal{L}_1$ sentences in Γ,

    $$\Phi = (\varphi_1 \wedge \dots \wedge \varphi_n),$$

    and a finite disjunction of $\mathcal{L}_1$ sentences in Δ,

    $$\Psi = (\psi_1 \vee \dots \vee \psi_n),$$

    such that $\Phi \models \Psi$. (You may assume the Compactness theorem.)

(c) A set of sentences of English is *compossible* just in case it is possible for them all to be true together. An analogue for the compactness theorem in English would be: for every infinite set E of English sentences not all compossible, there is a finite subset of E whose members are not all compossible.

    i. Show that this analogue of compactness fails for English.

    ii. What does this show about translations from English into $\mathcal{L}_1$?

11. Intuitively, any effective procedure must be able to be written down by a finite string of sentences in some language – say, English.

    (a) Give an argument that the set of effective procedures is countable.

    (b) Let a *recipe* be any finite set of English sentences. Consider the set E of recipes. (It is obvious that the set of effective procedures in English which compute one-place functions whose domain is (a subset of) the natural numbers will be a subset of E.) This set is countable; let f_n be the function – if there is one – computed by the n-th recipe in E under some enumeration of E. Define

$$d(n) = \begin{cases} 2 & \text{if } f_n(n) \text{ is defined and equal to 1;} \\ 1 & \text{otherwise} \end{cases}$$

    i. Show that there is no effective procedure for computing d.

    ii. Show that there is no effective procedure for deciding if a recipe is an effective procedure, and therefore that while the set of recipes E can be effectively produced, some subsets of E (in particular, the one corresponding to the set of effective procedures) cannot be effectively produced.

Chapter 6

Natural Deduction Proofs in $\mathscr{L}_1$; Soundness and Completeness

6.1 Natural Deduction Proofs

Proof   In the last chapter, we saw that using truth tables was a decidable method for evaluating sequents, but it is cumbersome and inelegant. Here, we consider elegant and powerful *proof systems* for sentential logic.

A *proof* in the informal sense is something which establishes the truth of some claim. Our proofs are structures made out of sentences of $\mathscr{L}_1$, with the following desirable properties:

- Each step in the proof involves an obviously correct rule.

- Whenever the earlier sentences are true, the later sentences will be also; so the terminal conclusion of a proof should always be true whenever the assumptions are.

- All the logical truths should have a proof.

In this chapter, we'll see that there do exist proofs that meet these conditions for $\mathscr{L}_1$; and thus that we can establish various claims in $\mathscr{L}_1$ without needing to consider all the $\mathscr{L}_1$-structures.

Natural Deduction: Assumptions and ⊢   The first proof system is the *natural deduction* system introduced in The Logic Manual.

In this system, every proof begins with *assumptions. Any sentence may be an assumption.* A proof terminates with a *conclusion*.

Whenever there is a correctly constructed proof that φ, on the basis of assumptions $\gamma_1, \ldots, \gamma_n$, we write this *syntactic sequent*:

$$\gamma_1, \ldots, \gamma_n \vdash \varphi.$$

This is to be read 'φ is *provable from* $\gamma_1, \ldots, \gamma_n$', or '$\gamma_1, \ldots, \gamma_n$ *syntactically entail(s)* φ'. If φ can be proved with no assumptions at all, we write $\vdash \varphi$; in that case, we say that φ is a *theorem*.

Any sentence φ may be assumed; applying no rules at all, the proof therefore terminates with φ. So $\varphi \vdash \varphi$; not perhaps the most interesting result, but nevertheless intuitively correct.

Natural Deduction Rules   As any assumption is a proof on its own, our rules are rules for constructing new proofs from existing proofs, often (but not always) retaining the assumptions of the existing proofs.

Some rules permit an assumption to be *discharged*. Suppose you have a proof of φ on the basis of assumption γ. You thus show that *if* you had a proof of γ, *then* you could construct a proof of φ; and this on its own is a proof of $\gamma \rightarrow \varphi$, *whether or not γ is actually true.* The assumption γ has been discharged. Note that γ needn't be used, or even occur in the proof, to be discharged: the rules entitle one to discharge *every* occurence of a sentence – even if there are none!

We can consider two types of obvious rules: those which construct a

$$\frac{\begin{array}{c}[\gamma] \\ \vdots \\ \varphi\end{array}}{\gamma \rightarrow \varphi} \rightarrow\text{Intro} \qquad \frac{\gamma \rightarrow \varphi \qquad \gamma}{\varphi} \rightarrow\text{Elim}$$

$$\frac{\begin{array}{cc}[\varphi] & [\varphi] \\ \vdots & \vdots \\ \psi & \neg\psi\end{array}}{\neg\varphi} \neg\text{Intro} \qquad \frac{\begin{array}{cc}[\neg\varphi] & [\neg\varphi] \\ \vdots & \vdots \\ \psi & \neg\psi\end{array}}{\varphi} \neg\text{Elim}$$

Table 6.1: Natural Deduction Rules for $\mathscr{L}_{\neg,\rightarrow}$

proof of a new sentence of less complexity than sentence proved by the existing proof(s), and those which construct a proof of a sentence of increased complexity. We call the first type of rule *elimination* rules, and the second type *introduction* rules.

Natural Deduction Rules for $\mathscr{L}_{\neg,\rightarrow}$ Let us consider for the moment the fragment of $\mathscr{L}_1$ involving only $\neg$ and $\rightarrow$, $\mathscr{L}_{\neg,\rightarrow}$. We have in this fragment four rules, laid out in Table 6.1, elimination rules and introduction rules for each connective. I write $[\gamma]$ to show that the rule permits the assumption γ to be discharged. Call this system of rules, plus the rules about assumptions, $ND_{\neg,\rightarrow}$.

Deduction Theorem for $ND_{\neg,\rightarrow}$

Theorem 54 (Deduction Theorem)
$\Gamma \vdash \psi \rightarrow \varphi$ *iff* $\Gamma, \psi \vdash \varphi$.

Suppose $\Gamma, \psi \vdash \varphi$. Then we can apply $\rightarrow$Intro as follows:

$$\frac{\begin{array}{c}\Gamma, [\psi] \\ \vdots \\ \varphi\end{array}}{\psi \rightarrow \varphi} \rightarrow\text{Intro}$$

Therefore, $\Gamma \vdash \psi \to \varphi$. Similarly, suppose $\Gamma \vdash \psi \to \varphi$:

$$\frac{\overset{\Gamma}{\vdots} }{\dfrac{\psi \to \varphi \qquad \psi}{\varphi}} \to\text{Elim}$$

That is, $\Gamma, \psi \vdash \varphi$.

¬Elim can be replaced by ¬¬Elim   Consider the rule:

$$\frac{\neg\neg\varphi}{\varphi} \ \neg\neg\text{Elim}$$

This rule can obviously be derived from our rules: the correctness of the syntactic sequent $\neg\neg\varphi \vdash \varphi$ is demonstrated by this proof:

$$\frac{\neg\neg\varphi \quad [\neg\varphi]}{\varphi} \ \neg\text{Elim}$$

But with ¬¬Elim we can show the rule ¬Elim is redundant, using ¬Intro:

$$\frac{\dfrac{\overset{[\neg\varphi]}{\vdots}}{\psi} \qquad \dfrac{\overset{[\neg\varphi]}{\vdots}}{\neg\psi}}{\dfrac{\neg\neg\varphi}{\varphi} \ \neg\neg\text{Elim}} \ \neg\text{Intro}$$

6.2   Soundness

Soundness and Completeness   A proof system P in a given language is *sound* with respect to a semantic interpretation of the language just in case whenever there is a proof which establishes $\Gamma \vdash_P \varphi$, it is the case that $\Gamma \vDash \varphi$.

A proof system P in a given language is *complete* with respect to a

semantic interpretation of the language just in case whenever it is the case that $\Gamma \vDash \varphi$, there is a proof which establishes $\Gamma \vdash_P \varphi$.

In this section and the next, we will show that the natural deduction system $ND_{\neg,\rightarrow}$ just introduced is *sound and complete* for the standard semantics of $\mathscr{L}_{\neg,\rightarrow}$. Given the expressive adequacy of $\mathscr{L}_{\neg,\rightarrow}$, that also shows the completeness of the natural deduction system with respect to $\mathscr{L}_1$; to show the soundness of that system, we also need to show that the other rules of the system ND are also sound.

Soundness of $ND_{\neg,\rightarrow}$

Theorem 55
The Rules of $ND_{\neg,\rightarrow}$ are sound.

Proof. Suppose that $\Gamma \vdash \varphi$. We show by induction on the complexity of proofs that $\Gamma \vDash \varphi$.

Base case: The least complex proof is a single node, φ, establishing $\varphi \vdash \varphi$. In this case, any structure which assigns T to φ clearly assigns T to φ, so $\varphi \vDash \varphi$, i.e., $\Gamma \vDash \varphi$.

Induction step: Suppose $\Gamma \vdash \varphi$, established by applying one of the natural deduction rules to some previously obtained proofs to obtain a proof of φ. There are four cases:

$\rightarrow$**Elim** From a proof of ψ on assumptions Δ, and a proof of $\psi \rightarrow \varphi$ on assumptions Θ, obtain a proof of φ on assumption $\Gamma = \Delta \cup \Theta$. By the induction hypothesis, $\Delta \vDash \psi$ and $\Theta \vDash \psi \rightarrow \varphi$, so every structure which makes all of the members of Γ true, makes ψ and $\psi \rightarrow \varphi$ true, and therefore φ must be true in all such structures, showing that $\Gamma \vDash \varphi$.

$\rightarrow$**Intro** From a proof of φ on assumptions $\Gamma \cup \{\psi\}$, apply $\rightarrow$Intro to obtain a proof of $\psi \rightarrow \varphi$ on assumption Γ (discharging ψ). But since by hypothesis $\Gamma, \psi \vDash \varphi$, by the deduction theorem, $\Gamma \vDash \psi \rightarrow \varphi$.

¬**Elim** From a proof of ψ on assumptions $\Gamma, \neg\varphi$ and a proof of $\neg\psi$ on assumptions $\Gamma, \neg\varphi$, obtain a proof of φ on assumptions Γ, discharging $\neg\varphi$. Since $\Gamma, \neg\varphi \vDash \psi$ and $\Gamma, \neg\varphi \vDash \neg\psi$, it follows that $\Gamma, \neg\varphi \vDash$; therefore $\Gamma \vDash \varphi$.

¬**Intro** From a proof of ψ on assumptions Γ, φ and a proof of $\neg\psi$ on assumptions Γ, φ, obtain a proof of $\neg\varphi$ on assumptions Γ, discharging φ. Since $\Gamma, \varphi \vDash \psi$ and $\Gamma, \varphi \vDash \neg\psi$, it follows that $\Gamma, \varphi \vDash$; therefore $\Gamma \vDash \neg\varphi$.

Thus all the rules preserve soundness; no proof in $ND_{\neg,\rightarrow}$ can be constructed other than using these rules; so $ND_{\neg,\rightarrow}$ is sound. □

6.3 Completeness

Consistency A set of sentences Γ is called *(syntactically) consistent* iff $\Gamma \nvdash \neg(\varphi \rightarrow \varphi)$. Otherwise it is *inconsistent*, also written $\Gamma \vdash$. Equivalently, consider this proof:

$$
\cfrac{\neg(\varphi \rightarrow \varphi) \quad \cfrac{\begin{array}{c}\Gamma\\ \vdots \end{array} \qquad \cfrac{[\varphi]}{\varphi \rightarrow \varphi}\rightarrow \text{Intro}}{\psi}}{} \; \neg\text{Elim}
$$

If we have a proof of $\neg(\varphi \rightarrow \varphi)$, we have a proof of arbitrary ψ from the assumption Γ; so we can say that Γ is consistent iff there is a sentence *not provable* from Γ.

A set of sentences Γ is *maximally consistent* iff it is consistent and for any φ, if $\varphi \notin \Gamma$, then $\Gamma, \varphi \vdash$ (one cannot add any more sentences while remaining consistent).

Properties of Maximal Consistent Sets: Deductive Closure

Lemma 56 (Deductive Closure)

If Γ is a maximal consistent set, then if $\Gamma \vdash \varphi$, then $\varphi \in \Gamma$.

Proof. Suppose that $\Gamma \vdash \varphi$, but $\varphi \notin \Gamma$. Since Γ is maximal consistent, $\Gamma, \varphi \vdash \neg(\varphi \to \varphi)$. Using $\to$Intro and discharging φ, we see that $\Gamma \vdash \varphi \to (\neg(\varphi \to \varphi))$. But now use $\to$Elim, and obtain $\Gamma \vdash \neg(\varphi \to \varphi)$ i.e., Γ is inconsistent. $\square$

Properties of Maximal Consistent Sets: Negation Completeness

Lemma 57 (Negation Completeness)

If Γ is a maximal consistent set, then $\neg\varphi \in \Gamma$ iff $\varphi \notin \Gamma$.

Proof. L to R: If both φ and $\neg\varphi$ were in Γ, we could prove $\neg(\varphi \to \varphi)$ by an application of $\neg$Elim, showing Γ inconsistent.

R to L: If $\varphi \notin \Gamma$, then $\Gamma, \varphi \vdash \neg(\varphi \to \varphi)$:

$$
\cfrac{\cfrac{\Gamma, [\varphi]}{\vdots} \qquad \cfrac{[\varphi]}{\varphi \to \varphi} \to \text{Intro}}{\neg(\varphi \to \varphi) \qquad \varphi \to \varphi} \; \neg\text{Intro}
$$
$$
\cfrac{}{\neg\varphi}
$$

Therefore, $\Gamma \vdash \neg\varphi$; by Deductive Closure Lemma, $\neg\varphi \in \Gamma$. $\square$

Properties of Maximal Consistent Sets: Conditional Completeness

Lemma 58 (Conditional Completeness)

If Γ is any maximal consistent set, then $\varphi \to \psi \in \Gamma$ iff whenever $\varphi \in \Gamma$, it follows that $\psi \in \Gamma$.

Proof. L to R: Suppose $\varphi \to \psi \in \Gamma$. Then $\Gamma \vdash \varphi \to \psi$; and if $\varphi \in \Gamma$, then $\Gamma \vdash \varphi$. Apply $\to$Elim, and establish $\Gamma \vdash \psi$, i.e., $\psi \in \Gamma$.

R to L: Suppose that whenever $\varphi \in \Gamma$, $\psi \in \Gamma$. Consider two cases:

1. $\varphi \in \Gamma$. So $\Gamma = \Gamma \cup \{\varphi\}$. Moreover, $\psi \in \Gamma$, so $\Gamma, \varphi \vdash \psi$. Apply $\to$Intro and discharge φ to show $\Gamma \vdash \varphi \to \psi$.

2. $\varphi \notin \Gamma$. Then $\Gamma, \varphi \vdash \neg(\varphi \to \varphi)$. But then $\Gamma, \varphi \vdash \psi$ (by a similar proof as just above); applying $\to$Intro and discharge φ, $\Gamma \vdash \varphi \to \psi$. $\qquad \square$

Properties of Maximal Consistent Sets: Satisfiability

Theorem 59 (Maximal Consistent Sets Satisfiable)

Every maximal consistent set is satisfiable (i.e., has a model).

Proof. Suppose Γ is a maximal consistent set in $\mathscr{L}_{\to,\neg}$. For s a sentence letter, define a structure $\mathscr{A}$: $\mathscr{A}(s) = T$ iff $s \in \Gamma$.

For any φ, $|\varphi|_{\mathscr{A}} = T$ iff $\varphi \in \Gamma$. Proof by induction on complexity of sentences. The base case, φ a sentence letter, is immediate.

Suppose φ is complex. There are two cases:

1. $\varphi = \neg\psi$. $\varphi \in \Gamma$ iff $\psi \notin \Gamma$, by Negation Completeness; iff by the induction hypothesis, $|\psi|_{\mathscr{A}} = F$; iff by the clause on $\neg$, $|\varphi|_{\mathscr{A}} = T$.

2. $\varphi = (\psi \to \chi)$. $\varphi \in \Gamma$ iff if $\psi \in \Gamma$ then $\chi \in \Gamma$ (by Conditional Completeness); iff if $|\psi|_{\mathscr{A}} = T$ then $|\chi|_{\mathscr{A}} = T$; iff $|\varphi|_{\mathscr{A}} = T$. $\qquad \square$

Completeness

Theorem 60 (Completeness for $\mathscr{L}_{\neg,\to}$)

If $\Gamma \vDash \varphi$, then $\Gamma \vdash \varphi$.

Proof. We prove the *contrapositive*. So assume that $\Gamma \nvdash \varphi$. (Assuming that Γ itself is consistent; if it is not, then clearly the theorem holds fairly trivially.)

We'll construct a maximal consistent set Γ^+, a *superset* of Γ, such that $\Gamma^+ \nvdash \varphi$. By consistency of Γ^+, $\varphi \notin \Gamma^+$; by the Negation Completeness lemma, $\neg\varphi \in \Gamma^+$.

But by the Satisfiability Theorem, Γ^+ has a model, $\mathscr{A}$. Since $\Gamma \subseteq \Gamma^+$, for all $\gamma \in \Gamma$, $|\gamma|_{\mathscr{A}} = T$; but since $\neg\varphi \in \Gamma^+$, $|\neg\varphi|_{\mathscr{A}} = T$, so $|\varphi|_{\mathscr{A}} = F$. So $\Gamma \nvDash \varphi$, as required.

We now show how to construct Γ^+ from Γ, completing the proof. $\qquad \square$

The Construction of Maximal Consistent Γ^+   Supposing $\Gamma \nvdash \varphi$, construct Γ^+ as follows.

1. Enumerate the sentences of $\mathscr{L}_{\neg,\to}$, $\{\sigma_1, \sigma_2, \dots\}$.

2. Let $\Gamma_0 = \Gamma$.

3.

$$\Gamma_{n+1} = \begin{cases} \Gamma_n \cup \{\neg\sigma_{n+1}\} & \text{if } \Gamma_n, \sigma_{n+1} \vdash \varphi; \\ \Gamma_n \cup \{\sigma_{n+1}\} & \text{otherwise (i.e., if } \Gamma_n, \sigma_{n+1} \nvdash \varphi). \end{cases} \text{☒}$$

4. Let $\Gamma^+ = \bigcup_i \Gamma_i$.

It is clear that $\Gamma \subseteq \Gamma^+$. It is also clear that $\Gamma^+ \nvdash \varphi$, for at no stage was a sentence added to any Γ_n that permitted a proof of φ.

Finally, Γ^+ is a *maximal consistent set*, since (by construction), for every σ_i, either $\sigma_i \in \Gamma^+$ or $\neg\sigma_i \in \Gamma^+$. Thus Γ^+ has the properties required to prove Completeness stated just above.

Completeness for $\mathscr{L}_1$ We've shown the Completeness Theorem for the restricted language $\mathscr{L}_{\neg,\to}$. But by Expressive Adequacy, any sentence of $\mathscr{L}_1$ can be expressed in $\mathscr{L}_{\neg,\to}$. So we can assure ourselves that any truth function expressible in $\mathscr{L}_1$ can be expressed in $\mathscr{L}_{\neg,\to}$, and any valid argument in $\mathscr{L}_1$ has a corresponding provably correct sequent in $\mathscr{L}_{\neg,\to}$.

For completeness of $\mathscr{L}_1$, we cannot rest there. What if the rules of $\mathscr{L}_1$ don't suffice to show the equivalence of some arbitrary $\mathscr{L}_1$ sentence φ with a sentence only involving $\neg$ and $\to$? In that case, while a sentence logically equivalent to φ is provable, φ may not be.

We thus need additional lemmas showing that maximal consistent sets

- Contain $\varphi \wedge \psi$ iff they contain both φ and ψ;

- Contain $\varphi \vee \psi$ iff they contain at least one of φ or ψ;

- Contain $\varphi \leftrightarrow \psi$ iff they contain φ iff they contain ψ.

Then the Satisfiability Theorem must be extended to include these new cases; and the remainder of the theorem is proved as above.

Compactness Again   Consideration of the nature of natural deduction proofs should convince you that any conclusion φ drawn on the basis of assumptions Γ can be constructed by some finite application of the rules to finitely many sentences in Γ. Provability is essentially a finite notion: proofs are things that can be *constructed*.

   This observation, informal though it is, *can* be made precise, to yield another proof of the *Compactness theorem*. Since, by Completeness, every correct semantic sequent $\Gamma \vDash$ is provable, and every proof is a finite object using only finitely many members of Γ, then by Soundness there is a correct semantic sequent $\Gamma' \vDash$ where Γ' is a finite subset of Γ.

6.4   A Little More Proof Theory

'Proof theory' is the mathematical examination of proofs conceived of as mathematical objects in their own right. In this section we explore a little of what this involves. We've already seen one some proof theory in section 1 of this chapter, where we showed the redundancy of the proof rule ¬Elim in the presence of ¬¬Elim, and vice versa.

The Language $\mathscr{L}_{DP}$ Consider the language $\mathscr{L}_{DP}$ which contains all sentence letters, and all sentences involving only the connectives $\wedge, \rightarrow$, and also contains a *sentential constant* $\bot$. A sentential constant is syntactically like a sentence letter, but (unlike sentence letters) it gets a constant interpretation in every structure – in this case, for any structure $\mathscr{A}$, $|\bot|_{\mathscr{A}} = F$. The atomic sentences of $\mathscr{L}_{DP}$ thus include $\bot$ and all sentence letters; complex sentences are then built up in the usual way.

$$
\begin{array}{cc}
[\varphi] & [\varphi] \\
\vdots & \vdots \\
\psi & \psi \to \bot \\
\end{array}
$$

$$\dfrac{}{\varphi \to \bot}\ \neg\text{Intro}_{DP} \qquad \dfrac{}{\varphi}\ \neg\text{Elim}_{DP}$$

$$
\begin{array}{cc}
[\varphi \to \bot] & [\varphi \to \bot] \\
\vdots & \vdots \\
\psi & \psi \to \bot \\
\end{array}
$$

Table 6.2: Negation rules in $\mathscr{L}_{DP}$

Translating between $\mathscr{L}_{DP}$ and $\mathscr{L}_1$ It's clear that the $\mathscr{L}_1$ sentence '$\neg\varphi$' expresses the same truth function as the $\mathscr{L}_{DP}$ sentence '$\varphi \to \bot$' (consideration of a truth table will suffice to show this). Using this translation, we can give $\mathscr{L}_{DP}$-versions of the natural deduction rules for $\neg$ in $\mathscr{L}_1$, as seen in Table 6.2.

With the negation rules as specified, we can prove this:

Theorem 61 (Alternative Negation Rules for $\mathscr{L}_{DP}$)
Π *is a proof of some sequent in the language $\mathscr{L}_{DP}$ that makes a use u of one of the Table 6.2 versions of Halbach's rules iff there is another proof Π' of the same sequent which replaces that use u by a construction which only uses this rule for $\neg$ (together with the rules for $\to$):*

$$
\begin{array}{c}
[\varphi \to \bot] \\
\vdots \\
\bot \\
\hline
\varphi
\end{array}\ \bot_{C}
$$

Proof. If: There are two cases. Case 1: Suppose Π is a proof that makes use of $\neg\text{Intro}_{DP}$. Then we can replace that use by the following construction, which yields another proof Π' which uses only the rules for $\to$:

$$
\begin{array}{cc}
[\varphi] & [\varphi] \\
\vdots & \vdots \\
\psi & \psi \to \bot \\
\end{array}
$$
$$\dfrac{}{\bot}\ \to \text{Elim}$$
$$\dfrac{}{\varphi \to \bot}\ \to \text{Intro}$$

Case 2: Suppose Π is a proof that makes use of $\neg\mathrm{Elim}_{DP}$. Then we can replace that use by the following construction, which yields another proof Π' which uses only the new rule $\bot_C$ and the rule $\rightarrow\mathrm{Elim}$:

$$
\cfrac{\cfrac{\begin{array}{cc} [\varphi \rightarrow \bot] & [\varphi \rightarrow \bot] \\ \vdots & \vdots \\ \psi & \psi \rightarrow \bot \end{array}}{\bot} \rightarrow \mathrm{Elim}}{\varphi} \bot_C
$$

Only if: Suppose Π is a proof which makes use of the new rule $\bot_C$. Then we may replace that use by the following construction which uses only the old rules for $\rightarrow$ and $\neg\mathrm{Elim}_{DP}$.

□

$$
\cfrac{\cfrac{\begin{array}{c} [\varphi \rightarrow \bot] \\ \vdots \\ \bot \end{array} \qquad \cfrac{[\bot]}{\bot \rightarrow \bot} \rightarrow \mathrm{Intro}}{}}{\varphi} \neg\mathrm{Elim}_{DP}
$$

Reducing Complexity of Proofs   Return now to the $\mathscr{L}_1$-fragment containing only $\rightarrow, \wedge, \neg$. Let the *(degree of) complication* of a sentence be the number of connectives occurring in the sentence (a sentence letter has degree of complication 0, if φ and ψ have degree of complication m and n respectively, then $(\varphi \wedge \psi)$ has degree of complication $m + n + 1$, etc.) We can now show:

Theorem 62 (Reducing Complexity of Proofs)
If Π is a proof in Halbach's system for $\mathscr{L}_{\rightarrow,\wedge,\neg}$ that $\Gamma \vdash \varphi$, in which the most complicated sentence resulting from an application of $\neg$Elim is ψ, then (when ψ is not atomic) there is a distinct proof Π' of $\Gamma \vdash \varphi$ in which the results of all applications of $\neg$Elim are less complicated *than ψ. Indeed, there exists a proof $\Pi^{\dagger}$ of that sequent, in which the result of every application of $\neg$Elim is atomic.*

Proof. Suppose that the most complicated result of ¬Elim in Π is ψ. So part of Π looks like this:

$$
\begin{array}{ccc}
\Gamma \quad [\neg\psi] & \Delta \quad [\neg\psi] \\
\vdots & \vdots \\
\chi & \neg\chi \\
\hline
\multicolumn{2}{c}{\psi} \quad \text{¬Elim}
\end{array}
$$

ψ is complicated (not a sentence letter), so given there are three connectives in this language, there are three cases:

1. $\psi = \neg\pi$. In which case replace that part of the proof sketched above with this one:

$$
\begin{array}{cc}
\cfrac{[\pi] \quad [\neg\pi]}{\neg\neg\pi}\text{¬Intro} & \cfrac{[\pi] \quad [\neg\pi]}{\neg\neg\pi}\text{¬Intro} \\
\Gamma & \Delta \\
\vdots & \vdots \\
\chi & \neg\chi \\
\hline
\multicolumn{2}{c}{\neg\pi} \quad \text{¬Intro}
\end{array}
$$

2. $\psi = \pi \wedge \xi$. Proof left for Exercise.

3. $\psi = \pi \rightarrow \xi$. In which case replace that part of the proof sketched above with this one:

$$
\begin{array}{cc}
\cfrac{\cfrac{[\pi \rightarrow \xi] \quad [\pi]}{\xi}\rightarrow\text{Elim} \quad [\neg\xi]}{\neg(\pi \rightarrow \xi)}\text{¬Intro} & \cfrac{\cfrac{[\pi \rightarrow \xi] \quad [\pi]}{\xi}\rightarrow\text{Elim} \quad [\neg\xi]}{\neg(\pi \rightarrow \xi)}\text{¬Intro} \\
\Gamma & \Delta \\
\vdots & \vdots \\
\chi & \neg\chi \\
\hline
\multicolumn{2}{c}{\cfrac{\xi}{\pi \rightarrow \xi}\rightarrow\text{Intro}} \quad \text{¬Elim}
\end{array}
$$

Notice that applying these replacements repeatedly to uses of ¬Elim in a proof will eventually replace all complex sentences in uses of ¬Elim

by simpler ones. All sentences of $\mathscr{L}_1$ are finitely complex, and all proofs involve only finitely many applications of a given rule, this process terminates after a finite time with only atomic sentences as the conclusions of uses of ¬Elim. □

It is clear that we can put the results of Theorems 61 and 62 together, to establish that, in the system which involves the normal rules for ∧ and →, and has ⊥$_C$ as its only other rule, if Π is a proof in the system DP that $\Gamma \vdash_{DP} \varphi$, then there exists a proof of that sequent, $\Pi^\dagger$, in which the result of any application of ⊥$_C$ is atomic. (I leave the proof as a problem.)

6.5   Axioms

Axiomatic Systems   Another proof system we will now briefly consider is an *axiomatic* (or Hilbert-style) proof system.

This is familiar in mathematics: one begins with some *axioms* (normally interpreted as 'obvious' or 'certain' truths), and by applications of some very simple truth-preserving rules of inference, one derives further truths.

If one's axioms are well chosen, one should be able to prove all and only truths about a given subject matter from the axioms characterising that subject matter. This was Euclid's method in the *Elements*: he showed that all and only the truths of Euclidean geometry could be established on the basis of his well chosen axioms.

Axioms for Sentential Logic   In an axiomatic system, one begins with theorems, the axioms, and applies the rules of inference to preserve theoremhood.

Our axiom system Ł is very simple. It has three *axiom-schemata*, in the language $\mathscr{L}_{\neg,\rightarrow}$:

A1 $\vdash_{\text{Ł}} (\varphi \rightarrow (\psi \rightarrow \varphi))$;

A2 $\vdash_{Ł} ((\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi)))$;

A3 $\vdash_{Ł} ((\neg\psi \to \neg\varphi) \to (\varphi \to \psi))$.

Any sentence of $\mathscr{L}_{\neg,\to}$ can be substituted into these axiom schemata to generate an instance of an axiom. Any axiom is a theorem.

The system has one rule (apart from substitution): Modus Ponens, then rule that if $\vdash_{Ł} \varphi$ and $\vdash_{Ł} \varphi \to \psi$, then $\vdash_{Ł} \psi$.

A *proof* in Ł is any sequence of theorems, each of which is either an axiom or follows by *modus ponens* from earlier theorems.

A Proof in Ł   The system Ł is elegant, but proofs in it can be unnatural to say the least: (I annotate lines of the proof with the axiom scheme used.)

1. $\vdash_{Ł} (P \to ((P \to P) \to P)) \to ((P \to (P \to P)) \to (P \to P))$ (A2)

2. $\vdash_{Ł} (P \to ((P \to P) \to P))$ (A1)

3. $\vdash_{Ł} ((P \to (P \to P)) \to (P \to P))$ (MP 1, 2)

4. $\vdash_{Ł} (P \to (P \to P))$ (A1)

5. $\vdash_{Ł} (P \to P)$ (MP 3, 4)

This is (apparently) the *shortest* proof of $P \to P$ in Ł!

Soundness of Ł   It is a trivial exercise (and hence left for an exercise) to show that Ł is sound. That is, show:

- All the axioms of Ł are tautologies.

- The rule of Ł preserves tautologousness.

Equivalence of Ł and $ND_{\neg,\to}$ We now show that, perhaps surprisingly, the proof systems Ł and $ND_{\neg,\to}$ are equivalent – everything that can be proved in one can be proved in the other. This shows, too, that Ł is sound and complete.

Everything provable in Ł is provable in ND

Theorem 63

If $\vdash_{\overline{Ł}} \varphi$ then $\vdash \varphi$.

Proof. The proof is straightforward: show that each axiom is provable in ND, and show that the rule *modus ponens* is a valid rule in ND.

The second part is easy; if $\vdash \varphi$ and $\vdash \varphi \rightarrow \psi$, then appending an instance of →Elim to the preceding proofs is a proof of ψ.

Then it is just a matter of proving the axioms. I show A1:

$$\frac{\dfrac{[\varphi]}{\psi \rightarrow \varphi} \rightarrow \text{Intro}}{\varphi \rightarrow (\psi \rightarrow \varphi)} \rightarrow \text{Intro}$$

$\square$

Proofs from Assumptions in Ł   The ugly proof of $\vdash_{\overline{Ł}} P \rightarrow P$ two paragraphs back shows the benefits of working with *assumptions*. A proof of φ from assumptions Γ in Ł is a sequence of axioms, *members of* Γ, or follows from preceding members by *modus ponens*, and the last line of which is φ. This is a legitimate notion, because one can show:

Theorem 64 (Deduction Theorem for Ł)

$\Gamma \vdash_{\overline{Ł}} \varphi$ *iff for some finite set of sentences* $\gamma_1, \dots, \gamma_n$ *actually used in the proof of φ,* $\vdash_{\overline{Ł}} (\gamma_1 \rightarrow (\gamma_2 \rightarrow \dots (\gamma_n \rightarrow \varphi) \dots))$.

It suffices to show that $\Gamma, \psi \vdash_{\overline{Ł}} \varphi$ iff $\Gamma \vdash_{\overline{Ł}} \psi \rightarrow \varphi$. I leave the proof for an exercise.

Everything Provable in ND is provable in Ł

Theorem 65

If $\vdash \varphi$, then $\vdash_{\overline{Ł}} \varphi$.

Proof. Induction on length of proofs: we show that the shortest proofs of ND are provable in Ł, and that the rules are provable transitions.

Certainly the shortest proof, the single node φ, shows that $\varphi \vdash \varphi$; but $\varphi \vdash_{\text{Ł}} \varphi$ is shown by the deduction theorem given $\vdash_{\text{Ł}} \varphi \to \varphi$.

Suppose we extend some existing proofs by applying the rules of ND:

→**Elim** We have $\Gamma \vdash \varphi$ and $\Gamma \vdash \varphi \to \psi$, and extend by →Elim. But this is obviously *modus ponens*, so $\Gamma \vdash_{\text{Ł}} \psi$.

The case of →Intro is shown by the deduction theorem. I leave you to show the cases for negation. □

Further Reading

Natural deduction was invented by Gentzen (1969), the founder of proof theory. The proof theory of natural deduction systems discussed in section 4 of this chapter is based on Prawitz (2006: 39–41); he uses Theorems 61 and 62, and the last problem, and proves the famous 'cut elimination' theorem for natural deduction, unfortunately a topic beyond the scope of these notes.

The axiom system here was developed by Jan Łukasiewicz, simplifying Frege's system: see Borkowski and Słupecki (1958: p. 25). (The Polish notation used in this article is initially confusing, but has the wonderful feature that it is entirely unambiguous without the use of parentheses.) See also Bostock (1997: ch. 5–6); he gives a direct proof of the completeness of a system based on Ł at pp. 217–9.

Exercises

1. (a) Devise introduction and elimination rules for a natural deduction system for the language $\mathscr{L}_\uparrow$ with the Sheffer Stroke as its only connective, being sure to fully justify your answer. How many such rules do we need?

   (b) Suppose in a system of natural deduction ND for $\mathscr{L}_1$ we replaced the ¬Elim rule by the following, *ex falso quodlibet*:

$$\frac{\psi \qquad \neg\psi}{\varphi} \ \text{EFQ}$$

i. Is the resulting system ND' equivalent (in terms of what can be proved) to the original system? (Justify your answer; it is somewhat difficult to prove conclusively, but give some reasons for your view.)

ii. What happens if we, in addition, replace ¬Intro by the following rule, *tertium non datur*, yielding the system ND''?

$$\frac{[\varphi] \quad [\neg\varphi]}{\underset{\psi}{\vdots} \quad \underset{\psi}{\vdots}} \text{ TND}$$

(c) Consider the natural deduction system $ND_{\to,\lor}$ for the language $\mathscr{L}_{\to,\lor}$ in which the only connectives are $\to, \lor$. Suppose we introduce to the language the zero-place sentential constant $\bot$, and add the following rules to the proof system, yielding $ND_{\to,\lor,\bot}$:

$$\frac{\bot}{\varphi} \bot \qquad \frac{[\varphi \to \bot]}{\underset{\bot}{\vdots}} \text{ C}$$

i. Give a natural translation between $\mathscr{L}_{\to,\lor,\bot}$ and $\mathscr{L}_{\to,\lor,\neg}$.

ii. Using that translation, can you show that the proof system $ND_{\to,\lor,\bot}$ just introduced is equivalent to the natural deduction system $ND_{\to,\lor,\neg}$ involving just Halbach's rules for $\to$, $\lor$ and $\neg$?

(d) A mystery connective $\oplus$ has the following introduction and elimination rules:

$$\frac{\varphi \oplus \psi}{\psi} \oplus \text{Elim} \qquad \frac{\varphi}{\varphi \oplus \psi} \oplus \text{Intro}$$

i. Show that no system containing a connective defined by these rules is sound.

ii. What limits should be placed on the ability of introduction and elimination rules to define or characterise the inferential role of

a connective in light on this result? (This question invites discussion, not a definitive answer. You may wish to consult Prior (1960).)

(e) Relatedly to 1.(d).ii:

i. Check, by means of truth tables, that $(((P \rightarrow Q) \rightarrow P) \rightarrow P)$ is a tautology.

ii. Give a natural deduction proof of $(((P \rightarrow Q) \rightarrow P) \rightarrow P)$.

iii. Every valid argument involving only sentences containing $\wedge$ as their only connective can be proved valid using just the rules $\wedge$Intro and $\wedge$Elim. In that sense, $\wedge$ is completely characterised by its introduction and elimination rules. What do the previous results show about $\rightarrow$ in this connection?

2. Using the model soundness proof for the restricted natural deduction system $ND_{\neg,\rightarrow}$, show, by analysing proofs constructed using the remaining rules of the full Halbach system ND, that the system ND is sound.

3. The Completeness proof for $\mathscr{L}_{\neg,\rightarrow}$ relied on various lemmas concerning the behaviour of maximal consistent sets with regard to the connectives $\neg$ and $\rightarrow$. Show that

(a) analogous lemmas hold for $\wedge$, $\vee$ and $\leftrightarrow$.

(b) the Satisfiability Theorem holds for the full language $\mathscr{L}_1$.

(c) the full Natural Deduction system is complete for $\mathscr{L}_1$ with respect to the intended semantics.

4. (a) Show that the axiomatic system Ł introduced is sound.

(b) Show that the axioms A2 and A3 of Ł are provable in natural deduction:

i. A2. $\vdash_{\text{Ł}} ((\varphi \rightarrow (\psi \rightarrow \chi)) \rightarrow ((\varphi \rightarrow \psi) \rightarrow (\varphi \rightarrow \chi)))$;

ii. A3. $\vdash_{\text{Ł}} ((\neg\psi \rightarrow \neg\varphi) \rightarrow (\varphi \rightarrow \psi))$.

5. Suppose that $\Gamma, \psi \vdash_{\text{Ł}} \varphi$, shown by a proof Π. Define the ψ-*transform* of Π, written Π^{ψ}, as the sequence that results from prefixing '$\psi \rightarrow$' to the front of every sentence in Π. Show that each sentence in Π^{ψ} is provable from assumptions Γ in Ł. Since obviously ψ appears on Π, $\psi \rightarrow \psi$ appears on Π^{ψ}, and that is provable from Γ. So you need to show that

(a) the prefixed axioms (i.e., those $\psi \to \delta$ in Π^ψ where δ is an axiom) are provable from Γ (hint: use A1);

(b) the prefixed assumptions (i.e., those $\psi \to \gamma$ in Π^ψ where $\gamma \in \Gamma$) are provable from Γ;

(c) If χ followed by *modus ponens* from early claims in Π, $\psi \to \chi$ follows from earlier prefixed claims in Π^ψ (hint: use A2).

Using these results, show the deduction theorem for Ł.

6. Show the other cases for the equivalence of ND and Ł:[1]

(a) Show that if $\Gamma, \varphi \vdash_{\text{Ł}} \psi$ and $\Gamma, \varphi \vdash_{\text{Ł}} \neg\psi$, then $\Gamma \vdash_{\text{Ł}} \neg\varphi$.

(b) Show that if $\Gamma, \neg\varphi \vdash_{\text{Ł}} \psi$ and $\Gamma, \neg\varphi \vdash_{\text{Ł}} \neg\psi$, then $\Gamma \vdash_{\text{Ł}} \varphi$. (Equivalently, show that if $\Gamma \vdash_{\text{Ł}} \neg\neg\varphi$, then $\Gamma \vdash_{\text{Ł}} \varphi$.)

7. A fellow student argues as follows:

> If the Natural Deduction system introduced by Halbach is complete, then whenever $\Gamma \vDash \varphi$, there is a proof that shows $\Gamma \vdash \varphi$. But proofs are finite; so if Γ is infinite, there is no proof that $\Gamma \vdash \varphi$ – so the system is not complete!

How should you respond?

8. Prove the omitted case (where the complex sentence is a conjunction) in the proof of Theorem 62.

9. Show that, in the system which involves the normal rules for $\wedge$ and $\to$, and has $\bot_C$ as its only other rule, if Π is a proof in the system DP that $\Gamma \vdash_{DP} \varphi$, then there exists a proof of that sequent, $\Pi^\dagger$, in which the result of any application of $\bot_C$ is atomic.

---

[1]These proofs are quite elusive; I propose treating this question as optional.

Chapter 7

The Syntax and Semantics of $\mathscr{L}_2$

7.1 Syntax of $\mathscr{L}_2$

$\mathscr{L}_2$ We now have a good grip on $\mathscr{L}_1$. But many valid arguments in English *can't* be captured by a valid argument in sentential logic.

One reason for this is the lack of *non-truth-functional* connectives in $\mathscr{L}_1$ (consider English 'possibly', 'it will be that', 'it ought to be that', or even 'S knows that'). We are not going to remedy this lack now.

Another reason for the inability of $\mathscr{L}_1$ to capture valid arguments is that the analysis $\mathscr{L}_1$ provides of an English sentence stops at the truth-functional structure of the sentence. But English sentences have additional *internal structure* which arguments may make use of.

Consider: 'John loves James; therefore, John loves someone'. This is valid, but the best $\mathscr{L}_1$ formalisation is the incorrect sequent: $P \vDash Q$. The validity of this argument depends on the internal structure of the sentences. So now we look at a logic with the resources to formalise this additional structure.

The Syntax of $\mathscr{L}_2$: $\mathscr{L}_1$ alphabet and Predicate Letters   The alphabet of $\mathscr{L}_2$ – the constituents out of which $\mathscr{L}_2$-sentences may be formed – includes the alphabet of $\mathscr{L}_1$. So it contains infinitely many sentence letters, as well as the logical connectives $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$, and the parentheses (,).

The alphabet of $\mathscr{L}_2$ contains *predicate letters*. A predicate letter has *argument places* (intuitively, representing the number of designators or variables required to make a grammatical formula). The number of argument places of a predicate letter is called its *arity*. For each arity $n \geqslant 0$, we introduce infinitely many predicate letters: $P^n, Q^n, R^n, P_1^n, Q_1^n, R_1^n, \ldots$. Where $n = 0$, we omit the index. (So any expression of the form P_k^n, Q_k^n, or R_k^n, where n, k are either missing or a numeral '1', '2', etc. will be a predicate letter.)

We permit zero-place predicate letters, and identify these with the sentence letters of $\mathscr{L}_1$.

The Syntax of $\mathscr{L}_2$: Constants, Variables, Quantifiers   The alphabet of $\mathscr{L}_2$ includes infinitely many *constants*, $a, b, c, a_1, b_1, c_1, \ldots$.

$\mathscr{L}_2$ also contains infinitely many *variables*, $x, y, z, x_1, y_1, z_1, \ldots$.

And $\mathscr{L}_2$ contains infinitely many *quantifiers*, $\forall v$ and $\exists v$, where v is any variable. $\forall$ is the *universal* quantifier; $\exists$ is the *existential* (sometimes, *particular*) quantifier.

Variables and constants are collectively known as *terms*.

Atomic Formulae and Formulae of $\mathscr{L}_2$

Definition 57 (Atomic Formula). If Φ^n is a predicate letter of arity $n \geqslant 0$, and $\tau_1, \ldots, \tau_n$ are n terms (not necessarily all distinct), then $\Phi\tau_1 \ldots \tau_n$ is an *atomic formula* of $\mathscr{L}_2$.

It follows that all sentence letters are atomic formulae.

Definition 58 (Formula).

1. Any atomic formula is a formula of $\mathscr{L}_2$.

2. If φ and ψ are formulae of $\mathscr{L}_2$, then so are $\neg\varphi$, $(\varphi \lor \psi)$, $(\varphi \land \psi)$, $(\varphi \to \psi)$, and $(\varphi \leftrightarrow \psi)$.

3. If $\forall v$ and $\exists v$ are quantifiers and φ is a formula, then so are $\forall v \varphi$ and $\exists v \varphi$.

4. Nothing else is a formula of $\mathscr{L}_2$.

We permit brackets and the arity of predicate letters to be dropped in accordance with Halbach, §4.4.

Free and Bound Occurrences of Variables

Definition 59 (Free and Bound Variable Occurrences).

1. All variable occurrences in atomic formulae are free.

2. If some variables occur freely in φ and ψ, they remain free in $\neg\varphi$, $(\varphi \lor \psi)$, $(\varphi \land \psi)$, $(\varphi \to \psi)$, and $(\varphi \leftrightarrow \psi)$.

3. No occurrence of v is free in $\forall v\varphi$ or $\exists v\varphi$; all other variable occurrences in φ remain free in $\forall v\varphi$.

A variable occurrence is *bound* iff it is not free.

A variable *occurs freely* in φ iff there is at least one free occurrence of that variable in φ. If a variable occurs freely in φ, φ is an *open formula*.

Definition 60 (Sentence of $\mathscr{L}_2$). An $\mathscr{L}_2$-formula φ is a *sentence* (or 'closed formula') of $\mathscr{L}_2$ iff no variable occurs freely in φ.

7.2   Semantics of $\mathscr{L}_2$

Semantics for $\mathscr{L}_2$: Domains   As in $\mathscr{L}_1$, the idea is to specify meanings (or *extensions*) for the basic parts of the language, and systematically show how the meaning of complex expressions depends on the meanings of the parts and their relations. But now we have sub-sentential constituents of formulae, which cannot be assigned truth-values as their extensions, so $\mathscr{L}_1$ structures cannot suffice for $\mathscr{L}_2$.

We begin with the notion of a *domain D*. This is a non-empty set of objects – any objects – from which are drawn the *semantic values* of constants. (So the meaning of a constant will be an object.) The meaning of a predicate letter will be a *property* or *relation* on the domain. For our purposes, a subset of the domain will suffice for a property (intuitively, those things in the domain which have the property); and a subset of $D \times D$ will suffice for a 2-place relation. In general, then, the semantic value of a *n*-ary predicate letter will be a subset of $D \underbrace{\times \ldots \times}_{n} D = D^n$.

The attentive reader will have noticed a potential oddity here. The semantic value assigned to a 2-place predicate letter will be a set of pairs, and that assigned to a 3-place predicate letter will be a set of triples, etc. Shouldn't the semantic value assigned to a 1-place monadic predicate letter be a set of 1-tuples, rather than a subset of the domain? That is, shouldn't the semantic value of a predicate be something like $X = \{\langle x_1 \rangle, \ldots, \langle x_n \rangle, \ldots\}$, rather than $X' = \{x_1, \ldots, x_n, \ldots\}$? The still more attentive reader will recall that, by Definition 12, $\langle x_1 \rangle = x_1$, so that $X = X'$. A subset of the domain turns out to be a set of ordered 1-tuples, so the interpretation of all *n*-ary predicate letters is uniform: a set of *n*-tuples.

$\mathscr{L}_2$ Structures and Variable Assignments

Definition 61 ($\mathscr{L}_2$-Structure).   An $\mathscr{L}_2$-structure $\mathscr{A}$ is an ordered pair $\langle D, I \rangle$ where D ('the domain') is any *non-empty* set, and I ('the interpretation')

is a function

- which assigns an element of D to every constant;

- which assigns a subset of D^n to every predicate letter of arity $n > 0$;

- and which assigns a truth value (T or F) to any predicate letter of arity 0 (sentence letter).

Variables do not have a fixed extension, even in a particular structure. But they may be 'temporarily' assigned semantic values:

Definition 62 (Variable Assignment). A *variable assignment* α over $\mathscr{A}$ is a function which assigns each variable in $\mathscr{L}_2$ a member of $D_\mathscr{A}$.

Semantic Values of Simple Expressions   Let us extend the notation $|\cdot|_\mathscr{A}$ to mean the semantic value of any $\mathscr{L}_2$ expression in structure $\mathscr{A}$.

For formulae not involving variables, the assignment of semantic values (in particular, truth values) to complex formulae in a given structure will be straightforward. But what to do about variables – in particular, assigning semantic values to open formulae (those with free variables) and quantified formulae – poses some delicate issues. Here I use an approach due to Tarski; later I explore some alternatives.

Begin by defining, for structure $\mathscr{A}$ and variable assignment α, $|\varphi|_\mathscr{A}^\alpha = I_\mathscr{A}(\varphi)$ for all φ in the domain of $I_\mathscr{A}$. (All constants, and predicate letters, including sentence letters, thus receive a semantic value at this stage.) The variables of $\mathscr{L}_2$ get the obvious interpretation: $|v|_\mathscr{A}^\alpha = \alpha(v)$. We now assign truth values to the formulae of $\mathscr{L}_2$.

Satisfaction

Definition 63 (Satisfaction). Supposing φ, ψ to be formulae of $\mathscr{L}_2$, v a variable, $\mathscr{A}$ a $\mathscr{L}_2$-structure, and α a variable assignment over $\mathscr{A}$. We say that α *satisfies* φ under $\mathscr{A}$, $|\varphi|_\mathscr{A}^\alpha = T$, as follows:

- $\left|\Phi\tau_1\ldots\tau_n\right|_{\mathscr{A}}^{\alpha} = T$ iff $\langle\left|\tau_1\right|_{\mathscr{A}}^{\alpha},\ldots,\left|\tau_n\right|_{\mathscr{A}}^{\alpha}\rangle \in |\Phi|_{\mathscr{A}}^{\alpha}$, where Φ is a n-ary predicate letter ($n > 0$), and each τ_i is a term.[1]

- $|\neg\varphi|_{\mathscr{A}}^{\alpha} = T$ iff $|\varphi|_{\mathscr{A}}^{\alpha} = F$.

- $|\varphi \wedge \psi|_{\mathscr{A}}^{\alpha} = T$ iff $|\varphi|_{\mathscr{A}}^{\alpha} = |\psi|_{\mathscr{A}}^{\alpha} = T$.

- $|\varphi \vee \psi|_{\mathscr{A}}^{\alpha} = T$ iff $|\varphi|_{\mathscr{A}}^{\alpha} = T$ or $|\psi|_{\mathscr{A}}^{\alpha} = T$.

- $|\varphi \rightarrow \psi|_{\mathscr{A}}^{\alpha} = T$ iff $|\varphi|_{\mathscr{A}}^{\alpha} = F$ or $|\psi|_{\mathscr{A}}^{\alpha} = T$.

- $|\varphi \leftrightarrow \psi|_{\mathscr{A}}^{\alpha} = T$ iff $|\varphi|_{\mathscr{A}}^{\alpha} = |\psi|_{\mathscr{A}}^{\alpha}$.

- $|\forall \upsilon\varphi|_{\mathscr{A}}^{\alpha} = T$ iff $|\varphi|_{\mathscr{A}}^{\beta} = T$ for all variable assignments β over $\mathscr{A}$ differing from α *at most* in their assignment to υ.

- $|\exists \upsilon\varphi|_{\mathscr{A}}^{\alpha} = T$ iff $|\varphi|_{\mathscr{A}}^{\beta} = T$ for at least one variable assignment β over $\mathscr{A}$ differing from α *at most* in their assignment to υ.

Explaining Satisfaction The definition of satisfaction may appear puzzling at first.

The clause for atomic formulae is fairly straightforward. A variable assignment satisfies an atomic formula in a structure just in case the ordered sequence of the objects which are the semantic values of the constants and variables, under that assignment, have the property or stand in the relation which is the semantic value of the predicate letter.

The quantifier clauses are also fairly clear:

- A universally quantified sentence $\forall \upsilon\varphi$ is satisfied just in case φ is satisfied in every variable assignment that holds everything fixed except possibly the semantic value of υ – i.e., no matter what υ denotes, φ is satisfied.

---

[1]Recall the discussion of ordered 1-tuples at the beginning of this section when applying this definition to monadic predicates.

- An existentially quantified sentence $\exists v\varphi$ is satisfied just in case φ is satisfied in at least one variable assignment that holds everything fixed except possibly the semantic value of v – i.e., there is a semantic value for v on which φ is satisfied.

Truth   The notion of truth is then defined in terms of satisfaction:

Definition 64 (Truth). A sentence φ is *true* in $\mathscr{A}$ iff $|\varphi|_{\mathscr{A}}^{\alpha} = T$ for every variable assignment α over $\mathscr{A}$.

Notice that while open formulae can be always satisfiable, they cannot be true – that property is reserved for sentences. (If we had not restricted the notion, the open formula $Px \vee \neg Px$, which is assigned true under every variable assignment, would count as true – yet intuitively, at most it expresses a truth conditional on any constant is substituted for x, not unconditionally as a sentence does.)

7.3   Some Semantic Theorems About $\mathscr{L}_2$

Semantic Sequents for $\mathscr{L}_2$ We now extend the semantic turnstile $\vDash$ to $\mathscr{L}_2$. With the new definition of satisfaction and truth in a model in hand, we can define the old notions in much the same way:

1. If φ is true in all $\mathscr{L}_2$-structures, φ is a *tautology* or *logical* truth, written $\vDash \varphi$.

2. If φ is true in no $\mathscr{L}_2$-structure, φ is a *contradiction*, written $\varphi \vDash$.

3. A set Γ of $\mathscr{L}_2$ sentences is *semantically consistent* iff there is some $\mathscr{L}_2$-structure in which each $\gamma \in \Gamma$ is true; otherwise Γ is inconsistent, written $\Gamma \vDash$.

4. φ is *logically equivalent* to ψ iff both sentences are true in the same $\mathscr{L}_2$-structures, written $\varphi \equiv \psi$. (Some logicians use this symbol to

represent the object language biconditional connective, which we represent by ↔ – do not be confused.)

Definition 65 (Validity in $\mathscr{L}_2$). An argument from Γ to φ is *valid* (Γ entails φ) iff in every $\mathscr{L}_2$-structure in which each $\gamma \in \Gamma$ is true, φ is true also. We write again $\Gamma \vDash \varphi$.

Some Theorems About Entailment in $\mathscr{L}_2$ Many of the theorems about $\vDash$ we proved for $\mathscr{L}_1$ in Chapter 2 still hold, with exactly the same proofs.

Structural Rules All the structural rules (Permutation, Weakening, and Contraction) still hold.

Cut If $\Gamma, \varphi \vDash \psi$ and $\Gamma \vDash \varphi$, then $\Gamma \vDash \psi$.

Transitivity If $\Gamma \vDash \varphi$ and $\varphi \vDash \psi$, then $\Gamma \vDash \psi$.

Contraposition If $\varphi \vDash \psi$ iff $\neg\psi \vDash \neg\varphi$.

Deduction Theorem $\Gamma, \varphi \vDash \psi$ iff $\Gamma \vDash \varphi \to \psi$.

We now prove some theorems that depend on the new machinery we've introduced.

Quantifier Interdefinability

Theorem 66 (Interdefinable Quantifiers)

For all structures $\mathscr{A}$, and all variable assignments α,

- $|\forall v\varphi|_{\mathscr{A}}^{\alpha} = |\neg\exists v\neg\varphi|_{\mathscr{A}}^{\alpha}$;

- $|\exists v\varphi|_{\mathscr{A}}^{\alpha} = |\neg\forall v\neg\varphi|_{\mathscr{A}}^{\alpha}$.

Proof. I show only the first case. $|\forall v\varphi|_{\mathscr{A}}^{\alpha} = T$ iff for *all* v-variant assignments β, $|\varphi|_{\mathscr{A}}^{\beta} = T$; iff $|\neg\varphi|_{\mathscr{A}}^{\beta} = F$; iff $|\exists v\neg\varphi|_{\mathscr{A}}^{\alpha} = F$ (if not, there would be a β where $|\neg\varphi|_{\mathscr{A}}^{\beta} = T$); iff $|\neg\exists v\neg\varphi|_{\mathscr{A}}^{\alpha} = T$. □

Substitution of Co-Designating Terms

Theorem 67 (Substitution of Co-Designators)

Let τ_1 and τ_2 be terms, and let $\varphi[\tau_2/\tau_1]$ be the formula which results from substituting τ_2 for at least some free occurrences of τ_1 in φ, where none of those occurrences is in the scope of $\forall\tau_2$ or $\exists\tau_2$ (let constants vacuously occur free). Then, for any structure $\mathcal{A}$ and any variable assignment α: If $\left|\tau_1\right|_{\mathcal{A}}^{\alpha} = \left|\tau_2\right|_{\mathcal{A}}^{\alpha}$, then $\left|\varphi\right|_{\mathcal{A}}^{\alpha} = \left|\varphi[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha}$.

Proof. By induction on length of formulae.

Base case: Suppose φ is an atomic formula, $\Phi\theta_1 \ldots \theta_n$, where perhaps some $\theta_i = \tau_1$. But $\left|\Phi\theta_1 \ldots \theta_n\right|_{\mathcal{A}}^{\alpha} = T$ iff $\langle\left|\theta_1\right|_{\mathcal{A}}^{\alpha}, \ldots, \left|\tau_1\right|_{\mathcal{A}}^{\alpha}, \ldots, \left|\theta_n\right|_{\mathcal{A}}^{\alpha}\rangle \in \left|\Phi\right|_{\mathcal{A}}^{\alpha}$; since $\left|\tau_1\right|_{\mathcal{A}}^{\alpha} = \left|\tau_2\right|_{\mathcal{A}}^{\alpha}$, $\langle\left|\theta_1\right|_{\mathcal{A}}^{\alpha}, \ldots, \left|\tau_2\right|_{\mathcal{A}}^{\alpha}, \ldots, \left|\theta_n\right|_{\mathcal{A}}^{\alpha}\rangle \in \left|\Phi\right|_{\mathcal{A}}^{\alpha}$, i.e., $\left|\varphi[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha} = T$.

Induction step: Suppose φ is complex, and the theorem holds for ψ, χ less complex than φ and *each* variable assignment α over $\mathcal{A}$. We consider some representative cases, leaving others for exercises:

- For the first case, suppose $\varphi = \neg\psi$. It follows from the induction hypothesis that $\left|\neg\psi\right|_{\mathcal{A}}^{\alpha} = T$ iff $\left|\neg\psi[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha} = T$; by the clause on negation, that holds iff $\left|(\neg\psi)[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha} = T$, i.e., iff $\left|\varphi[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha} = T$.

- Suppose $\varphi = \psi \vee \chi$. $\left|\psi \vee \chi\right|_{\mathcal{A}}^{\alpha} = T$ iff either $\left|\psi\right|_{\mathcal{A}}^{\alpha} = T$ or $\left|\chi\right|_{\mathcal{A}}^{\alpha} = T$; by the induction hypothesis, iff $\left|\psi[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha} = T$ or $\left|\chi[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha} = T$; by substitution, iff $\left|(\psi \vee \chi)[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha} = T$, i.e., iff $\left|\varphi[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha} = T$.

- Suppose $\varphi = \forall v\psi$. $\left|\forall v\psi\right|_{\mathcal{A}}^{\alpha} = T$ iff for any variable assignment β which agrees with α except perhaps on v, $\left|\psi\right|_{\mathcal{A}}^{\beta} = T$. By the induction hypothesis, and the fact that it holds for all variable assignments, that holds iff $\left|\psi[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\beta} = T$ (note here we use the assumption that τ_1 doesn't occur in the scope of any quantifier that would bind τ_2). That holds iff $\left|\forall v\psi[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha} = T$ (again, it is required $v \neq \tau_2$), i.e., iff $\left|\varphi[\tau_2/\tau_1]\right|_{\mathcal{A}}^{\alpha} = T$. $\square$

Satisfaction of Sentences

Theorem 68 (Satisfaction of Sentences)

If φ is a $\mathcal{L}_2$-sentence, then for every structure $\mathcal{A}$ and all variable assignments $\alpha, \beta, |\varphi|_{\mathcal{A}}^{\alpha} = |\varphi|_{\mathcal{A}}^{\beta}$.

Proof. by induction on complexity. *Base case:* φ is an atomic formula. It is a sentence, so no variables occur in $\varphi = \Phi\tau_1 \ldots \tau_n$; since constants receive a constant interpretation in A, if φ is satisfied by any variable assignment it is satisfied by any other.

    Induction step: Suppose the theorem holds for less complex sentences ψ and all variable assignments over $\mathcal{A}$. The only interesting cases are the quantifiers, so let $\varphi = \forall v \chi$.

1. χ is a sentence (so the quantification is vacuous). Then for any α, β $|\chi|_{\mathcal{A}}^{\alpha} = |\chi|_{\mathcal{A}}^{\beta}$, so $|\forall v \chi|_{\mathcal{A}}^{\alpha} = |\forall v \chi|_{\mathcal{A}}^{\beta}$.

2. χ is not a sentence; since φ is, only v occurs free in χ. Suppose for some $\alpha, \beta, |\varphi|_{\mathcal{A}}^{\alpha} \neq |\varphi|_{\mathcal{A}}^{\beta}$. Without loss of generality, suppose $|\varphi|_{\mathcal{A}}^{\alpha} = T$. Then for every variable assignment γ differing from α at most on $v, |\chi|_{\mathcal{A}}^{\gamma} = T$.

   But for at least one variable assignment δ differing from β on at most $v, |\chi|_{\mathcal{A}}^{\delta} = F$.

   Let τ not occur in χ, and let $|\tau|_{\mathcal{A}}^{\delta} = |v|_{\mathcal{A}}^{\delta}$. Then by the Substitution of Terms Theorem, $|\chi|_{\mathcal{A}}^{\delta} = |\chi[\tau/v]|_{\mathcal{A}}^{\delta} = F$. By the induction hypothesis, for *every* γ, $|\chi[\tau/v]|_{\mathcal{A}}^{\gamma} = F$. In particular, for some γ^* which is like α except that it assigns $|\tau|_{\mathcal{A}}$ to be the extension of v, $|\chi[\tau/v]|_{\mathcal{A}}^{\gamma^*} = F$. But since $|\tau|_{\mathcal{A}}^{\gamma^*} = |v|_{\mathcal{A}}^{\gamma^*}, |\chi|_{\mathcal{A}}^{\gamma^*} = F$; contradicting our assumption that every v-variant of α satisfied χ. So there is no case where $|\varphi|_{\mathcal{A}}^{\alpha} \neq |\varphi|_{\mathcal{A}}^{\beta}$. $\square$

Quantifiers: Universal Elimination/Existential Introduction

Theorem 69

Suppose φ is a formula in which only v occurs free, and τ any constant. Then:

- $\forall v\varphi \vDash \varphi[\tau/v]$;

- $\varphi[\tau/v] \vDash \exists v\varphi$.

Proof. I only show the second. Suppose that $\mathscr{A}$ is a structure in which $\varphi[\tau/v]$ is true. Then in every variable assignment α according to which $|\tau|_{\mathscr{A}}^{\alpha} = |v|_{\mathscr{A}}^{\alpha}$, $|\varphi[\tau/v]|_{\mathscr{A}}^{\alpha} = |\varphi|_{\mathscr{A}}^{\alpha}$ (by the substitution of co-designating terms). There will be such a variable assignment α; so by the clause for $\exists$, $|\exists v\varphi|_{\mathscr{A}} = T$.

Quantifiers: Universal Introduction/Existential Elimination

Theorem 70

Suppose φ is a formula in which at most v occurs free, and τ any constant not occurring in Γ or φ. Then:

- *If $\Gamma \vDash \varphi[\tau/v]$, then $\Gamma \vDash \forall v\varphi$.*

- *If $\Gamma, \varphi[\tau/v] \vDash \psi$ then $\Gamma, \exists v\varphi \vDash \psi$, provided τ does not occur in ψ.*

Proof. Suppose that every structure $\mathscr{A}$ which makes Γ true makes $\varphi[\tau/v]$ true. τ does not occur in Γ, so every structure $\mathscr{A}'$ just like $\mathscr{A}$ except perhaps in the value of $I_{\mathscr{A}'}(\tau)$ also makes all of Γ true. Then $\varphi[\tau/v]$ is true in *every* such τ-variant structure $\mathscr{A}'$. Suppose β is a variable assignment where $|\varphi|_{\mathscr{A}}^{\beta} = F$. There is some $\mathscr{A}'$ where $I_{A'}(\tau) = |v|_{\mathscr{A}}^{\beta}$, so $|\varphi[\tau/v]|_{\mathscr{A}'} = F$. Contradiction; there is no such variable assignment, so $\forall v\varphi$ is true in every structure $\mathscr{A}$ which makes all of Γ true.

Substitution of Equivalent Formulae

Theorem 71 (Substitution of Equivalent Formulae)

Suppose φ and ψ are formulae, in which $v_1, \ldots, v_n$ occur free. Suppose δ is a some sentence in which φ occurs as a subformula. Then $\forall v_1 \ldots \forall v_n(\varphi \leftrightarrow \psi) \vDash \delta \leftrightarrow \delta[\psi/\varphi]$.

Proof. In every structure $\mathscr{A}$ in which $\forall v_1 \ldots \forall v_n(\varphi \leftrightarrow \psi)$ is true, $\varphi \leftrightarrow \psi$ is satisfied by every variable assignment agreeing on variables except possibly the v_is. But clearly if for every variable assignment α $|\varphi|_{\mathscr{A}}^{\alpha} = |\psi|_{\mathscr{A}}^{\alpha}$, a straightforward induction on complexity of δ will establish the result. $\square$

7.4 Alternative Semantics for $\mathscr{L}_2$

Semantics only for sentences   *Objection:* '*Why are open formulae being assigned truth values at all? They couldn't express any proposition, and we should only assign truth to formulae which could express propositions: sentences.*'

The easiest way to secure this result is as follows:

- Redefine formulae of $\mathscr{L}_2$ so that no open formula is a formula at all. We will thus alter the formation rules so that we do not add quantifiers to bind variables in open formulae, but rather to (i) *substitute* a variable ξ for one or more occurrences of some constant τ in φ; (ii) say that $\forall\xi\varphi[\xi/\tau]$ and $\exists\xi\varphi[\xi/\tau]$ are formulae.

- Replace the semantic clauses involving satisfaction with something like this:

    - $|\forall\xi\varphi|_{\mathscr{A}} = T$ iff for every individual $i \in D_{\mathscr{A}}$, and every structure $\mathscr{A}'$ just like $\mathscr{A}$ except perhaps in the semantic value of τ, if $|\tau|_{\mathscr{A}'} = i$, then $|\varphi[\tau/\xi]|_{\mathscr{A}'} = T$.

This alternative semantics gives a *recursion on truth* rather than satisfaction.

'Substitutional' Quantification   *Objection*: '*I like the idea that universal quantification is kind of like an infinite conjunction of all of its instances. Do we need to make the confusing detour through variable assignments?*'

Here is the idea behind this objection. What $\forall x Px$ really means is that whatever constant a one substitutes for x, Pa will be true. This is made more precise as follows: where φ is a formula with at most v free,

- $|\forall v \varphi|_{\mathscr{A}} = T$ iff for any constant τ, $|\varphi[\tau/v]|_{\mathscr{A}} = T$.

These alternative truth conditions quickly go awry if there are unnamed objects in the domain. For instance, in an uncountable domain (like the real numbers), as there are only countably many constants in $\mathscr{L}_2$, there are entities which cannot be the semantic value of any constant. Suppose we enumerate the constants a_i, letting $I_{\mathscr{A}}(a_i) = i$. Then on the domain of the real numbers, since every constant denotes an integer, $\forall x(\text{Integer}(x))$ looks like it will be true in $\mathscr{A}$ on the substitutional reading; yet this is highly counterintuitive.

Truth for Open Formulae   *Objection*: '*Wait! We do have the resources to explain the truth of an open formula: it is true whenever it is true under every variable assignment. So why be so restrictive?*'

On this proposal, $\vDash \varphi$ iff φ is true in every structure under every variable assignment. In effect, therefore, an open formula φ with free variables $v_1, \ldots, v_n$ will be a tautology iff its *universal closure* (the result of binding all free variables with universal quantifiers) $\forall v_1 \ldots \forall v_n \varphi$ is a tautology.

Yet this is problematic; for then this will be a correct sequent:

$$Px \vDash \forall x Px.$$

But the deduction theorem then fails! For

$$\vDash Px \rightarrow \forall x Px \qquad \text{iff} \qquad \vDash \forall x(Px \rightarrow \forall x Px),$$

and the right hand side sequent is clearly incorrect.

Further Reading

Tarski's original definition of satisfaction and truth can be found in Tarski (1933). (This is a translation of his 1936 paper.) Bostock (1997: §§3.4–3.6) gives a semantics, and proves many theorems, in which the basic semantic clauses involve recursion on truth rather than satisfaction. Kripke (1976) contains an interesting though technical discussion of the merits and problems with the substitutional intepretation of the quantifiers.

Exercises

1. Suppose we define a sentence φ to be *true** in $\mathscr{A}$ iff $|\varphi|_{\mathscr{A}}^{\alpha} = T$ for at least one variable assignment α over $\mathscr{A}$. Show that φ is true in $\mathscr{A}$ iff it is true* in $\mathscr{A}$.

2. Prove that, for all structures $\mathscr{A}$ and variable assignments α, $|\exists v \varphi|_{\mathscr{A}}^{\alpha} = |\neg \forall v \neg \varphi|_{\mathscr{A}}^{\alpha}$. ∎

3. Prove that if v is not free in φ, $|\forall v \varphi|_{\mathscr{A}}^{\alpha} = |\varphi|_{\mathscr{A}}^{\alpha}$. If φ is a sentence, show that φ and $\forall v \varphi$ are logically equivalent.

4. Prove that, if v and ν are variables, φ a formula not containing v, and $\varphi[v/\nu]$ the formula that results from replacing *every* occurence of ν in φ with v – *even those that occur in quantifiers* $\forall \nu, \exists \nu$ – then for any structure $\mathscr{A}$ and any variable assignment α, there exists a variable assignment β on $\mathscr{A}$ such that $|\varphi|_{\mathscr{A}}^{\alpha} = |\varphi[v/\nu]|_{\mathscr{A}}^{\beta}$.

5. Prove that, for any structure $\mathscr{A}$, if Φ and Ψ are n-ary predicate letters such that $|\Phi|_{\mathscr{A}} = |\Psi|_{\mathscr{A}}$, φ a formula, and $\varphi[\Psi/\Phi]$ the formula that results from replacing *every* occurrence of Φ in φ with Ψ, then for any variable assignment α over $\mathscr{A}$, $|\varphi|_{\mathscr{A}}^{\alpha} = |\varphi[\Psi/\Phi]|_{\mathscr{A}}^{\alpha}$. Then prove that, if φ is a sentence, φ is true in $\mathscr{A}$ iff $\varphi[\Psi/\Phi]$ is true in $\mathscr{A}$.

6. Suppose φ is a formula in which only v occurs free, and τ any constant. Prove that $\forall v \varphi \vDash \varphi[\tau/v]$.

7. Suppose φ is a formula in which at most v occurs free, and τ any constant *not occurring in* Γ *or* φ. Prove that if $\Gamma, \varphi[\tau/v] \vDash \psi$ then $\Gamma, \exists v \varphi \vDash \psi$, provided τ does not occur in ψ.

8. Show that, if τ and θ are any constants:

    (a) If $\Gamma \vDash \varphi$ then $\Gamma[\tau/\theta] \vDash \varphi[\tau/\theta]$.

    (b) If $\Gamma \vDash$ then $\Gamma[\tau/\theta] \vDash$.

9. Suppose φ and ψ are any formulae, in which possibly (but not necessarily) $v_1, \ldots, v_n$ occur free. Suppose δ is a some sentence in which φ occurs as a subformula. Using induction on complexity of δ, establish the result that $\forall v_1 \ldots \forall v_n(\varphi \leftrightarrow \psi) \vDash \delta \leftrightarrow \delta[\psi/\varphi]$. (Hint: you might want to consider an induction showing joint satisfiability of any formulae γ and $\gamma[\psi/\varphi]$.)

10. Show that $\forall \xi(\varphi \leftrightarrow \psi) \vDash \forall \xi \varphi \leftrightarrow \forall \xi \psi$.

11. Give a full specification of the alternative formation rules for a language $\mathscr{L}_2{}'$ in which there are no open formulae. Show that the sentences of $\mathscr{L}_2$ are the formulae of $\mathscr{L}_2{}'$.

12. Give a semantic clause for $\exists$ on the alternative semantics using *recursion on truth* as specified on page 110.

13. Show the following:

    (a) $\forall x(Px \wedge Qx) \equiv \forall x Px \wedge \forall x Qx$;

    (b) $\exists x(Px \wedge Qx) \vDash \exists x Px \wedge \exists x Qx$. Why not vice versa?

    (c) $\exists y \forall x Pxy \vDash \forall x \exists y Pxy$. Why not vice versa?

14. If $v_1 \ldots v_n$ is a sequence of variables, let $\forall v_1 \ldots v_n$ be the string of quantifiers $\forall v_1 \ldots \forall v_n$. Let $\mathscr{P}$ be an arbitrary permutation of the order of a sequence. Show that $\forall v_1 \ldots v_n \varphi \equiv \forall \mathscr{P} v_1 \ldots v_n \varphi$.

15. Explain why the 'objectual' reading of the quantifiers (i.e., the standard one) has no difficulty with universal quantification over an uncountable domain.

Chapter 8

Natural Deduction Proofs in $\mathscr{L}_2$; Soundness

8.1   Proofs in $\mathscr{L}_2$

Proofs in $\mathscr{L}_2$ In the case of $\mathscr{L}_1$, a proof procedure was not strictly speaking necessary. Every sentence of $\mathscr{L}_1$ has only finitely many sentence letters, so a truth table will need only to specify finitely many rows to capture what is true in every $\mathscr{L}_1$-structure.

But in $\mathscr{L}_2$ we don't have anything like a truth-table: one cannot simply survey a finite number of cases to represent all the structures. (Intuitively, this is because of the introduction of quantification – now a sentence may depend for its truth on the way infinitely many things are, so surveying only finitely many of them won't suffice.) Brute force checking of semantic sequents is not possible.

Thus the introduction of a proof technique, to enable the derivation of some $\mathscr{L}_2$ claims from others, is crucial. Once more, we use natural deduction; again, we abuse notation at let $\vdash$ stand for *provability in $\mathscr{L}_2$*.

$$\frac{\forall v\varphi}{\varphi[\tau/v]} \ \forall\text{Elim}$$

$$\begin{array}{c} \vdots \\ \varphi[\tau/v] \\ \hline \forall v\varphi \end{array} \ \forall\text{Intro}$$

Provided τ doesn't occur in φ or in any undischarged assumption in the proof of $\varphi[\tau/v]$.

$$\frac{\varphi[\tau/v]}{\exists v\varphi} \ \exists\text{Intro}$$

$$\begin{array}{c} [\varphi[\tau/v]] \\ \vdots \qquad \vdots \\ \psi \qquad \exists v\varphi \\ \hline \psi \end{array} \ \exists\text{Elim}$$

Provided τ doesn't occur in $\exists v\varphi$, in ψ, or in any undischarged assumption other than $\varphi[\tau/v]$ in the proof of ψ.

Table 8.1: New Natural Deduction Rules for $\mathscr{L}_2$

Natural Deduction   The rules for the natural deduction system ND_2 include all of those from the original system ND. As $\mathscr{L}_2$ includes all of $\mathscr{L}_1$, everything provable in ND is also provable in ND_2. But we include new rules to cover the quantifiers, laid out in Table 8.1. Throughout, let $\varphi[\tau/v]$ be the result of replacing *free* occurrences of variable v in φ by a constant τ.

Some Elementary Theorems About ND_2   Many of the theorems for ND continue to hold. For example, the proof of the *deduction theorem* in ND relied only on the arrow rules, so carry directly over to ND_2.

Theorem 72 (Cut)
If $\Gamma \vdash \varphi$ *and* $\varphi, \Delta \vdash \psi$ *then* $\Gamma, \Delta \vdash \psi$.

Proof. We cannot, as in ND, simply chain the proofs together, because Γ may contain a constant τ which, in the proof that $\varphi, \Delta \vdash \psi$, is used in an application of $\exists$Elim or $\forall$Intro (of course τ cannot appear in φ or Δ given the correctness of $\Delta, \varphi \vdash \psi$). So we need to show that there is *another* proof of $\Delta, \varphi \vdash \psi$ which doesn't involve any such τ. I leave this for an exercise.□

Proofs and Connectives   The natural deduction proofs of many sequents involving sentences with given connectives involve only the rules for those connectives. For example, the sequent $\forall x \exists y (Px \land Qx) \vdash \exists y \forall x (Px \land Qx)$ can be proved entirely using the rules for conjunction and the quantifiers.

But this is not true in general. Consider this: $\forall x (P \lor Qx) \vdash P \lor \forall x Qx$. (Proof below.)

This proof cannot be carried out using just the rules for disjunction and the quantifiers.

The situation is very similar to a case in an earlier problem (Chapter 4, 1.(e), page 97) where the natural deduction rules for $\rightarrow$ did not suffice to prove every tautology involving $\rightarrow$ as its only connective.

Proof of $\forall x (P \lor Qx) \vdash P \lor \forall x Qx$

$$
\begin{array}{c}
\cfrac{
\cfrac{
\cfrac{
\cfrac{\forall x(P \lor Qx)}{P \lor Qa}\,\forall\text{Elim} \quad [P] \quad \cfrac{[\neg(P \lor \forall xQx)] \quad \cfrac{[P]}{P \lor \forall xQx}\,\lor\text{Intro}}{\neg P}\,\neg\text{Elim}
}{Qa}\,\neg\text{Elim} \qquad\qquad [Qa]
}{\cfrac{\cfrac{Qa}{\forall xQx}\,\forall\text{Intro}}{P \lor \forall Qx}\,\lor\text{Intro}}\,\lor\text{Elim} \qquad\qquad\qquad [\neg(P \lor \forall xQx)]
}{P \lor \forall xQx}\,\neg\text{Elim}
\end{array}
$$

Some Proof Theory: Uniqueness of $\exists$ Some philosophers have wanted to maintain that there are at least two senses of the English word 'exists'. (They have wanted to do this typically to distinguish a committal from a non-committal sense, so they can accept the truth of claims like 'there are at lest two prime numbers less than 8' and 'there are at least two cities smaller than London', while being committed to cities, and avoiding com-mittment to such mysterious entities as numbers.) But these philosophers

have typically also desired to speak a language including both quantifiers; or at least to be able to translate between the two languages. There is however a fairly straightforward theorem which shows that this situation is not possible: if there are two quantifiers which both meet the minimal requirements to be an existential quantifier, those quantifiers are logically equivalent.

Theorem 73 (Harris 1982)

If there are two quantifiers $\exists_1$ and $\exists_2$ in a language, both governed by the standard introduction and elimination rules for the quantifiers, then $\exists_1 v\varphi$ is logically equivalent to $\exists_2 v\varphi$.

Proof. Assuming that we have introduction and elimination rules for the two quantifiers, we can give the following natural deduction proof:

$$\cfrac{\exists_1 v\varphi \qquad \cfrac{[\varphi[\tau/v]]}{\exists_2 v\varphi}\ \exists_2\text{Intro}}{\exists_2 v\varphi}\ \exists_1\text{Elim}$$

Obviously a precisely similar proof will show that $\exists_2 v\varphi \vdash \exists_1 v\varphi$. □

Unless, then, we wish to introduce two different sorts of names, two different sorts of predicates, etc., we cannot simply introduce two quantifiers into a language – not, at least, if they are to meet the minimal conditions that quantifiers should, namely, being characterised by our introduction and elimination rules.

8.2 Soundness of $\mathscr{L}_2$

Soundness   We are now in possession of a proof system ND_2, which characterises a provability turnstile $\vdash$. We now wish to show that this system is *sound* with respect to the semantics for $\mathscr{L}_2$ we introduced in the last chapter. That is, we wish to prove

Theorem 74 (Soundness)

If $\Gamma \vdash \varphi$ then $\Gamma \vDash \varphi$.

As ND_2 extends ND, and the rules of ND remain sound (verification of this is left to an exercise), we just need to show that the new rules – the quantifier rules – are sound. We also still know that the base case $\varphi \vdash \varphi$ is sound. So we just consider proofs extended by the quantifier rules.

Soundness of $\mathscr{L}_2$: ∀Elim and ∃Intro Suppose $\Gamma \vdash \varphi$, where φ is obtained from an earlier proof by the use of the ∀Elim rule, and so is of the form $\psi[\tau/v]$. Then there was a proof on the assumption Γ of $\forall v\psi$.

By the induction hypothesis, $\Gamma \vDash \forall v\psi$. So every structure which makes all the members of Γ true, makes $\forall v\psi$ true. By theorem 69 (proved on page 109), $\forall v\psi \vDash \psi[\tau/v]$. Since $\psi[\tau/v] = \varphi$, φ will be true in every structure which makes Γ true, i.e., $\Gamma \vDash \varphi$.

The proof for ∃Intro is similar. I leave it for an exercise.

Soundness of $\mathscr{L}_2$: ∀Intro and ∃Elim The trickier case is if $\Gamma \vdash \varphi$ which was obtained from ∃Elim. From a proof of $\exists v\psi$ on assumptions Δ, and a proof of φ on assumptions $\Theta, \psi[\tau/v]$, we obtain a proof of φ on assumption $\Gamma = \Delta \cup \Theta$ (with appropriate restriction on where τ occurs). Since $\Delta \subseteq \Gamma$, we have $\Gamma \vdash \exists v\psi$, and by the induction hypothesis, $\Gamma \vDash \exists v\psi$.

Since $\Theta \subseteq \Gamma$, we have $\Gamma, \psi[\tau/v] \vdash \varphi$, and by the induction hypothesis $\Gamma, \psi[\tau/v] \vDash \varphi$. By theorem 70 (proved on page 109), $\Gamma, \exists v\psi \vDash \varphi$. And by the Cut theorem, $\Gamma \vDash \varphi$.

The soundness of ∀Intro is similarly demonstrated; I leave it for an exercise.

8.3   Completeness of $\mathscr{L}_2$

Completeness

Theorem 75 (Completeness of $\mathscr{L}_2$)

If $\Gamma \vDash \varphi$, $\Gamma \vdash \varphi$.

The proof extends the proof of completeness for $\mathscr{L}_1$, roughly like this:

- First, show that, e.g., if Γ is a maximal consistent set, then $\exists v\varphi \in \Gamma$ iff $\varphi[\tau/v] \in \Gamma$ for some constant τ. (This is analogous to the other closure properties we showed.)

- Second, show that every maximal consistent set of $\mathscr{L}_2$-sentences is satisfiable.

- Finally, show that if $\Gamma \nvdash \varphi$, there is a maximal consistent set Γ^+ such that $\Gamma \cup \{\neg\varphi\} \subseteq \Gamma^+$ – since it is maximal, Γ^+ is satisfiable, so there is a model where all of Γ is true and φ is false, so $\Gamma \nvDash \varphi$. By contraposition, this shows completeness.

We won't prove this theorem here; the details are involved.

Compactness

Theorem 76 (Compactness)

If Γ is any set of $\mathscr{L}_2$ sentences, then Γ is satisfiable iff every finite subset of Γ is satisfiable.

Again, compactness is a consequence of the finitude of proofs and the Completeness Theorem. A direct proof, like that we gave for $\mathscr{L}_1$ in Chapter 3, is more difficult in $\mathscr{L}_2$, and again beyond the scope of this course.

Compactness has the consequence that certain relations are undefinable. In English, the infinitely many sentences of the form 'x is *not* a predecessor of…a predecessor of y' entail 'x is not less than y', but no finite subset does – but their translation into the compact language $\mathscr{L}_2$ must therefore fail. This is because the relation between predecessor and less than is not definable in $\mathscr{L}_2$.

8.4   Decidability and Undecidability

Completeness and Decidability
Theorem 77 ($\mathscr{L}_2$ is positively decidable)
If $\vDash \varphi$, then there is an effective procedure demonstrating that.

Proof. Completeness shows that if φ is a theorem, there is a proof of it. By brute force, we can generate finite proofs and spit out a proof of φ if one is to be had. Since the set of sentences of $\mathscr{L}_2$ is countable, and since every proof is of finite length, consider the proofs of length n which involve only the first m $\mathscr{L}_2$ sentences. For each pair $\langle m, n \rangle$ there are only finitely many such proofs; we can effectively produce them. Since the set of pairs of integers are countable (exercise), there is an enumeration of those pairs, hence we can effectively produce any finite proof. But every theorem φ is proved by some such finite proof; so at some finite point in the enumeration, we will reach an m greater than the numbers of all the members of a proof of φ, and at some finite point spit out a proof of φ of length less than n. □

Undecidability
Theorem 78
$\mathscr{L}_2$ is negatively undecidable: *if $\varphi \nvDash$, no effective procedure exists which will show that for any φ in a finite time.*

Considering the effective procedure for positive decidability, we can see why this won't stop after a finite time (the proof of $\neg \varphi$ might always be achieved using the next pair to be considered); but maybe there is another method that will terminate?

There is not. One way of helping to make this thought persuasive is to consider a particular incorrect sequent: $\forall x \exists y Pxy \vDash Paa$. One can certainly give a counterexample to this (let the domain be the natural numbers, and 'P' be interpreted as '$<$'). But is there an effective procedure for producing such counterexamples?

The Halting Problem   If S is a finite set of instructions, can we give an effective procedure that determines whether it is the case that those instructions, carried out on some acceptable input, will *terminate* after a finite time with some output?

Consider some enumeration of the set of finite sets of instructions; the halting problem, if soluble, will entail the existence of an effectively computable function $h(i)$ that yields 1 if the i-th set of instructions halts with a defined output on input i, and 0 otherwise (i.e., $h(i) = 0$ if the i-th set of instructions determines a function which is undefined on input i). Can there be such a function h?

Sketch of Insolubility of the Halting Problem   If there were such a function h, we could effectively compute the function $g(i) = 0$ if the i-th set of instructions is undefined on input i, and undefined otherwise. We could define g by adding to the instructions that compute h the last instruction: if $h(i)$ outputs 1, then check if $h(i)$ outputs 1. This last instruction will send give an infinite loop if $h(i)$ is equal to 1, rendering $g(i)$ undefined; but if $h(i) = 0$, then $g(i) = 0$ too.

But g, if effectively computable, will be specified by a finite set of instructions; let it be j-th in our list. Then $g(j) = 0$ iff $g(j)$ is undefined, so it is undefined; but then $g(j)$ must be defined. Contradiction; so there is no such g, and therefore no such h. The *self-halting problem* is insoluble, and therefore so is the halting problem.

For a more humorous proof sketch, see Pullum (2008).

Undecidability and the Halting Problem   It turns out that the undecidability of $\mathscr{L}_2$ is closely related to the Halting problem. For there is a way of associating to each set of instructions i an *argument* in $\mathscr{L}_2$ such that the i-th set of instructions halts on input i iff the argument is valid. So if there was an effective test for invalidity of an argument in $\mathscr{L}_2$, the halting problem would be solvable. Since it is insoluble, we have

Theorem 79 (Undecidability of $\mathscr{L}_2$)
$\mathscr{L}_2$ *is not decidable.*

The translation essentially involves *binary* predicates (Jeffrey §8.7) and sentences in which (sometimes) a universal quantifier is essentially in the scope of an existential quantifier. So we might wonder: are there fragments of $\mathscr{L}_2$ that are decidable?

Prenex Normal Form   A sentence φ of $\mathscr{L}_2$ is in *prenex normal form* iff no quantifier in φ occurs in the scope of any truth-functor (all quantifiers are in a block).

Theorem 80 (PNF)
Every sentence of $\mathscr{L}_2$ is equivalent to a sentence in prenex normal form.

Proof. The proof will be work by showing that if φ is some sentence, any quantifier which occurs somewhere in the middle of φ can be 'moved' to the left. The idea is to show these equivalences:

1. $\neg \forall v \varphi \equiv \exists v \neg \varphi$; (basically shown by Theorem 66)

2. $\neg \exists v \varphi \equiv \forall v \neg \varphi$; (basically shown by Theorem 66)

3. $\psi \wedge \forall v \varphi \equiv \forall v (\psi \wedge \varphi)$ (v not free in ψ).

4. $\psi \wedge \exists v \varphi \equiv \exists v (\psi \wedge \varphi)$ (v not free in ψ).

Etc. The full proof is left for an exercise. $\qquad\qquad\square$

Decidability of $\forall \exists$-sentences   An $\forall \exists$-*sentence* is a sentence in PNF in which all universal quantifiers precede all existential quantifiers.

Theorem 81 (Decidability of $\forall \exists$-sentences)
There exists an effective procedure which establishes the validity of any valid $\forall \exists$-sentence, and the invalidity of any invalid $\forall \exists$-sentence.

The proof is as follows. We show how to reduce the question of validity of a ∀∃-sentence to the question of validity of a related quantifier-free sentence; and such quantifier-free sentences of $\mathscr{L}_2$ are decidable. For convenience, I abbreviate $\forall x_1 \dots \forall x_n$ as $\forall x_1 \dots x_n$, and similarly for ∃.

Lemma on Quantifier-Free Sentences

Lemma 82 (Quantifier-Free Sentences)

If φ is a sentence of $\mathscr{L}_2$, which is not a (substitution instance of a) truth-functional tautology, then there is a structure $\mathscr{A}$ in which $|\varphi|_{\mathscr{A}} = F$ and either (i) if φ contains no constants, the domain of $\mathscr{A}$ has one element; or (ii) otherwise, $\mathscr{A}$ has an element in its domain for each distinct constant occurring in φ and no other elements, and each constant in φ is assigned a distinct element of the domain.

Proof. Hint: if φ is not a truth-functional tautology, there is an assignment of truth-values to atomic sentences in φ that mimics an $\mathscr{L}_1$-structure; and we can construct an $\mathscr{L}_2$-structure $\mathscr{A}$ which agrees with that pseudo-$\mathscr{L}_1$-structure on its assignment to atomic sentences. □

A corollary is that quantifier-free sentences of $\mathscr{L}_2$ are decidable, since the truth-table test decides truth-functional tautologies.

∃-sentences

Theorem 83 (Constant-free ∃-sentences)

If φ is a quantifier-free formula which contains only the variables $v_1, \dots, v_n$ and which contains no constants, then $\models \exists v_1 \dots v_n \varphi$ iff $\models \varphi[\tau/v_1, \dots, \tau/v_n]$.

Proof. R to L: Repeated applications of Existential Introduction (Theorem 69) yield $\varphi[\tau/v_1, \dots, \tau/v_n] \models \exists v_1 \dots v_n \varphi$. By Cut, the result follows.

L to R: If $\nvDash \varphi[\tau/v_1, \dots, \tau/v_n]$, then by Lemma 82, there is a structure with a one-element domain $\mathscr{A}$ where it is false. If $\left|\exists v_1, \dots v_n \varphi\right|_{\mathscr{A}} = T$, there is a variable assignment α such that $|\varphi|_{\mathscr{A}}^{\alpha} = T$; but this variable assignment

must assign $|\tau|_{\mathscr{A}}^{\alpha}$ to each variable (there is only one element in the domain), so $\left|\varphi[\tau/v_1, \dots, \tau/v_n]\right|_{\mathscr{A}}^{\alpha} = T$, contradiction. (Using Theorem 68.) $\qquad\square$

These $\exists$-sentences are *equi-decidable* with quantifier-free sentences.

Theorem 84 (Arbitrary $\exists$-sentences)
If φ is a quantifier free formula containing only $\tau_1, \dots, \tau_n$ and $v_1, \dots, v_m$, then

$$\vDash \exists v_1 \dots v_m \varphi \quad \text{iff} \quad \vDash \bigvee (\varphi[\tau_i/v_j]),$$

where $\bigvee(\varphi[\tau_i/v_j])$ is the disjunction of all the ways of substituting the constants $\tau_1, \dots, \tau_n$ for the variables $v_1, \dots, v_m$.

I leave the proof of this for an exercise; it is along the same lines as the proof of the proof for constant-free $\exists$-sentences.

Now we have shown that an arbitrary sentence involving only existential quantification is valid iff some quantifier-free sentence is valid, and since the latter is decidable, so is the former. Now we just need to show that any $\forall\exists$-sentence is valid iff some $\exists$-sentence is valid.

$\forall\exists$-Sentences
Theorem 85 ($\forall\exists$-sentences and $\exists$-sentences)
If $\tau_1, \dots, \tau_m$ don't occur in φ, then

$$\vDash \forall v_1 \dots v_m \exists v_1 \dots v_n \varphi \quad \text{iff} \quad \vDash \exists v_1 \dots v_n \varphi[\tau_1/v_1, \dots, \tau_m/v_m].$$

Proof. L to R: By repeated applications of $\forall$-Elimination (Theorem 69), we have

$$\forall v_1 \dots v_m \exists v_1 \dots v_n \varphi \vDash \exists v_1 \dots v_n \varphi[\tau_1/v_1, \dots, \tau_m/v_m],$$

from which the result follows by Cut.

R to L: Because Γ is empty, and $\tau_1, \dots, \tau_m$ don't occur in φ, we can use $\forall$-Introduction (Theorem 70) repeatedly to show the theorem. $\square$

So any $\forall\exists$-sentence is equi-valid with some quantifier-free sentence.

Elementary Quantifications and Monadic Sentences   A sentence is an *elementary quantification* iff no quantifier occurs in the scope of any other. Some sentences won't be equivalent to any elementarily quantified sentence – e.g., $\forall x\exists y Rxy$ can't be put into that form.

An $\mathscr{L}_2$ sentence is *monadic* iff it contains at most unary (one-place) predicates.

Lemma 86 (Monadic and Elementary Sentences)
Every monadic sentence of $\mathscr{L}_2$ is equivalent to some elementary quantification.

Proof. Left as a problem. The general idea is: since we have no binary predicates, each bound variable in a sentence is associated with a unique quantifier, and we can 'drive in' the quantifiers so that it binds only its own variable. $\square$

Corollary: Monadic $\mathscr{L}_2$ decidable

Theorem 87 (Monadic $\mathscr{L}_2$ sentences)
If φ is a monadic sentence of $\mathscr{L}_2$, it is equivalent to a $\forall\exists$-sentence.

Proof. Suppose φ is a monadic sentence. We can effectively produce a an elementary quantification φ', logically equivalent to φ. But switching the order of any quantifiers in that driven in sentence can't change the truth-value (exercise). And so we can reorder the quantifiers in φ' to get another logically equivalent sentence φ'' that is $\forall\exists$. Each step is effective: monadic $\mathscr{L}_2$ is decidable. $\square$

Further Reading

Further discussion of the Harris Theorem is in McGee (2006).

The completeness proof was first established by Gödel in his 1929 doctoral thesis; the proof sketched here follows the later proof of Henkin (1949). The technique of associating (in)valid arguments with (un)defined computable functions mentioned in the proof of undecidability is involved. An elementary version is in Jeffrey (2006: ch. 7–8). A more sophisticated version is in Boolos *et al.* (2007: ch. 1–8). The connection between decidability and the halting problem was shown by Church and Turing, independently: Turing's result is in Turing (1937). This technique for proving decidability of ∀∃-sentences is in Bostock (1997: §3.9). He provides an alternative decision procedure for monadic $\mathcal{L}_2$ in §3.8.

Exercises

1. Prove, using the hints in the chapter, the Cut Theorem: that if $\Gamma \vdash \varphi$ and $\Delta, \varphi \vdash \psi$, then $\Gamma, \Delta \vdash \psi$.

2. Explain why we could have used this rule instead of our ∀Intro:

$$\frac{\overset{\vdots}{\varphi}}{\forall v \varphi[v/\tau]} \text{ ∀Intro}$$

   Provided τ doesn't occur in any undischarged assumption in the proof of φ.

3. Let ψ be the result of substituting v for *all* occurrences of τ in φ. Show that this rule is not sound (i.e., not truth-preserving)

   If $\Gamma, \varphi \vdash \chi$ then $\Gamma, \exists v \psi \vdash \chi$, provided τ does not occur in Γ or in χ.

   Can you think of a way of amending the rule to make it sound? *(Hint: Theorem 7.3.)*

4. Verify that all the rules of *N D* are sound with respect to the $\mathcal{L}_2$ semantics.

5. (a) Show that the rule ∃Intro is sound with respect to the $\mathcal{L}_2$ semantics.

   (b) Show that the rule ∀Intro is sound with respect to the $\mathcal{L}_2$ semantics.

6. Explain why one cannot define the relation 'x is an ancestor of y' in terms of the relation 'x is a parent of y'. What resources could be added to the language to allow this to be defined?

7. Sketch the proof that for any finite set of $\mathscr{L}_2$ sentences, we can effectively produce every proof that uses only sentences in that set.

8. Assuming that $\wedge, \neg, \vee$ are the only truth functional connectives in φ:

    (a) Show that, where v is not free in ψ, that $\psi \wedge \forall v \varphi \equiv \forall v(\psi \wedge \varphi)$.

    (b) Show that, where v is not free in ψ, that $\psi \wedge \exists v \varphi \equiv \exists v(\psi \wedge \varphi)$.

    (c) Show that, where v is not free in ψ, that $\psi \vee \forall v \varphi \equiv \forall v(\psi \vee \varphi)$.

    (d) Show that, where v is not free in ψ, that $\psi \vee \exists v \varphi \equiv \exists v(\psi \vee \varphi)$.

    (e) Show that $\forall v \varphi \wedge \forall v \psi \equiv \forall v(\varphi \wedge \psi)$.

    (f) Show that $\exists v \varphi \vee \exists v \psi \equiv \exists v(\varphi \vee \psi)$.

9. (a) Using the results in problem 8, plus the results on the interdefinability of quantifiers from week 5, prove the PNF theorem for the fragment of $\mathscr{L}_2$ without arrow or biconditional.

    (b) Prove the PNF in the following stronger form:

    > For every sentence φ in the arrow and biconditional-free fragment of $\mathscr{L}_2$, there is a logically equivalent sentence φ' in PNF *that has no more connectives than φ*.

    (c) Can we extend the stronger result to the full language $\mathscr{L}_2$?

10. Prove the Lemma on Quantifier-Free Sentences.

11. Prove the Theorem for Arbitrary $\exists$-Sentences.

12. Suppose that φ is a monadic $\mathscr{L}_2$ sentence containing only $\wedge, \vee, \neg$ as connectives in addition to quantifiers.

    (a) Show that we can construct a sentence φ' logically equivalent to φ in which no quantifier occurs in the scope of any other – i.e., an elementary quantification. (Consider exercises 8.(a–f), and recalling facts about DNF and CNF, thinking particularly about giving a CNF 'equivalent' of an open formula in DNF and *vice versa*.)

    (b) Show that $\forall v \varphi \wedge \forall v \psi \equiv \forall v \varphi \wedge \forall v \psi[v/v]$ (and similarly for disjunction, and other combinations of universal and existential quantifiers).

(c) Show that for any elementary quantification φ' we can construct a logically equivalent sentence φ'' in which any quantifier in φ' is brought to the front and has scope over the others. (Tip: remember to use the result proved in 12.(b).)

(d) Show, using these results, that there is a $\forall\exists$ sentence equivalent to any such monadic φ.

13. Prove that there is no sentence φ which contains just one occurrence of some two-place connective, just one occurrence of any quantifier, which is in PNF and is equivalent to $P \leftrightarrow \forall x Q x$. Is there any connective other than $\leftrightarrow$ for which this also holds? Why?

14. The *Cantor Pairing Function* is this function from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$:

$$\pi(\langle x, y \rangle) = \frac{(x + y)(x + y + 1)}{2} + y.$$

Show that this function has an inverse (i.e., a function from natural numbers to pairs of natural numbers). Show therefore that π is one-one and onto, and that the set of pairs of natural numbers is countable. (Hint: show that, given some z, we can reconstruct a unique x and y from it such that $\pi(\langle x, y \rangle) = z$.)

Chapter 9

Syntax, Semantics, and Proofs in $\mathscr{L}_=$

9.1 Identity

Identity: Syntax and Semantics   The language $\mathscr{L}_2$ doesn't have a privileged predicate for identity. Yet such a predicate is very useful in discussing binary relations.

We *add* to $\mathscr{L}_2$ a binary predicate '='. We add new atomic formulae: where τ_1 and τ_2 are any terms, '$\tau_1 = \tau_2$' is an atomic formula. We keep the same structures as $\mathscr{L}_2$, but give this new predicate a constant semantic value, so that in every $\mathscr{L}_2$-structure $\mathscr{A}$,

(=) $I_\mathscr{A}(=) = \{\langle x, x \rangle : x \in D_\mathscr{A}\}$

Of course the property that we intend this predicate to express could exist in an $\mathscr{L}_2$-structure already; the change is that we now give it a logically privileged meaning. The satisfaction conditions are:

- $\left|\tau_1 = \tau_2\right|^\alpha_\mathscr{A} = T$ iff $\left\langle \left|\tau_1\right|^\alpha_\mathscr{A}, \left|\tau_2\right|^\alpha_\mathscr{A} \right\rangle \in \left|=\right|^\alpha_\mathscr{A}$, i.e., iff $\left|\tau_1\right|^\alpha_\mathscr{A} = \left|\tau_2\right|^\alpha_\mathscr{A}$; i.e., if $\left|\tau\right|^\alpha_\mathscr{A}$ is the very same object as $\left|\kappa\right|^\alpha_\mathscr{A}$.

Call the language which keeps all the old atomic formulae and forma-
tion rules of $\mathscr{L}_2$, but adds these atomic formulae, $\mathscr{L}_=$.

Some Theorems About Identity

Theorem 88 (Theoremhood of Identity)

For any constant τ, $\vDash \tau = \tau$.

Proof. Suppose $\nvDash \tau = \tau$ for some τ. Then there is a structure $\mathscr{A}$ such that
$|\tau = \tau|_{\mathscr{A}} = F$, iff $|\tau|_{\mathscr{A}}^{\alpha} \neq |\tau|_{\mathscr{A}}^{\alpha}$; impossible. $\qquad\qquad\square$

Theorem 89 (Substitution of co-designating terms II)

$\tau = \kappa \vDash \varphi[\tau/v] \leftrightarrow \varphi[\kappa/v]$.

Proof. Fairly immediate from the Substitution of Co-Designating Terms
theorem (chapter 5, page 107).

The Relation of Identity

Theorem 90 (Identity an Equivalence Relation)

- $\vDash \forall x\, x = x;$

- $\vDash \forall x \forall y (x = y \rightarrow y = x);$

- $\vDash \forall x \forall y \forall z ((x = y \land y = z) \rightarrow x = z).$

Theorem 91 ('Leibniz' Law')

For any φ in which at most v occurs free, $\vDash \forall v \forall v (v = v \rightarrow \varphi \leftrightarrow \varphi[v/v])$.

Proofs left for exercises. *Leibniz' Law* usually refers to this *definition* of
identity in second order logic: $\forall v \forall v (v = v \leftrightarrow \forall \Phi(\Phi(v) \leftrightarrow \Phi[v/v](v)))$.

Proofs in $\mathscr{L}_=$ To obtain a natural deduction system for $\mathscr{L}_=$, $ND_=$, we *add*
to the rules of ND_2 the rules in Table 9.1, keeping all other rules. In =Elim-l
and =Elim-r, φ is a formula in which at most v occurs free.

$$\frac{[\tau = \tau]}{\vdots} \text{=Intro}$$

$$\frac{\varphi[\tau/v] \quad \tau = \kappa}{\varphi[\kappa/v]} \text{=Elim-r} \qquad \frac{\varphi[\tau/v] \quad \kappa = \tau}{\varphi[\kappa/v]} \text{=Elim-l}$$

Table 9.1: New Natural Deduction Rules for $\mathscr{L}_=$

Dispensibility of =Elim-l/=Elim-r The two rules =Elim-l and =Elim-r obviously offer very similar resources in proofs. If we could show that $\tau = \kappa \vdash \kappa = \tau$ using just *one* of these rules, we could show the other to be dispensible.

Theorem 92 (Dispensibility of =Elim-r (or =Elim-l))
In the presence of the rule =Elim-l (respectively, =Elim-r), the functionality of =Elim-r (=Elim-l) can be derived.

Proof. $\dfrac{\varphi[\tau/v] \qquad \dfrac{[\kappa = \kappa] \quad \tau = \kappa}{\kappa = \tau} \text{=Elim-l.}}{\varphi[\kappa/v]} \text{=Elim-l}$

This is obviously equivalent to =Elim-r; a similar proof would show the dispensibility of =Elim-l. (Note the use of =Intro.) □

Soundness of $\mathscr{L}_=$

Theorem 93 ($\mathscr{L}_=$ is Sound)
If φ is a sentence of $\mathscr{L}_=$, then if $\vdash_{\mathscr{L}_=} \varphi$ then $\vDash \varphi$.

Proof. We only need show that proofs extended by the new rules are sound.

- We have a proof of φ on assumption $\tau = \tau$, and apply =Intro to discharge the assumption. Since by the Theoremhood of Identity, $|\tau = \tau|_{\mathscr{A}} = T$ in every structure, and by the induction hypothesis $|\varphi|_{\mathscr{A}} = T$ if $|\tau = \tau|_{\mathscr{A}} = T$, clearly $|\varphi|_{\mathscr{A}} = T$.

- We have a proof of $\tau = \kappa$ on assumptions Γ, and a proof of $\varphi[\tau/\upsilon]$ on assumptions Δ, and we extend to a proof of $\varphi[\kappa/\upsilon]$ using =Elim-r on assumptions $\Gamma \cup \Delta$. By the induction hypothesis, for any $\mathscr{A}$ where $\Gamma \cup \Delta$ is satisfied, $|\tau = \kappa|_{\mathscr{A}} = T$, and similarly $|\varphi[\tau/\upsilon]|_{\mathscr{A}} = T$. By the Substitution of co-designating constants theorem of page 130 and logic, in any such $\mathscr{A}$, $|\varphi[\kappa/\upsilon]|_{\mathscr{A}} = T$. $\qquad\square$

Further metalogical results   It turns out that the proof of completeness of $\mathscr{L}_2$ can be extended to provide a completeness proof for $\mathscr{L}_=$.

Theorem 94 (Completeness of $\mathscr{L}_=$)

If $\vDash \varphi$ then $\vdash \varphi$.

In the now familiar way, this can be adapted to a compactness proof.

Since $\mathscr{L}_2$ is not decidable, and $\mathscr{L}_=$ contains $\mathscr{L}_2$ as a part, $\mathscr{L}_=$ is not decidable either. Obviously the monadic fragment of $\mathscr{L}_=$ is decidable (as '=' is binary, the monadic fragment is the same as that of $\mathscr{L}_2$). A more interesting fragment is *the pure theory of identity* – the fragment where the only atomic formulae are of the form $\tau = \kappa$. This, it turns out, is decidable (Bostock, 1997: 329–31).

9.2   Numerical Quantification and the Theory of Definite Descriptions

Numerical Quantification: 'At Least'   The ordinary interpretation of the sentence $\exists x P x$ is that there are one or more things which satisfy P. Obviously, $\exists x \exists y P x \wedge P y$ does not express that there are two or more things which satisfy P, because x and y may take the same values under a variable assignment. To express that there are at least two Ps, we need to guarantee the *distinctness* of the values of the variables. We can do this using identity:

$$\exists x \exists y ((P x \wedge P y) \wedge \neg x = y).$$

Similarly,

$$\exists x \exists y \exists z((Px \wedge Py \wedge Pz) \wedge (\neg x = y \wedge \neg y = z \wedge \neg x = z))$$

expresses that there are *at least three* distinct Ps. Let $\exists_n \varphi$ express that there are at least n things which satisfy φ.

Numerical Quantification: 'At Most', 'Exactly'   Just as for 'at least', so we can formalise 'there are at most n Ps'. Under what circumstances are there at most n things? Just in case, if we chose $n + 1$ times, we would have to have chosen some of the things twice: some of things chosen would have to be identical. So we formalise 'there are at most 2 Ps' as

$$\forall x \forall y \forall z((Px \wedge Py \wedge Pz) \rightarrow (x = y \vee x = z \vee y = z)).$$

Formalise 'there are at most n φs' as $\forall_n x \varphi$. Then we can say that *there are exactly n φ* as

$$\exists_n x \varphi \wedge \forall_n x \varphi.$$

There will be more or less concise ways of expressing this claim.

Definite Descriptions   The $\mathscr{L}_=$ sentence Pa is often a good translation of simple English sentences of subject-predicate form. Yet some sentences, with this apparent form, cannot be translated in this way. Consider 'The fourth-oldest college is a college'. While 'the fourth-oldest college' is a referring expression, denoting a particular item in the domain, it seems a bad idea to formalise this by a constant a. For one thing, the English sentence is a tautology, but the proposed formalisation is not. So some English expressions which, like constants, denote an object in the domain, shouldn't be formalised by constants. It is notable that 'the fourth-oldest college' doesn't

name a particular college (Exeter) – it *describes* it. A *definite description* designates a particular thing by giving a description that the thing – and *only* that thing – satisfies. We will look again at the relation between constants of $\mathscr{L}_2/\mathscr{L}_=$ and English referring expressions in Chapter 8.

So 'the strongest man in the world' designates that person who satisfies the property of being a man stronger than any other. 'Strongest' entails uniqueness, and this is generally true for definite descriptions: they fail to refer if more than one thing satisfies the description (so 'the college on Turl St' fails, because there are three such). *Indefinite* descriptions, such as 'a strong man', do not require uniqueness.

Definite descriptions also seem to require existence of the thing they describe: the definite description 'the present king of France' fails because *nothing* satisfies that description.

Definite Descriptions and Numerical Quantification   Putting these ideas together, we might say that the definite description 'the P' refers to the unique existing thing that is P. But we know how to express that there is one and only one P, using numerical quantification – so we can say that 'the P is Q', because we can say that there is one and only one P, and it is Q. Where φ, ψ are expressions in which at most v occurs free, let us introduce the notation $\imath v : \varphi$ to denote 'the v such that φ', or simply, 'the φ'. One could permit this expression can take the place of a constant, so if $\psi\tau$ is well-formed, so is $\psi(\imath v : \varphi)$; the latter says 'the φ is ψ'. Given our resources, though, we needn't add this to the language, or deal with the complications that ensue, for this expression can be defined:

Definition 66 (Definite Descriptions). $\psi(\imath v : \varphi) =_{\mathrm{df}} \exists v(\psi \wedge \forall v \varphi[v/v] \leftrightarrow v = v)$.

This, essentially, is *Russell's analysis of the logical form of definite descriptions*.

Descriptions and Scope   One reason to suspect that definite descriptions aren't just ordinary names – and a corresponding reason to reject the descriptivist account of the meaning of the constants of natural language – is that descriptions, unlike names, have *scope*. Consider

(*) The prime minister has always been Australian.

In $\mathscr{L}_=$, there is no non-truth-functional operator 'it always has been that φ'. We can mimic this (or maybe it is not mimicking) by *quantifying over times*. Let t be 'now', $\prec$ be 'is earlier than', P be 'is prime minister at', and A be 'is Australian'. (*) has two readings:

$$\forall x(x \prec t \rightarrow \exists y((Pyx \wedge \forall z(Pzx \leftrightarrow y = z))) \wedge Ay);$$
$$\exists y((Pyt \wedge \forall z(Pzt \leftrightarrow y = z)) \wedge \forall x(x \prec t \rightarrow Ay)).$$

The first says: always the prime minster, whoever they've been, has been Australian at the time of their prime ministership; the second says the present prime minister has always been Australian.

  (Very similar results can be seen in *modal* contexts of possibility and necessity. We can again mimic the behaviour of the operator 'possibly' as an existential quantifier over 'possible worlds', and 'necessarily' (like 'always') as a universal quantifier.)

  The problem is that in the case of names we seem *not* to see these scope problems. For this claim has only one reading:

(†) Kevin Rudd has always been Australian.

The existence of scopes for definite descriptions, is some evidence for the fact that they are really quantifier expressions.

  Yet there are some disanalogies. Consider

(‡) The present king of France is bald.

According to Russell's theory, this is *false* – existence fails. But we may not share that intuition; this claim may strike us as neither true nor false, or having a false presupposition and thus unassessable.

Alternative Theories of Descriptions   Strawson used our uncertain intuitions about (‡) to argue that descriptions are often used referentially,

> to mention or refer to some individual person or single object…, in the course of doing what we should normally describe as making a statement about that person [or] object.

When uttering a referring expression, one *presupposes* the existence of the thing. But when our intentions go awry, and we have presupposed something false, the utterance fails to have a truth value.

One truly weird way of dealing with (‡) is due to Meinong. This is to suppose that definite descriptions always pick out something, but sometimes the things picked out do not exist! Rather, according to Meinong, they *subsist*. Meinongianism apaprently handles one type of problematic definition description with ease:

(¶) The present king of France does not exist.

But one might legitimately query whether a real success has been achieved here. For Meinong, apparently empty definite descriptions, like the present king of France, do function referentially: they pick out merely subsisting entities. But then Meinong offers us an explanation as to why we cannot use existential generalisation (the English analogue of the $\mathscr{L}_2$ rule ∃-intro) on such singular terms. For we manifestly cannot use it, else the true claim (¶) would entail this false claim:

(!) There is something that does not exist.

9.3   Compactness and Cardinality

'There are n things' and infinity   We saw before that there is a set of $\mathscr{L}_=$ sentences satisfiable only in infinite domains. If R is a transitive asymmetric relation (both of which properties are expressible by an $\mathscr{L}_=$ sentence), then $\forall x \exists y Rxy$ will be satisfiable in no finite domain. Can we give a set of sentences satisfiable in *all and only* infinite domains?

Using the resources for numerical quantification, define $\mathbf{n}$ as $\exists_n x(x = x)$. Then $\mathbf{3}$ says there are at least three things.

Consider the set $\mathbf{N} = \{\mathbf{n} : n \in \mathbb{N}\}$. $\mathbf{N}$ is not satisfiable in any finite domain: suppose it were satisfiable in domain of size m, then where $n = m + 1$, $\mathbf{n}$ would be false and because $\mathbf{n} \in \mathbf{N}$, $\mathbf{N}$ is not satisfied. So $\mathbf{N}$ is satisfiable only in infinite domains. Moreover, in every infinite domain, $\mathbf{N}$ is satisfied. (exercise).

$\mathbf{N}$ is thus an 'axiomatisation of infinity'. It is a consequence of compactness that there is no single $\mathscr{L}_=$ sentence which is an axiom of infinity (exercise).

Finitude Undefinable   Since there is at least one sentence of $\mathscr{L}_=$ true only in infinite domains, perhaps there is a sentence of $\mathscr{L}_=$ true only in *finite* domains? Perhaps surprisingly, there is not.

Theorem 95 (Finitude Undefinable)

There is no set of $\mathscr{L}_=$-sentences satisfiable in all and only finite domains.

Proof. Suppose FIN is a set of $\mathscr{L}_=$ sentences true in all and only finite domains. $FIN \cup \mathbf{N}$ must be unsatisfiable; by compactness, where Γ is a finite subset of FIN and Δ is a finite subset of $\mathbf{N}$, $\Gamma \cup \Delta$ is unsatisfiable. As Δ is a finite subset of $\mathbf{N}$, there is some greatest $\mathbf{n} \in \Delta$; in any model $\mathscr{A}$ with a domain of size greater than n, Δ is satisfiable in $\mathscr{A}$. Since $\Gamma \cup \Delta$ is unsatisfiable, Γ is unsatisfiable in any such model $\mathscr{A}$, so there is a model with a finite domain (any such $\mathscr{A}$) in which FIN is unsatisfiable, contrary to assumption. So there is no such FIN.

Löwenheim-Skolem Theorem

Theorem 96 (The Löwenheim-Skolem Theorem)

If Γ is a set of sentences of $\mathscr{L}_=$ which has an infinite model $\mathscr{A}$ (of size ω_α), then

Upward Γ *has a model for every infinite size $\omega_\beta > \omega_\alpha$.*

Downward Γ *has a countable model.*

Proof. Upward: Extend the language for Γ to $\mathscr{L}^+$ by adding new constants C such that size of C is ω_β. Consider the set of $\mathscr{L}^+$ sentences $\Gamma^* = \Gamma \cup c_i \neq c_j : i \neq j, c_i, c_j \in C$. Since Γ is satisfiable, every finite subset of Γ^* is satisfiable – by compactness (for this uncountable language!), Γ^* has a model, and the model must have cardinality ω_β.

Downward: too complex to be covered here, unfortunately – see, e.g., Boolos *et al.* (2007: chs. 12–13). □

Overspill

Theorem 97 (Overspill)

If a set of sentences has an arbitrarily large finite model, it has a countable model.

Proof. Let Γ have arbitrarily large finite models. Let $\Gamma' = \Gamma \cup \mathbf{N}$. Any finite subset of Γ' is a subset of $\Gamma \cup \Delta$, where Δ is a finite subset of $\mathbf{N}$, with a greatest $\mathbf{n} \in \Delta$. Since Γ has arbitrarily large models, $\Gamma \cup \Delta$ has a model. But then every finite subset of Γ' has a model; by Compactness, Γ' has a model. And by the downward Löwenheim-Skolem theorem, it has a countable model (Boolos *et al.*, 2007: 147).) □

Skolem's Paradox   The Löwenheim-Skolem theorem says that no first-order theory can restrict itself to models of one infinite cardinality. But set theory is a first-order theory: the Löwenheim-Skolem theorm says it has a

(presuambly unintended) countable model, but set theory itself entails that there are uncountable sets! This is *Skolem's paradox*.

The resolution of the 'paradox' is to note that set theory

> only "says" that [some set] S is nondenumerable in a *relative* sense: the sense that the members of S cannot be put in one-to-one correspondence with a subset of $\mathbb{N}$ by any ***R in the model***. A set S can be "nondenumerable" in this *relative* sense and yet be denumerable "in reality". This happens when there *are* one-to-one correspondences between S and $\mathbb{N}$ but all of them lie outside the given model. What is a "countable" set from the point of view of one model may be an uncountable set from the point of view of another model. (Putnam, 1980: 465)

9.4 Further Directions of Research and Study in Logic

We could now proceed to extend, or alter our logic, in quite a few ways:

Enriching We could add still further to $\mathscr{L}_=$ or $\mathscr{L}_1$, introducing *modal operators* like 'necessarily' and 'possibly', *temporal operators* like 'it was the case that', or even moral operators like 'it ought to be that', or stronger conditionals like the *counterfactual* 'If it had been the case that φ, it would have been the case that ψ'. Then we can prove results about these enriched systems.

Altering Some complain that $\neg\neg\varphi \nvDash \varphi$; these *intuitionists* want to revise our logic. *Free logic* permits constants to lack a designation in a model, and *universally free logic* abandons the requirement that the domain be non-empty; these of course will also revise our logic as given.

Advancing We could augment our languages in another way too, perhaps introducing *function symbols* – operators on terms that yield terms – like + or ×. We can then formalise languages of interest – most

famously *arithmetic*, in the *Peano axioms*, and prove various things about them. In this way we can build up to one of the most celebrated results of twentieth century logic:

Theorem 98 (Gödel Incompleteness Theorems)

If T is a consistent, finitely axiomatisable theory able to express elementary arithmetic, then

- *T cannot prove every arithmetical truth;*

- *T cannot prove its own consistency (i.e., cannot prove an arithmetical sentence that is true iff T is consistent). (This second result obviously entails the first given that T is consistent.)*

The proof is sadly beyond us.

In the next chapter, we will look at some of the philosophical issues arising for the logic we have introduced. In particular, we will look at how well the operator $\rightarrow$ matches the English conditional, and how well the semantics of designators and relations for $\mathscr{L}_2$ matches similar constructions in English. We'll also reexamine the sequent $\varphi, \neg\varphi \vdash \psi$.

Further Reading

Russell's famous paper on definite descriptions is Russell (1956); the nearly as famous response by Strawson is Strawson (1950). A good overall source is Ludlow (2013).

One much-discussed philosophical analysis of the consequences of the Löwenheim-Skolem theorem is Putnam (1980). This article is very controversial! One response is Lewis (1984).

On modal, temporal, etc., logics, a good treatment is by Beall and van Fraassen (2003); a very elegant recent examination of these topics an the topicof nonclassial logic generally is Burgess 2009. Free logics are treated by Bostock (1997: §§8.4–8.7), and Lambert (2001).

A good text on the further development of metamathematics is Boolos *et al.* 2007.

Exercises

1. Prove the theorem called 'Substitution of co-designating constants' from page 130.

2. Prove the following

   (a) $\models \forall x x = x$;

   (b) $\models \forall x \forall y (x = y \rightarrow y = x)$;

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

3. (a) Show that in the presence of =Elim, ∃Elim and ∃Intro, the rule =Intro can be replaced by this rule (where τ is a constant):

   $$\frac{[\exists v\, v = \tau]}{\vdots} \text{=Intro†}$$

   (b) What would be the effect of adopting, instead of our rule =Intro, this rule (in the presence of the other rules of ND_2):

   $$\frac{\varphi}{\tau = \tau} \text{=Intro}^*$$

   (c) Can we replace both =Intro and =Elim by this pair of rules, in the presence of the other rules of ND_2 (φ is a formula in which at most v occurs free):

   $$\frac{\varphi[\tau/v]}{\exists v(v = \tau \wedge \varphi)} \text{=Intro‡} \qquad \frac{\exists v(v = \tau \wedge \varphi)}{\varphi[\tau/v]} \text{=Elim‡}$$

4. Suppose φ is a formula in which at most v occurs free, and in which ν does not occur at all.

   (a) Show that $\models \forall v \forall \nu (v = \nu \rightarrow \varphi \leftrightarrow \varphi[\nu/v])$.

   (b) Suppose for *every* formula φ with just v free, not itself involving identity, and any constants τ, κ, we could show that $|\varphi[\tau/v]|_{\mathscr{A}} = |\varphi[\kappa/v]|_{\mathscr{A}}$. Could we then show that $|\tau = \kappa|_{\mathscr{A}} = T$? (This is intended to mimic as best we can the second-order quantification over properties in the definition of identity: the question is, does it work in the right way?)

5. (a) Show that $\exists_{n+1}xPx \equiv \exists x(Px \land \exists_n y(Py \land \neg x = y))$.

   (b) Define $\forall_{\tilde{n}}x\varphi$ as $\neg\exists_n x\neg\varphi$. Give an English translation of $\forall_{\tilde{n}}xPx$.

6. In the analysis of definite descriptions, $\exists v(\psi \land (\forall v\varphi[v/v] \leftrightarrow v = v))$, could we replace the '$\leftrightarrow$' by '$\rightarrow$'? If not, why? What about in this formulation: $\exists v(\psi \land (\forall v v = v \leftrightarrow \varphi[v/v]))$?

7. Give appropriate analyses in $\mathscr{L}_=$ of these claims, being sure to point out any cases of ambiguity, and commenting on other issues:

   (a) The number of planets is necessarily eight.

   (b) The prime minister must be an honest person.

   (c) There used to be a Labor prime minister.

8. What is the best way to formalise 'x exists' in $\mathscr{L}_=$? What are the difficulties with a Russellian analysis of (¶) 'The present king of France does not exist'?

9. Can the Russellian analysis of descriptions deal with the following sentences?

   (a) 'The beer is all gone.'

   (b) 'He is tall, handsome, and the love of my life.'

   (c) 'She is the proud owner of a Rolls Royce.'

10. Can we use identity and quantification to express 'there are exactly as many Ps as Qs'?

11. (a) Show that the infinite set of sentences $\mathbf{N}$ is satisfied in every infinite domain.

   (b) Show that there is no single $\mathscr{L}_=$-sentence which is satisfied in $\mathscr{A}$ iff $\mathbf{N}$ is satisfied in $\mathscr{A}$.

12. A *permutation* on S is a one-one function π from S to S. If $\mathscr{A}$ is a model, where π is a function from the domain of $\mathscr{A}$ to itself, let the permutation $\pi(\mathscr{A})$ be this model

   - $D_{\mathscr{A}} = D_{\pi(\mathscr{A})}$.
   - For any term τ, $I_{\pi(\mathscr{A})}(\tau) = \pi(I_{\mathscr{A}}(\tau))$.

- For any n-ary predicate Φ,

$$I_{\pi(\mathscr{A})}(\Phi) = \{\langle \pi(x_1), \ldots, \pi(x_n)\rangle : \langle x_1, \ldots, x_n\rangle \in I_{\mathscr{A}}(\Phi)\}.$$

(a) Show the *Permutation Theorem*: for any sentence φ, and permutation π, $|\varphi|_{\mathscr{A}} = |\varphi|_{\pi(\mathscr{A})}$.

(b) Consider this passage:

> the Permutation Theorem shows that, even if we can somehow discount the problems raised by the [Skolem's paradox and the Löwenheim-Skolem Theorem], and settle on models with one fixed domain as those which can truly claim to be abstractions of the world – even then, there are as many different reference relations holding between the subsentential expressions of T's language and the world as there are nontrivial permutations of the fixed domain, all of which underwrite the truth of exactly the same sentences, and, hence, equally subserve the truth of the theses of T. There being no basis for preference between these reference relations, realism's doctrine that there is a unique, privileged such relation is apparently discredited. (Taylor, 2006: 53–4)

Assess and evaluate this argument. Do you accept the conclusion? If not, why not?

Chapter 10

Logic and Natural Language: Conditionals, Entailment, Designators, and Relations

10.1 Conditionals

'If φ, ψ' and $\rightarrow$ So far we have rendered the English conditional construction 'If φ, ψ' using the *material conditional* $\rightarrow$. You may be worried, as others have worried before you, that this is not an accurate rendering. For, you may think, it is implausible that 'If φ, ψ' is true whenever ψ is true. Consider:

(1) The Tigers will not lose every game this season;

(2) The Tigers will make it to the finals;

(3) Therefore: If the Tigers lose every game this season, they will make it to the finals.

If 'if φ, ψ' is $\varphi \rightarrow \psi$, this argument should be valid (it is an instance of the valid form $\varphi, \psi \vDash \neg\varphi \rightarrow \psi$). But this argument does not strike us as valid.

Can we do better in our logic? The first question we should address is: is there a better option in $\mathscr{L}_1$? And it seems that there is not. Suppose that 'If φ, ψ' is expressed by some truth function $f(|\varphi|, |\psi|)$. One thing is definitely true; if $|\varphi| = T$ and $|\psi| = F$, we want $f(|\varphi|, |\psi|) = F$. This leaves us with 8 possible two-place truth-functions. Another thing we want is *asymmetry*: there exists at least one structure $\mathscr{A}$ where $f(|\varphi|_\mathscr{A}, |\psi|_\mathscr{A}) \neq f(|\psi|_\mathscr{A}, |\varphi|_\mathscr{A})$. This rules out four further functions (it rules out all those where $f(F, T) = F$), leaving us with four. Since we've just suggested in our discussion of the boxed argument above that it is undesirable for the truth of a conditional to follow from the truth of the consequent or the falsity of the antecedent, it would surely be even worse if the conditional were *equivalent* to the consequent or the negation of the antecedent. This rules out two of our remaining truth functions, leaving us with just two: $\rightarrow$ and a function g (expressed by $\neg(\psi \rightarrow \varphi)$). But $g(T, T) = F$, hardly what we want from a conditional. So the only truth-function which satisfies minimal conditions necessary to count as a conditional is $\rightarrow$.

Meaning as characterised by derivation rules Of course, this is not a strong argument – that $\rightarrow$ is better than other candidates is no argument that $\rightarrow$ is itself adequate! But we can offer better considerations. Given that our derivation system for $\mathscr{L}_1$ is sound and complete, there is a sense in which those rules – collectively – govern or fix the meaning of the connectives. No other rules are necessary to ensure the right things are provable, the things which correspond exactly to the meaning of the connectives.

Things may not be much different in natural languages. Of course we don't have soundness and completeness proofs, because we have neither a fully spelled out semantics nor a formal derivation system for natural language. But we do have characteristic patterns of inference for the connectives of English. Some of these patterns of inference look remarkably like those for the formal connectives of $\mathscr{L}_1$, notably the rules governing 'and'. (This is some motivation behind the nomenclature of 'natural deduc-

tion' – it is natural in the sense that it formalises intuitively fundamental characteristic patterns of inference for the corresponding operators.)

Whatever else we might want to say about the English conditional connective 'if φ, ψ', it seems *prima facie* it should obey these rules:

Modus Ponens If you have established φ, and you have established 'if φ, ψ', then you can on that basis establish ψ.

Conditional Proof If ψ can be established, conditional on the assumption that φ, then you can establish 'if φ, ψ' without need of that assumption.

It is obvious that something which didn't obey these rules would be quite unlike a conditional. A conditional helps us neatly summarise reasoning from assumptions, store it, and then use it when those assumptions come true – conditional proof captures the first part, and *modus ponens* the last. But if, as seems very plausible, these two rules capture the English conditional, then we can offer an argument that the English conditional is $\rightarrow$. More precisely, the argument is that 'If φ, ψ' is true iff $\varphi \rightarrow \psi$ is:

> *Only If*: Suppose that if φ, ψ. Assume φ. By *modus ponens*, ψ. By $\rightarrow$Intro, $\varphi \rightarrow \psi$, discharging the assumption.
>
> *If*: Suppose that $\varphi \rightarrow \psi$. This is equivalent to $\neg\varphi \vee \psi$. Assume φ. By disjunctive syllogism, ψ. By conditional proof, it follows that if φ, ψ, without need of the assumption.

'Disjunctive syllogism' is the rule that from $\varphi \vee \psi$ and $\neg\varphi$ one can infer ψ. (A problem asks you to show that this is valid derived rule in the system $\mathscr{L}_1$; it is obviously a good rule for the English 'or'.)

Replying to the apparent counterexamples – Grice If this argument works, we're wrong to think that the cases where $\rightarrow$ doesn't seem to behave like 'If...then' are genuine counterexamples. But we need to say something about those cases.

The explanation offered is that, even though these sentences are true, they *conversationally implicate* something false, and that is what we are responding to. The notion of conversational implicature was introduced by Grice, who noted that many utterance seem to communicate more than what is strictly said, and gave a series of principles that he argued govern what is communicated by an utterance, based on what strictly it means (Grice, 1989). One classic example is this: 'some philosophers have beards'. An utterance of this sentence conversationally implicates that not all philosophers have beards, and that is what something most people will assume the speaker to be committed to in their utterance. Yet this is not a consequence of the utterance: 'some philosophers have beards; in fact, all of them do!' can perfectly well be true, and is no contradiction.

The Maxim of Quantity   The principle Grice invokes to explain this implicature is his 'Maxim of Quantity' – simply put, this is the principle that cooperative speakers be as informative as they can be. A hearer, without good reason to think otherwise, will assume that a speaker is being cooperative, and hence that when the speaker utters a claim, it is also the most informative claim the speaker is in a position to responsibly utter. One measure of informativeness is this: If φ entails ψ, then φ is at least as informative than ψ (for it carries the information that ψ, and perhaps additional information too if ψ does not entail φ). Since 'All philosophers have beards' entails 'some philosophers have beards', someone who utters the latter utters something less informative than they could utter, if in fact they thought that all philosophers have beards. Since they are cooperative and did not utter the stronger claim, therefore, the hearer infers that the speaker did not believe the stronger claim. So the hearer now knows that the speaker must believe that some but not all philosophers have beards, and the speaker therefore communicates that not all philosophers have beards to the hearer.

Applying Quantity to conditionals The material conditional is equivalent to a disjunction $\neg\varphi \vee \psi$. Any disjunct of a disjunction entails the disjunction, so contains more information. So an utterance of a disjunction implicates that neither disjunct is believed by the speaker. As Thomson put it,

> In saying 'if p then q' a speaker will say something which is in general anyway true or false. But by the act of making the statement he will do other things, too. He will encourage us to think that he has some or other reason for thinking that if p then q and that his reasons are not such as to allow him to assert not-p nor such as to allow him to assert q. (Thomson, 1990: 67–8)

So if it is apparent that the speaker *does* believe either disjunct – either through being committed to the consequent, or rejecting the antecedent, of the relevant material conditional – we should be struck by the fact that the speaker has communicated an apparent contradiction.

In the case of the Tigers, the speaker who utters both the premises and conclusion of this argument communicates that they believe the consequent, and also utters the conditional which conversationally implicates that the speaker does not believe the consequent. So the hearer has a contradiction communicated to them – little wonder, then, that these sentences sound so bad! But – and this is the crucial part – nothing in this explanation of the badness of the conditional is *semantic*. Nothing in this explanation undermines the idea that the truth conditions of the English conditional construction are the same as those for the material conditional. This is all to do with what we standardly take ourselves to be able to infer from an utterance of a sentence with those semantics, and what goes wrong here is that inference, not the semantics.

Trickier Cases   All is not smooth sailing for the material conditional account of 'If φ, ψ'. I mention two trickier cases:

1. There seem to be cases where conditional proof goes wrong – most noticeably, in the case of assumptions made that explicitly contradict other things we believe. This is especially apparent in so called *counterfactual* conditionals – those (at least at first glance) that involve conditional reasoning about circumstances that do not obtain. So consider the (apparently true) counterfactual 'If Univ had never existed, Exeter would have been the third oldest college'. The conditional proof rationale for this is: if we can derive the consequent from the assumption that Univ never existed and other things, then we can derive the conditional from those other things alone. Yet since Univ's existence is known to us, adding that assumption to our prior beliefs gives an inconsistent set of assumptions, from which any consequent whatever follows. But 'If Univ had never existed, Catz would have been the third oldest college' is clearly false, though it should be derivable on exactly the same grounds. What we need to do, apparently, is modify which of our prior beliefs can be retained when reasoning from this contrary-to-fact assumption.

   Many philosophers are agreed, however, that such cases should be treated by analysing the 'if' involved in counterfactuals as distinct from the 'if' of the so-called *indicative* conditionals we've been dealing with so far – if that suggestion is adopted, we may retain a material conditional analysis of indicative 'if' while giving an alternative treatment of counterfactual 'if'. The classic early account of counterfactual 'if' is Lewis (1973).

2. There seem to be cases where the Gricean explanation should make a conditional sound bad, but in which it does not. Consider: 'You won't fail EDL. Even if you do, you will pass on the re-sit.' The first sentence

of this little speech is a denial of the antecedent of the following conditional, so the conditional should sound bad. But it sounds fine. The word 'even' may be held responsible; but it is now up to the Gricean to explain how it interferes with the ordinary Gricean mechanisms.

The Gibbard Argument

Theorem 99 (Gibbard)

Any conditional operator $\Rightarrow$ *which satisfies these three conditions is equivalent to the material conditional (Gibbard, 1981):*

Import-Export $\varphi \Rightarrow (\psi \Rightarrow \chi)$ *and* $(\varphi \wedge \psi) \Rightarrow \chi$ *are logically equivalent.*

Lower $\varphi \Rightarrow \psi$ *entails* $\varphi \to \psi$ *(i.e., the conditional entails the material conditional).*

Upper $\varphi \vDash \psi$ *entails* $\varphi \Rightarrow \psi$ *(i.e., entailment entails the conditional)*

Proof. Left for exercise. □

The relevance of the theorem is that the English conditional 'if' does seem to satisfy Upper, Lower, and Import-Export. The only mildly controversial one is apparently the latter, but support for it can be found by considering cases in which the application of Import-Export seems unproblematic. Consider this pair, which seem to express the same claim:

(4) a. If you drink one more can of beer then if I drink one more can of beer then we'll be completely out of beer.

　　 b. If you drink one more can of beer and I drink one more can of beer then we'll be completely out of beer. (Kratzer, 2012: 88)

The Trickiest Case　Here's another problem case:

Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat

> Jimmy Carter, with the other Republican in the race, John An-
> derson, a distant third. (McGee, 1985)

There are only two Republicans in the race, so this is trivially true:

(5)   If a Republican wins and it's not Reagan, then it's Anderson who
wins.

(5) is a conditional with a conjunctive antecedent; if Import-Export is valid
for the English 'if', then we can derive

(6)   If a Republican wins the election, then if it's not Reagan who wins
it will be Anderson.

And since the polls overwhelmingly favour Reagan, it is almost certain he
will win, and so almost certain that the antecedent of this conditional (6)
- i.e., 'A Republican will win the election' – is true. It was certainly widely
believed by people at the time that a Republican would win, and it turned
out to be true in the end.

Assuming that if someone believes a conditional, and believes the ante-
cedent of that conditional, they can validly reason (using *modus ponens*) to
the consequent, this follows:

(7)   If it's not Reagan who wins, it will be Anderson.

But this is false: Anderson was a distant third in the polls, and the most
likely alternative to Reagan was Carter.

What has gone wrong? We've derived, from claims we believe to be
true ((5) and that a Republican will win), using Import-Export and *modus
ponens*, a claim we believe to be false: 7. The defender of the material con-
ditional account of 'if' cannot object to either of these rules, since both
Import-Export and *modus ponens* are provably correct derivation rules for
the material conditional. So they have to argue that the apparently false
conclusion is in fact true; or else deny either the trivial truth (5), or deny
that it is acceptable to reason from the highly probable claim – that in

fact turned out to be true, just as the polls indicated – that a Republican would win the election. None of these options looks particularly appealing, and most philosophers and linguists have concluded that the material conditional account of 'if' is thereby refuted. However, there has been no consensus about what to offer in its place.

10.2   Entailment

Another problematic feature of $\mathscr{L}_1$ (somewhat related to the foregoing by the deduction theorem) are the so-called *paradoxes of entailment*. These are the following valid sequents:

- $\varphi, \neg\varphi \vDash \psi$;

- $\varphi \vDash \psi \vee \neg\psi$.

But, some have objected, these are terrible inferences – we should never conclude, on the basis of contradictory premises, anything; and while it may be safe to conclude a tautology from any premises, that is hardly a good argument to persuade someone that the conclusion is a tautology.

Relevance   Critics of these sequents have maintained that the failure here is one of *relevance*. The arguments may be valid in the narrow sense, because there is no structure which satisfies the premises without also satisfying the conclusion (in the former case, because no structure satisfies the premises; in the latter, because every structure satisfies the conclusion). But validity in this narrow sense is too narrow – what we ordinarily understand to be a successful valid argument excludes these trivial cases of validity. The conclusion in a *genuinely* valid argument – these critics say – should follow from, and be relevant to, the premises.

The Lewis Argument   Yet there is a very simple argument, using ordinary English reasoning which is apparently impeccable, to support these 'problematic' sequents. It is known as the *Lewis argument*, after C. I. Lewis, one of its modern rediscoverers.

(8)	φ	Assumption
(9)	φ or ψ	Disjunction introduction, 1
(10)	$\neg\varphi$	Assumption
(11)	ψ	Disjunctive syllogism, 2, 3

This argument seems to show that, even ordinarily, this form of apparently irrelevant argument is valid. And who could object to disjunction introduction or disjunctive syllogism, which seem like about the most basic things one can safely say about disjunction.

The relevantist response – reject Disjunctive Syllogism   The relevantist has a response to this argument – reject disjunctive syllogism. For, they say, it goes wrong in precisely this circumstance. The inference works normally because, given a true disjunction with one false disjunct, we can infer to the truth of the other disjunct *without* checking to see whether it is true. (Disjunctive syllogism was called 'the dog' by its Stoic inventors, because 'even a dog uses this form of inference when it comes to a fork in the road, sniffs down one branch, and not finding the scent there immediately takes off down the other branch, without stopping to sniff' (Burgess, 2009: 99–100).) But in this case, φ is also an assumption – so the fact that $\varphi \vee \psi$ has a disjunct whose denial is an assumption doesn't entail ψ. If there are cases where we can have inconsistent assumptions, like this, we should expect disjunctive syllogism to go wrong.

Dialethism   Some truly radical support for this line of argument can be found from the *dialethist* position that some sentences are *both* true and

false. If φ is one such sentence, then both assumptions can be true (φ because it is true, $\neg\varphi$ because φ is false). But ψ, since it is arbitrary, can be chosen to be a plain falsehood, giving a situation which is a counter-example to disjunctive syllogism. Dialethists motivate their non-standard semantics by appeal to particular problem cases that don't seem to admit of any other solution. The best known is

L   L is false.

The sentence L displayed above says of itself that it is false. If it is true, then L is false (as that is what it says). But if it is false, then what it says is false, so it must be true. So L is true iff it is false. The dialethist takes this proof at face value – L is both true and false. But, despite the great difficulties more standard approaches to the Liar paradox L have, few have found dialethism compelling (though it has surprisingly able defenders).

Inconsistent belief   Dialethism is too incredible to convincingly motivate the rejection of disjunctive syllogism. A weaker motivation is found in reasoning about what we believe. If you are like me, you probably have some inconsistent beliefs that you are not aware of. It is absurd to say that I therefore believe everything, just because I have beliefs that entail everything. So it seems, the set of my beliefs is not plausibly closed under classical entailment (that is, the set of my beliefs does not contain everything which is entailed by some things in my set of beliefs). Disjunctive syllogism seems to fail for my beliefs; just because $\varphi \vee \psi$ and $\neg\varphi$ are among my beliefs does not mean that ψ is one of my beliefs; for φ may be one of my beliefs as well. The logic of belief, it may be said, is relevant. (The original application of this kind of idea was to simple inference-drawing computers – if they are not aware of when the body of information input contains a contradiction, you should not allow them to draw conclusions using full classical logic.) The problem with all this is that it doesn't really seem like logic any more. There are no counterexamples to disjunctive syl-

logism, that is, structures where $\varphi \lor \psi$ and $\neg\varphi$ are true but ψ is false. Rather, there are cases where *according to my beliefs* $\varphi \lor \psi$ and $\neg\varphi$ are true, and also *according to my beliefs* ψ is not true. But it is already well-known that, when we have a non-truth-functional operator like 'according to my beliefs', it won't generally be true that classically valid sequents remain correct when each sentence in the sequent is in the scope of the operator. (Think of the non-truth-functional operator 'possibly': $\varphi, \psi \vDash \varphi \land \psi$ is correct, 'possibly φ, possibly $\psi \vDash$ possibly $\varphi \land \psi$' is not correct – think of the case where ψ = 'the coin toss lands heads', φ = 'the coin toss lands tails'.)

10.3   Designators

Designators in $\mathscr{L}_2$ and English   The formal properties of $\mathscr{L}_2$ have been established; now we begin a relatively informal investigation into the possible connections between the semantics of $\mathscr{L}_2$ and the semantics of English. We begin in this section with a discussion of *designation* in both languages.

Direct Reference $\mathscr{L}_2$ embodies a particularly simple view of the meaning of a constant.

Definition 67 (Direct Reference). A referring expression *directly refers* iff the meaning of the name is just its referent, the object denoted by the expression.

Constants in $\mathscr{L}_2$ directly refer, because the meaning (semantic value) of a in $\mathscr{A}$ is $|a|_{\mathscr{A}}$, which is just an item in the domain of $\mathscr{A}$.

   This is to be contrasted with any thesis of *indirect* reference, according to which the semantic value of a denoting expression is something other than the referent. The primary example here are definite descriptions; if Russell's account (recall chapter 9) is correct, for example, the meaning of a description is somehow to be extracted from a complex quantified

sentence; the referent (if any) is determined by the meaning, but is not the meaning. This can clearly be seen if we accept Russell's treatment of 'The present king of France is bald', where the sentence is meaningful even though the definite description fails to denote anything. Given that the meaning of the sentence is determined by the meanings of its constituents, the definite description noun phrase must have a meaning, even though it lacks a referent, and hence cannot be a directly referential expression.

Rigid Designation   Direct reference is related to, but distinct from, the following notion:

Definition 68 (Rigid Designator). A referring expression a is a *rigid designator* iff the referent of a is the same object in every possible situation (possible world) in which that object exists.[1]

It may seem that this doesn't hold in $\mathcal{L}_2$, because if $\mathscr{A} \neq \mathscr{B}$, then it may be that $I_{\mathscr{A}}(a) \neq I_{\mathscr{B}}(a)$. But for all that it may hold; the crucial fact to recognise is whether, *in a given situation*, when we consider what is possible for a, we consider what is possible for what a refers to – in every situation in which that thing exists. Since $\mathcal{L}_2$ has no resources for discussing possibility, it is a moot point whether or not rigid designation holds for it.

If an expression is directly referential, in a language with resources for discussing possibility and necessity, then it looks like it will also be a rigid designator. Assume that we evaluate 'Possibly, $F(a)$' by seeing whether there is a possible world in which $F(a)$. If a directly refers, the evaluation of 'Possibly $F(a)$' will proceed by evaluating whether actual meaning of a

---

[1]Be careful not to confuse this with the claim that every possible use of 'a' refers to the same thing – this would be false. The idea is rather, when we actually evaluate the possibility of various things about a in some possible world w, we continue to hold fixed the actual meaning of a, rather than (for example) what the residents of w themselves would mean by their own utterance of 'a'. Suppose we consider the possibility in which Aristotle hadn't been called 'Aristotle': we are still talking about the person actually called 'Aristotle', even though residents of that possibility do not use that name to denote anyone.

– which is just some object **a**, given direct reference – satisfies F at each possibility in which **a** exists. So the actual meaning of a, evaluated at each world, will remain constant.

The classic example of a non-rigid designator is also a simple definite description, like 'the tallest person'. Suppose Jacques is the tallest person. Still, I can perfectly well say 'Possibly, Gill is the tallest person', which is *not* synonymous with 'Possibly, Gill is Jacques'. At the possible world in which Gill is the tallest person, the expression 'the tallest person', with its actual meaning, denotes her, though it doesn't denote her actually.

Things are complicated by the existence of *rigidified descriptions*. As the name suggests, they are rigid designators, but they are also descriptions which refer indirectly. Consider 'the actual tallest person': intuitively, the meaning of this description evaluated at every possibility determines it to have a constant referent, of Jacques. Because of this, 'Possibly, Gill is the actual tallest person' does seem to entail that possibly, Gill is Jacques.

Names and Direct Reference in English   For names in English, like 'Antony Eagle' or 'Aristotle', there is no straightforward semantics for names as there is for constants in $\mathscr{L}_2$.

On the one hand, there is considerable syntactic and semantic evidence that names in English function rather like the constants of $\mathscr{L}_2$. Consider

(12)   Antony Eagle is lecturing now.

(12) is true iff the thing **Antony Eagle**, designated by 'Antony Eagle', has the property **lecturing** denoted by 'is lecturing', **now**, i.e., at the time designated by (this use of) 'now'. These seem remarkably like the clauses for the satisfaction of atomic formulae of $\mathscr{L}_2$; in that sense it looks like the meaning of 'Antony Eagle' is just the thing it refers to.

Objection to Direct Reference: Informative Identities   If 'Hesperus' denotes Hesperus, and 'Phosphorus' denotes Phosphorus, then, since Hes-

perus *is* Phosphorus (both are names for Venus), then the meanings of 'Hesperus' and 'Phosphorus' are the same. So then why aren't these sentences synonymous?:

(13)   Hesperus is Hesperus.

(14)   Hesperus is Phosphorus.

This objection trades on the fact that we receive information when told the second claim, but we do not when told the first; hence they must yield different information, and not be synonymous after all. One possible response is to say that *for all we know*, we are in a situation where 'Phosphorus' names something other than Venus; and what we are told by the second sentence is a contingent fact about English, namely, that 'Phosphorus' means Phosphorus!

Objection to Direct Reference: Empty Names   If the meaning of a name is its referent, then 'Santa Claus is jolly' is meaningless, yet we judge it to be true. And don't say it is true 'in a fiction' – for fictional English is the same as non-fictional English, so if it is meaningless outside of the fiction, it is just as meaningless inside. But set that aside: *negated existentials* show that 'Santa Claus' is meaningful, because 'Santa Claus does not exist' is straightforwardly true.

To be more explicit: 'Santa Claus does not exist' apparently expresses a truth. This can be rendered in the language $\mathscr{L}_=$ by this translation: '$\neg \exists x (x = a)$', where 'a' denotes Santa Claus. But since '$a = a$' expresses a logical truth, it follows that '$\exists x (x = a)$' expresses a logical truth of $\mathscr{L}_=$, denying the apparent truth which we translated. But this looks like a problem in English, as well as in formal languages: since 'Santa Claus is identical to Santa Claus' also expresses a truth, the English claim 'Something is identical to Santa Claus' is also true. But if something is identical to Santa Claus, Santa Claus does exist. To deny this is to deny claims that seem tautologous in English.

This is a very serious objection to direct reference theories. Many have chosen to bite the bullet, and accept that such sentences are in fact meaningless. Others have chosen to accept some secondary notion of meaning, like a canonical description associated with the empty name, to play a name-like role in these kind of claims. But there is no consensus on how to deal with this problem.

Objection to Direct Reference: The Role of Descriptions Many names are simply introduced by descriptions; and certainly what most of us aim to communicate when we use a name is a certain individual satisfying a certain canonical description that the hearer is familiar with – if there were no such description, 'how do people ever use names to refer to things at all?' (Kripke, 1980: 28).

Reference Fixing Some descriptions are used to establish the reference: 'Julius' may be introduced into the language as 'the inventor of the zipper'. But it is no part of the meaning: consider the counterfactual situation where Julius gets pipped by his rival.

Background knowledge If one knows a lot about Julius, and one's hearers do too, then one will expect that background knowledge to be useable when one uses the name 'Julius'. It may be difficult if not impossible for a speaker themselves to separate the meaning of 'Julius' from what they know about Julius.

Opaque Contexts Consider

(15)   Lois Lane believes that Superman is Superman;

(16)   Lois Lane believes that Superman is Clark Kent.

It is fairly clear that in the context 'Lois Lane believes that Superman is x', one cannot straightforwardly substitute co-referring expressions into x and get a true sentence. Such *opaque contexts* – those which do not permit

substitution *salva veritate* – are a problem for the direct reference theory, of course, but also for other theories that think the objects of belief are propositions concerning the entities one has the belief about.

Obviously $\mathscr{L}_2$ lacks opaque contexts, as theorems about substitution of co-referring constants proved in Chapter 7 show. This is an expressive weakness of $\mathscr{L}_2$ compared with English.

Other Designators in English   There are several other natural language expressions other than names and definite descriptions which have a designator role in English.

Indexicals such as 'I', 'you', 'here', 'now', 'actually'. This class may also include *demonstratives* like 'this', 'that'.

Count Nouns such as 'chair', 'boy', 'piece of furniture'.

Non-Count Nouns such as 'gold', 'dirt', 'intelligence'.

Count nouns can be handled by the apparatus of quantification in $\mathscr{L}_2$, rather than as designating expressions directly.

Indexicals   The characteristic of an indexical expression is *context sensitivity*. 'I' refers to different people when uttered by different people: to me when I say 'I am tired', to you if you utter that same string of words. Similarly with other expressions in this category, as they are sensitive to features of context – where the context of an utterance includes the speaker, time, place, etc., of the utterance – and thus may have different referents in different contexts.

Despite the context-sensitivity of indexical *reference*, their meaning appears constant. Distinguish (following Kaplan):

Content The content of an expression is its semantic value;

Character The character of an expression is a function from context to content.

The thesis is that indexicals have a constant character, but a variable content; and character is plausibly a kind of meaning.

Pronouns – 'she', 'they', 'it' – may perhaps be analysed similarly.

Non-Count Nouns Non-count nouns are a certain type of *non-plural* referring expression. One reasonable test is whether 'μ' can be grammatically substituted into the expression 'I want some μ'. These fall into at least two kinds:

Mass Mass nouns, like 'gold' or 'dirt', denote a kind of stuff or substance, rather than a particular thing. They typically display *cumulative reference*: if 'being gold' is true of a and b, then it is true of the *sum $a+b$* which has just a and b as parts.

Abstract Abstract nouns, like 'intelligence' or 'generosity', which seem to denote *qualities* that something may possess, but can themselves have things predicated of them.

A full treatment of these nouns cannot be carried out in $\mathcal{L}_2$; constants referring to qualities, that we naturally translate as predicates like 'being gold', may require a *second order* logic, which permits quantification over predicate variables as well as individual variables.

10.4 Binary Relations

Binary Relations Binary relations – any subset of $D_\mathcal{A} \times D_\mathcal{A}$ – have many interesting properties. We can often represent these properties *graphically*, if the domain on which the relation is defined is finite.
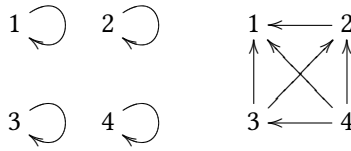
Figure 10.1: Reflexivity and Transitivity

Definition 69 (Finite Graph). A *finite graph* on $D_{\mathscr{A}}$ is a pair $\langle V, E \rangle$ where the set of *vertices* V is $D_{\mathscr{A}}$, $D_{\mathscr{A}}$ is finite, and E is a set of *edges*, directed 'arrows' from members of V to members of V.

A graph on $D_{\mathscr{A}}$ *corresponds to* a binary relation R on $D_{\mathscr{A}}$ just in case

1. If $\langle x, y \rangle \in R$, then $x \in V$ and $y \in V$.

2. If $\langle x, y \rangle \in R$, then there is an arrow from x to y in E.

Reflexivity and Transitivity Represented Graphically Consider the domain $D_{\mathscr{A}} = \{1, 2, 3, 4\}$.

Definition 70 (Reflexive). A relation on $D_{\mathscr{A}}$ is *reflexive* iff for all $x \in D_{\mathscr{A}}$, $\langle x, x \rangle \in R$.

Graphically, that means, as in the figure on the left in Figure 10.1, that each vertex has an arrow pointing to itself. A relation is *irreflexive* iff no vertex has an arrow pointing to itself.

Definition 71 (Transitive). A relation on $D_{\mathscr{A}}$ is *transitive* iff for all $x, y, z \in D_{\mathscr{A}}$, if both $\langle x, y \rangle \in R$ and $\langle y, z \rangle \in R$, then also $\langle x, z \rangle \in R$.

Graphically, that means, as in the figure on the right in Figure 10.1, that whenever there is an indirect sequence of arrows between two vertices, there is also a 'shortcut'. A relation is *intransitive* iff there is never a shortcut.
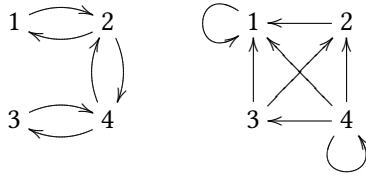
Figure 10.2: Symmetry

Symmetry Represented Graphically

Definition 72 (Symmetric). A relation R on $D_{\mathscr{A}}$ is *symmetric* iff whenever $\langle x, y \rangle \in R$, then $\langle y, x \rangle \in R$.

Graphically, this means (as on the left in Figure 10.2) that whenever there is an arrow from one vertex to another, there will be a return arrow. A relation is *asymmetric* iff there are no returning arrows at all.

Definition 73 (Anti-symmetric). A relation is *anti-symmetric* iff whenever $\langle x, y \rangle \in R$ and $\langle y, x \rangle \in R$, then $x = y$ (sometimes this is known as *weak asymmetry*).

Graphically, this means (as on the right in Figure 10.2) that the only 're-turning' arrows are loops.

We need to clearly distinguish non-reflexivity from irreflexivity: the former is just the failure of reflexivity, i.e., at least one object in the domain doesn't bear R to itself, while the latter involves every object in the domain not bearing R to itself. Likewise, we need to distinguish non-transitivity from intransitivity, and non-symmetry from asymmetry *and* antisymmetry.

The observant reader will have noticed that, while we defined *reflexive on D*, *symmetric on D*, etc., the only definition which actually mentions D is the definition of reflexivity. All the other definitions are *conditional* in form: they say, if certain pairs are in the relation, then certain other pairs will be too. We needn't specify a domain to check whether these conditionals

hold of any relation. But to check whether a relation is reflexive, we need not only the ordered pairs of the relation, but also what the domain is, so we can see if any members of the domain are missing from the relation. (Reflexivity is an extrinsic property of a relation, relative to a domain; the others are intrinsic properties.)

Equivalence Relations

Definition 74 (Equivalence Relation). If R on $D_{\mathscr{A}}$ is reflexive, transitive, and symmetric, then it is an *equivalence relation*.

An equivalence relation divides up the domain into *equivalence classes*, every member of which bears the relation to every other member of it. The relation 'is the same height as' is an equivalence relation on the domain of all people; we say that it *induces a partition* on the domain of people, sorting them into groups which are uniform with respect to height.

The most obvious equivalence relation is the identity relation; it is reflexive by definition, and moreover symmetric and transitive; since there are *never* arrows from any x to any distinct y, it follows that every time there are such arrows, there are return arrows and shortcuts. (Notice that the identity relation is both symmetric and anti-symmetric.)

Any relation on the empty set is also trivially an equivalence relation. While every empty relation is trivially symmetric and transitive, it will generally fail to be reflexive unless the domain is empty too.

Connectedness   First, we define a (directed) path:

Definition 75 (Path). There is a *path* from x to y in R iff there exists a sequence $z_1, \ldots, z_n$ such that $\langle x, z_1 \rangle \in R$ and $\langle z_1, z_2 \rangle \in R$ and ...and $\langle z_n, y \rangle \in R$.

Now we may define:

Definition 76 (Strong Connectedness). R on D is *connected* iff for any $x \in D$ and $y \in D$, such that x and y are distinct, then there is a path from x to y in R.

There is another property of binary relations, sometimes called *weak connectedness*, which holds iff for any nodes on the graph of R, x and y, there is some path that joins x to y moving only along arrows (perhaps in the 'backwards' direction). For any binary relation S, let S^* be the relation defined by $\langle x, y \rangle \in S^*$ iff $\langle x, y \rangle \in S$ or $\langle y, x \rangle \in S$.

Definition 77 (Weak Connectedness). A relation R is weakly connected on D iff R^* is connected on D.

There is an even stronger condition than connectedness:

Definition 78 (Totality). A relation R on D is *total* iff for any x and y, either $\langle x, y \rangle \in R$ or $\langle y, x \rangle \in R$.

Theorem 100
If R is total, then R is reflexive.

Proof. If R is total, then for any x and y, either $\langle x, y \rangle \in R$ or $\langle y, x \rangle \in R$. Suppose $x = y$; then either either $\langle x, x \rangle \in R$ or $\langle x, x \rangle \in R$. Therefore, $\langle x, x \rangle \in R$, i.e., R is reflexive. $\qquad\square$

Inverse and Complement Relations Given a relation R on D, we can define two interesting further relations:

Definition 79 (Inverse). Given a relation R, its *inverse R^{-1}* is defined to be this set: $\{\langle y, x \rangle : \langle x, y \rangle \in R\}$.

Definition 80 (Complement). Given a relation R on D, its *complement R'* on D is defined to be this set: $D \times D \setminus R$ (i.e., the set of all ordered pairs of elements of the domain *not* in R.)
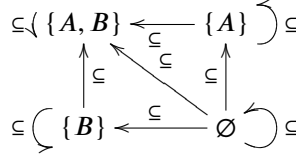
Figure 10.3: Partial ordering of a set of sets by $\subseteq$

Note that complement is always with respect to the underlying domain, whereas we can define an inverse relation just with reference to the original relation.

Theorem 101 (Property Preservation)

If R is reflexive then R^{-1} is reflexive and R' is irreflexive; if R is symmetric then R^{-1} and R' are symmetric; if R is transitive then R^{-1} is transitive; if R is connected then R^{-1} is connected.

Proof. I prove one case: reflexivity. If R is reflexive, then for all $x \in D_{\mathscr{A}}$, $\langle x, x \rangle \in R$. By definition of inverse, then, all $\langle x, x \rangle \in R^{-1}$; i.e., R^{-1} is also reflexive. By definition of complement, *no* $\langle x, x \rangle \in R'$; so R' is irreflexive.□

Orderings

Definition 81 (Ordering). A relation R on $D_{\mathscr{A}}$ is a

Partial order iff it is reflexive, transitive, and anti-symmetric (e.g., $\subseteq$ on a set of sets; see Figure 10.3);

Strict Partial Order iff it is transitive and asymmetric (i.e., $\subset$);

Total Order iff it is a total partial ordering (i.e., $\leqslant$ on $\mathbb{N}$). Also sometimes known as a *linear order*;

Strict Total Order iff it is a total strict partial order (i.e., $<$).

As defined above, a total order is reflexive, transitive, antisymmetric, and total. By Theorem 100, a total order is reflexive. So we could have defined a total order as one that is total, transitive, and antisymmetric.

Well-Orderings   Suppose R is a total order on $D_{\mathscr{A}}$. If $\langle x, y \rangle \in R$, then we say x *precedes* y. If $\langle x, y \rangle \in R$ and $x \neq y$ and there does not exist a z such that $\langle x, z \rangle \in R$ and $\langle z, y \rangle \in R$, then x *immediately precedes* y. (We can give exactly similar definitions for *succeeds*.)

An element x of $D_{\mathscr{A}}$ is *minimal* iff there is no distinct y that precedes x. An element x is *least* if it precedes *every distinct* $y \in D_{\mathscr{A}}$. (Corresponding definitions can be given for maximality and greatest.)

Definition 82 (Well-ordering). R is a well-ordering of D iff it is total and every non-empty subset d of D has a least member under R.

Example: $\leqslant$ is a well-ordering of the natural numbers $\mathbb{N}$. Here is another well-ordering of $\mathbb{N}$:

$$
(10.1) \qquad x \preccurlyeq y \quad \text{iff} \quad
\begin{cases}
x \text{ is even and } y \text{ is odd; or} \\
x \text{ and } y \text{ are even and } x \leqslant y; \text{ or} \quad \boxtimes \\
x \text{ and } y \text{ are odd and } x \leqslant y.
\end{cases}
$$

The order induced by $\preccurlyeq$ yields $0, 2, 4, \ldots, 1, 3, 5, \ldots$. In this ordering, every subset of $\mathbb{N}$ has a $\preccurlyeq$-least element. Notice, however, that 1 is not least, but it has no immediate predecessor.

While $\leqslant$ well-orders $\mathbb{N}$, it does not well-order the positive and negative integers $\mathbb{Z}$. Here is a well-ordering of $\mathbb{Z}$: $x \leqslant_{\mathrm{abs}} y$ iff either (i) $|x| \leqslant |y|$ or (ii) $|x| = |y|$ and $x \leqslant y$, where '$|x|$' here denotes the absolute value of x. This induces the order $0, -1, 1, -2, 2, \ldots$.

Expressing Properties of Binary Relations in $\mathscr{L}_{2=}$ We can use formulae of $\mathscr{L}_2$ to express various properties of relations on an entire domain. For example

Reflexivity A relation R is reflexive on $D_{\mathscr{A}}$ iff $R = \left| P^2 \right|_{\mathscr{A}}$ and
$\left| \forall x P^2 x x \right|_{\mathscr{A}} = T.$

Transitivity A relation R is transitive on $D_{\mathscr{A}}$ iff $R = \left|P^2\right|_{\mathscr{A}}$ and

$$\left|\forall x \forall y \forall z (P^2 xy \wedge P^2 yz) \rightarrow P^2(xz)\right|_{\mathscr{A}} = T.$$

Symmetry A relation R is symmetric on $D_{\mathscr{A}}$ iff $R = \left|P^2\right|_{\mathscr{A}}$ and
$\left|\forall x \forall y (P^2 xy \rightarrow P^2 yz)\right|_{\mathscr{A}} = T.$

Relations in English and $\mathscr{L}_2$: Intension and Extension Is the semantic value of an English predicate expression a property or relation as those are interpreted in $\mathscr{L}_2$-structures?

The property 'is a person' on the domain of Oxford students is the same property as 'is a student'; but even though they coincide in their members, we do not think these predicates are synonymous. We may distinguish the property given

- *in extension*, by the list of things which satisfy it; or

- *in intension*, by a rule.

On this domain, two different rules give the same extension; and intuitively in English the meaning is the property given intensionally.

Relations in English and $\mathscr{L}_2$: Sparse and Abundant Properties *Any* set of pairs is a binary relation. Yet this means that 'is identical to' is just as legitimate a relation as 'is within 21 metres of'; whereas we think the former is a more fundamental and non-accidental relation than the latter. We may think that the 'real' properties are sparse and fundamental, and the abundant predicates like 'grue' and 'within 21 metres of' are less so, and do not correspond to genuine properties – but $\mathscr{L}_2$ hardly allows us to draw that distinction.

So despite the fact that 'grue' and 'green' are both meaningful predicates of English, we want to give quite different things to be their semantic

values – perhaps just a set of things for 'grue', but a real property **greenness** to be the semantic value of 'green'.

Despite these problems, the case against English having a similar semantics for predicates to $\mathscr{L}_2$ is rather weaker than that against English having a similar semantics for designators.

Further Reading

Beall and van Fraassen (2003) and Burgess (2009), cited in the last chapter, provide useful material on relevant and conditional logics. On the topic of the English indicative conditional, a good guide is Edgington (2014). In addition to the work of Grice and Thomson cited above, Davis (2014) is a useful source on conversational pragmatics. Stalnaker (1975) gives an account of the English conditional which invalidates Import-Export; McGee (1985) argues that in fact the English 'if' doesn't satisfy *modus ponens*. Edgington and Kratzer (2012) discuss more exotic accounts.

The best defence of dialethism is Priest (2006); the application of relevant logic to inconsistent data sets is Belnap (1977).

A wonderful book on names and description is Kripke (1980). A modern classic: amazingly, it is more or less a transcription of a series of lectures originally delivered without notes. A useful source on empty names is Caplan (2006). The topic of empty names is one main topic of another famous series of lectures by Kripke (2013).

Exercises

1. (a) Show that the following derivation rules are acceptable in our framework:

    i. *Modus Ponens*: If $\vDash \varphi$ and $\vDash \varphi \rightarrow \psi$, then $\vDash \psi$.

    ii. *Disjunctive Syllogism*: If $\vDash \varphi \vee \psi$ and $\vDash \neg\varphi$, then $\vDash \psi$.

    iii. *Reductio ad Absurdum*: If $\varphi \vDash \psi$ and $\varphi \vDash \neg\psi$ then $\vDash \neg\varphi$.

   (b) In our system each acceptable rule of inference has a corresponding correct semantic sequent (for example, the sequent $\varphi, (\varphi \rightarrow \psi) \vDash \psi$ corresponds to *modus ponens*). Consider now this rule of inference in English: if φ is a truth of logic, then 'Necessarily, φ' is too. Is

this rule intuitively correct? Is the corresponding sequent '$\varphi \vDash_{\text{English}}$ Necessarily φ' intuitively correct?

2. Give the proof of Gibbard's theorem (Theorem 99). (*Hint:* begin with the fact that all instances of this schema for disjunctive syllogism are valid: $((\neg\varphi \vee \psi) \wedge \varphi) \vDash \psi$.)

3.   (a) John argues that, since φ and ψ together entail (in English) ψ, if follows that ψ entails 'if φ, ψ'. Evaluate John's argument.

  (b) Mary claims that 'if φ, ψ' entails 'if $\neg\psi$, $\neg\varphi$'. Evaluate Mary's claim.

  (c) Show how it would be possible to use John's conclusion and Mary's claim to argue that '$\varphi \rightarrow \psi$' entails 'if φ, ψ'.

  (d) Give an argument that 'if φ,ψ' is a truth-functional connective in English. Do you see any difficulties with your argument?

4. Say that $\varphi \vDash \psi$ is a *perfect* sequent iff it is correct, and φ is satisfiable, and ψ is not a tautology. Say that a sequent is *perfectible* iff it is a substitution instance of a perfect sequent.

  (a) Show that there must be a sentence letter in common between φ and ψ is a perfectible sequent. (Hint: use the Craig Interpolation Theorem (page 69), plus properties of substitution.)

  (b) Give examples to show that perfectible entailment is not transitive (i.e., that there are cases where φ perfectibly entails ψ, where ψ perfectibly entails χ, but where φ does not perfectibly entail χ.)

  (c) Which of the structural rules of §4.5 (page 48) are perfectibly inappropriate (i.e., take perfectible sequents to non-perfectible sequents.)?

5. Which of the relations expressed by the following English predicates are equivalence relations:

  (a) 'x and y attend the same lectures', on the domain of Oxford students.

  (b) 'x is studying the same subject as y', on the domain of Oxford students.

  (c) 'x entails y', on the domain of sentences of $\mathscr{L}_2$.

  (d) 'x is logically equivalent to y', on the domain of sentences of $\mathscr{L}_2$.

(e) 'If $\{x\}$ is consistent, then $\{x\} \cup \{y\}$ is consistent' on the domain of sentences of $\mathcal{L}_2$.

6. Prove the following:

(a) If R is irreflexive, then R' is reflexive.

(b) If R is symmetric, then R^{-1} is symmetric.

(c) If R is symmetric, then R' is symmetric.

(d) If R is transitive, then R^{-1} is transitive.

(e) If R is connected, then R^{-1} is connected.

(f) If R is asymmetric and non-empty, then R' is non-symmetric.

7. (a) If R is transitive, is R' transitive?

(b) If R is connected, is R' connected?

(c) If R is antisymmetric, is R' symmetric?

(d) If R is transitive and non-empty, is R' is transitive?

8. If R is defined on the empty domain, can it be reflexive? Can it be irreflexive? What about transitivity and symmetry?

9. What is wrong with the following argument that reflexivity is a consequence▐ of symmetry and transitivity?

> If $\langle x, y \rangle \in R$, then $\langle y, x \rangle \in R$ since we assume R is symmetric.
> If both $\langle x, y \rangle \in R$ and $\langle y, x \rangle \in R$, then since R is transitive,
> $\langle x, x \rangle \in R$ – so R is reflexive.

(After Partee *et al.* 1990: p. 52.)

10. (a) Can a relation be asymmetric and reflexive?

(b) Can a relation be transitive, non-symmetric and irreflexive?

(c) Can a relation be connected and irreflexive?

11. (a) Show that $\forall x \forall y (\neg x = y \rightarrow (Pxy \vee Pyx)) \equiv \forall x \forall y (x = y \vee Pxy \vee Pyx)$.

(b) A relation R satisfies *trichotomy* iff, for all x and y, at most one of these three holds: Rxy, Ryx, or $x = y$. Is every trichotomous relation connected? Under what circumstances is a connected relation trichotomous?

12. Show by suitable reasoning that in a finite domain, for any *partial order R*, $\exists x \forall y (Ryx \rightarrow x = y)$. Give a counterexample to this condition in an infinite domain.

13. Show that a well-ordering of D is a total ordering of D, but not *vice versa*.

14. Give a graph, on some non-empty domain, of a relation R which satisfies this condition: $\forall x \forall y (Rxy \rightarrow \exists z ((x \neq z) \wedge (y \neq z) \wedge Rxz))$. Can you give an example of a relation which satisfies this condition (being sure to specify the domain)?

15. Prove that

    (a) If x is a least member under an ordering R, then it is the unique least member.

    (b) The set of natural numbers $\mathbb{N}$ is well-ordered by $<$.

    (c) The set of (positive and negative) integers $\mathbb{Z}$ is *not* well-ordered by $<$.

    (d) There exists orderings with no maximal elements.

16. A relation R is *dense* iff whenever $\langle x, y \rangle \in R$, there exists a z such that $\langle x, z \rangle \in R$ and $\langle z, y \rangle \in R$. Prove that on the domain of the natural numbers $\mathbb{N}$, the greater-than relation $>$ is not dense; but that on the domain of the positive rationals (i.e., numbers of the form $\frac{n}{m}$ where $n, m \in \mathbb{N}$), it is dense.

17. Let $D = \{1, 2, 3, 5, 6, 10, 15, 30\}$. Let R be the relation on D defined by

$$R = \{\langle x, y \rangle : x \text{ divides } y \text{ without remainder}\}.$$

    (a) Show that R is a weak partial order but not a total order.

    (b) Draw the graph of R, and identify any minimal, maximal, least or greatest elements.

    (c) Do the same for the set $\wp\{a, b, c\}$, using the relation $\subseteq$ on that domain.

18. We might normally expect 'is similar to' to be a symmetric relation: after all, if there is a respect in which a is similar to b, then b must be similar to

a in that very same respect. But many people seem to judge that similarity is *not* symmmetric:

> When people are asked to make comparisons between a highly familiar object and a less familiar one, their responses reveal a systematic asymmetry: The unfamiliar object is judged as more similar to the familiar one than vice versa. For example, people who know more about the USA than about Mexico judge Mexico to be more similar to the USA than the USA is to Mexico. (Kunda, 1999: 520)

(You might think also of the fact that it is much more natural to say that children resemble their parents, than that parents resemble their children.)

Can you provide a rationale for these psychological results? Do they indicate that people are systematically mistaken about the meaning of the relational predicate 'is similar to', or do they indicate that our theory of similarity in terms of matching respects of similarity is incorrect?

Appendix A

An Alternative Proof of Compactness

n-Acceptable structures   Suppose that Γ is a finitely satisfiable set of sentences, and $P_1, P_2, \ldots$ is an enumeration of the sentence letters occurring as subsentences of sentences in Γ. Let $\mathbf{S}_n$ be the set consisting of the first n sentence letters in Γ under the enumeration, i.e., $\mathbf{S}_n = \{P_1, P_2, \ldots, P_n\}$. ($\mathbf{S}_0$ is the empty set.)

If S is a set of $\mathscr{L}_1$ sentences, we say that a structure $\mathscr{C}$ is an *S-agreeing variant* of a structure $\mathscr{B}$ iff $|\varphi|_{\mathscr{C}} = |\varphi|_{\mathscr{B}}$ for all $\varphi \in S$. Call an $\mathscr{L}_1$-structure $\mathscr{B}$ *n-acceptable* iff, for each finite subset $\Delta \subseteq \Gamma$, $\mathscr{B}$ has a $\mathbf{S}_n$-agreeing variant $\mathscr{C}$ which satisfies Δ.

Define a series of $\mathscr{L}_1$-structures as follows:

- $\mathscr{A}_0$ is some arbitrarily chosen structure.

- If $\mathscr{B}$ is an $\mathbf{S}_n$-agreeing variant of $\mathscr{A}_n$ which assigns T to P_{n+1} and $\mathscr{B}$ is $(n + 1)$-acceptable, let $\mathscr{A}_{n+1}$ be $\mathscr{B}$; otherwise let $\mathscr{A}_{n+1}$ be an $\mathbf{S}_n$-agreeing variant of $\mathscr{A}_n$ which assigns F to P_{n+1}.

Each $\mathscr{A}_n$ is n-acceptable   We now prove by induction on n that each $\mathscr{A}_n$ is n-acceptable (i.e, $\mathscr{A}_0$ is 0-acceptable, $\mathscr{A}_1$ is 1-acceptable, and so on).

Obviously $\mathscr{A}_0$ is 0-acceptable; to be 0-acceptable is for each finite subset of Γ to agree with $\mathscr{A}_0$ on the set of sentence letters in $\mathbf{S}_0$; since the latter is the empty set, *any* finite subset of Γ agrees with $\mathscr{A}_0$ on that set.

For the induction step, suppose there is some $\mathscr{A}_n$ which is n-acceptable, but $\mathscr{A}_{n+1}$ is not. Since by construction, if there was an $(n + 1)$-acceptable variant of $\mathscr{A}_n$ which assigned T to P_{n+1}, $\mathscr{A}_{n+1}$ would be such. So $\mathscr{A}_{n+1}$ must assign F to P_{n+1}. But from the induction hypothesis, some finite subset $\Delta \subset \Gamma$ is not satisfied by any $\mathbf{S}_{n+1}$-agreeing variant of $\mathscr{A}_{n+1}$; so Δ cannot be satisfied by *any* $\mathbf{S}_n$-agreeing variant of $\mathscr{A}_n$ (since it can't be satisfied by a variant which assigns T to P_{n+1}, or one that assigns F, and they are the only options), contrary to the assumption that $\mathscr{A}_n$ is n-acceptable. So $\mathscr{A}_{n+1}$ must in fact be $(n + 1)$-acceptable.

Compactness Again   If we let $\mathscr{A}$ be a structure such that, for all i, $|P_i|_{\mathscr{A}} = |P_i|_{\mathscr{A}_i}$ (that is, for all i, $\mathscr{A}$ agrees with the i-th structure $\mathscr{A}_i$ on the value it assigns to the i-th sentence letter P_i).

Suppose $\mathscr{A}$ isn't a model for Γ. Then there is a $\varphi \in \Gamma$ such that $|\varphi|_{\mathscr{A}} = F$. Let P_k be the highest numbered sentence letter occurring in φ. Then $\mathscr{A}_k$ must make φ true, since $\{\varphi\}$ is a finite subset of Γ and $\mathscr{A}_k$ is k-acceptable. But then every $\mathbf{S}_k$-agreeing variant of $\mathscr{A}_k$ makes φ true; and $\mathscr{A}$ is one such. Contradiction – $\mathscr{A}$ must satisfy Γ.

But since Γ was arbitrary, we've shown that if Γ is finitely satisfiable, then Γ is satisfiable. This is compactness.

Appendix B

Greek letters

α, A	Alpha	ν, N	Nu
β, B	Beta	ξ, Ξ	Xi
γ, Γ	Gamma	o, O	Omicron
δ, Δ	Delta	π, Π	Pi
ε, E	Epsilon	ρ, P	Rho
ζ, Z	Zeta	σ, Σ	Sigma
η, H	Eta	τ, T	Tau
θ, Θ	Theta	υ, Y	Upsilon
ι, I	Iota	φ, Φ	Phi
κ, K	Kappa	χ, X	Chi
λ, Λ	Lambda	ψ, Ψ	Psi
μ, M	Mu	ω, Ω	Omega

Table B.1: Table of Greek letters

Appendix C

List of Definitions

Bibliography

Beall, JC and van Fraassen, Bas C (2003), *Possibilities and Paradox*. Oxford: Oxford University Press.

—— and Restall, Greg, 'Logical Consequence'. In Edward N Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL plato.stanford.edu/entries/logical-consequence/.

Belnap, Jr., Nuel D (1977), 'A Useful Four-Valued Logic'. In J Michael Dunn and G Epstein (eds.), *Modern Uses of Multiple-Valued Logic*, Dordrecht: D Reidel, pp. 8–37.

Bennett, Jonathan (2003), *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Boolos, George S, Burgess, John P, and Jeffrey, Richard C (2007), *Computability and Logic*. Cambridge: Cambridge University Press, 5 ed.

Boolos, George (1971) 'The Iterative Conception of Set', *Journal of Philosophy* **68**: 215–232.

Borkowski, L and Słupecki, J (1958), 'The Logical Works of J Łukasiewicz'. *Studia Logica*, vol. 8: pp. 7–56.

Bostock, David (1997), *Intermediate Logic*. Oxford: Oxford University Press.

Burgess, John P (2009), *Philosophical Logic*. Princeton: Princeton University Press.

Caplan, Ben (2006), 'Empty names'. In Robert Stainton and Alex Barber (eds.), *The Encyclopedia of Language and Linguistics: Philosophy and Language*, vol. 4, Oxford: Elsevier, pp. 132–136.

Craig, William (1957), 'Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory'. *Journal of Symbolic Logic*, vol. 22: pp. 269–85.

Davis, Wayne (2014), 'Implicature'. In Edward N Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.

URL plato.stanford.edu/entries/
implicature/.

Dummett, Michael (1983), 'The Philosophical Basis of Intuitionist Logic'. In Paul Benacerraf and Hilary Putnam (eds.), *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2 ed., pp. 97–129.

Edgington, Dorothy (2014), 'Indicative Conditionals'. In Edward N Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL plato.stanford.edu/entries/conditionals/.

Gentzen, Gerhard (1969), 'Investigations into Logical Deduction'. In M E Szabo (ed.), *The Collected Papers of Gerhard Gentzen*, Amsterdam: North-Holland, pp. 68–131.

Gibbard, Allen (1981), 'Two Recent Theories of Conditionals'. In W L Harper, *et al.*, (eds.), *Ifs*, Dordrecht: D. Reidel, pp. 211–47.

Gowers, Timothy (ed.) (2008), *The Princeton Companion to Mathematics*. Princeton, NJ: Princeton University Press.

Grice, Paul (1989), 'Logic and Conversation'. In *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.

Halbach, Volker (2010), *The Logic Manual*. Oxford: Oxford University Press.

Harris, John H (1982), 'What's so logical about the 'logical' axioms?' *Studia Logica*, vol. 41: pp. 159–71.

Henkin, Leon (1949), 'The Completeness of the First-Order Functional Calculus'. *Journal of Symbolic Logic*, vol. 14: pp. 159–66.

Henle, James M, Garfield, Jay L, and Tymoczko, Thomas (2011), *Sweet Reason*. Chichester: Wiley-Blackwell, 2nd ed.

Heyting, A (1956), *Intuitionism: An Introduction*. Amsterdam: North-Holland.

Iemhoff, Rosalie (2015), 'Intuitionism in the Philosophy of Mathematics'. In Edward N Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. plato.stanford.edu/entries/logic-intuitionistic/.

Jeffrey, Richard C (2006), *Formal Logic: Its Scope and Limits*. Indianapolis: Hackett, 4th ed, with a supplement by John P Burgess.

Kratzer, Angelika (2012) 'Conditionals'. In her *Modals and Conditionals*, Oxford: Oxford University Press, pp. 85–108.

Kripke, Saul (1976), 'Is There a Problem About Substitutional Quantification?' In Gareth Evans and John McDowell (eds.), *Truth and Meaning*, Oxford: Oxford University Press, pp. 324–419.

—— (1980), *Naming and Necessity*. Cambridge, MA: Harvard University Press.

—— (2013), *Reference and Existence*. Oxford: Oxford University Press.

Kunda, Ziva (1999), *Social Cognition: Making Sense of People*. Cambridge, MA: MIT Press.

Lambert, Karel (2001), 'Free Logics'. In Lou Goble (ed.), *The Blackwell Guide to Philosophical Logic*, Oxford: Blackwell.

Lewis, David (1973), *Counterfactuals*. Oxford: Blackwell.

—— (1984), 'Putnam's Paradox'. *Australasian Journal of Philosophy*, vol. 62: pp. 221–36.

Ludlow, Peter (2013), 'Descriptions'. In Edward N Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL plato.stanford.edu/entries/descriptions/.

Machover, Moshé (1996), *Set Theory, Logic, and Their Limitations*.

Cambridge: Cambridge University Press.

McGee, Vann (1985), 'A Counterexample to *Modus Ponens*'. *The Journal of Philosophy*, vol. 82: pp. 462–71.

—— (2006), 'There's a Rule for Everything'. In Agustín Rayo and Gabriel Uzquiano (eds.), *Absolute Generality*, Oxford: Oxford University Press, pp. 179–202.

Partee, Barbara H, ter Meulen, Alice and Wall, Robert E (1990), *Mathematical Methods in Linguistics*. Dordrecht: Kluwer.

Potter, Michael (2004), *Set Theory and Its Philosophy*. Oxford: Oxford University Press.

Prawitz, Dag (2006), *Natural Deduction: A Proof-theoretic study*. New York: Dover.

Priest, Graham (2006), *In Contradiction*. Oxford: Oxford University Press, 2nd ed.

Prior, A N (1960), 'The Runabout Inference-Ticket'. *Analysis*, vol. 21: pp. 38–9.

Pullum, Geoffrey K (2008), 'Scooping the Loop Snooper: a proof that the Halting Problem is unsolvable'.

URL www.ling.ed.ac.uk/~gpullum/loopsnoop.pdf.

Putnam, Hilary (1980), 'Models and Reality'. *Journal of Symbolic Logic*, vol. 45: pp. 464–82.

Quine, W V O (1940) *Mathematical Logic*. Boston, MA: Harvard University Press.

Restall, Greg (2000) *An Introduction to Substructural Logics*. London: Routledge.

Richard, Mark (1986) 'Quotation, grammar, and opacity', *Linguistics and Philosophy* **9**: 383–403.

Russell, Bertrand (1956), 'On Denoting'. In *Logic and Knowledge*, London and New York: Routledge, pp. 41–56.

Shapiro, Stewart (2000) *Thinking About Mathematics*. Oxford: Oxford University Press.

—— (2013), 'Classical Logic'. In Edward N Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL plato.stanford.edu/entries/logic-classical/.

Stalnaker, Robert C (1975), 'Indicative Conditionals'. *Philosophia*, vol. : pp. 269–86.

Strawson, P F (1950), 'On Referring'. *Mind*, vol. 59: pp. 320–44.

Tarski, Alfred (1933), 'The Concept of Truth in Formalized Languages'. In *Logic, Semantics, Metamathematics*, Indianapolis: Hackett, 2nd ed., 1983, pp. 152–278.

—— (1936), 'On the Concept of Logical Consequence'. In *Logic, Semantics, Metamathematics*, Indianapolis: Hackett, 2nd ed., 1983, pp. 409–420.

Taylor, Barry (2006), *Models, Truth, and Realism*. Oxford: Oxford University Press.

Thomson, J F (1990) 'In Defense of ⊃'. *The Journal of Philosophy*, vol. 87: pp. 57–70.

Turing, A M (1937), 'On Computable Numbers, with an Application to the Entscheidungsproblem'. *Proceedings of the London Mathematical Society*, vol. 42: pp. 230–65.