Logical Forms

An Introduction to Philosophical Logic

SECOND EDITION

Mark Sainsbury

BLACKWELL
Publishers

Logical Forms

An Introduction to Philosophical Logic

SECOND EDITION

Mark Sainsbury

BLACKWELL
Publishers

First published 1991
Reprinted 1993, 1997, 1998, 1999

Second edition published 2001

2 4 6 8 10 9 7 5 3 1

Blackwell Publishers Ltd
108 Cowley Road
Oxford OX4 1JF
UK

Blackwell Publishers Inc.
350 Main Street
Malden, Massachusetts 02148
USA

British Library Cataloguing in Publication Data
A CIP catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data has been applied for

ISBN 0-631-21678-2 (hardback); 0-631-21679-0 (paperback)

Typeset in 11 on 13 pt Bembo
by Best-set Typesetter Ltd., Hong Kong
Printed in Great Britain by Biddles Ltd, Guildford, Surrey

This book is printed on acid-free paper.

Contents

Preface to the first edition

An early draft of the first three chapters of this book circulated in manuscript in 1980: I am very grateful to the many people – colleagues, friends and students – who commented upon that draft.

Most of the final version of the first edition was written in Belize, Central America, in the summer months of 1987. I am grateful to King's College London for making this possible by a grant of sabbatical leave; and to the Government of Belize for kindly granting me exemption from import duty on generating equipment required to run my word processor.

I would like to thank Danial Bonevac, who read the entire manuscript in a near-final draft, and whose comments led to many improvements; Marianne Talbot, who provided invaluable help with references; and Stephen Read, whose very penetrating and knowledgeable criticisms of what had been intended as the final draft led to some considerable rewriting, and saved me from many mistakes.

Preface to the second edition

The main changes in this edition relate to chapter **3** (on conditionals, retitled "Conditionals and probabilities"), chapter **4.19** (retitled "Predicate quantifiers and second order logic"), chapter **4.20** (retitled "Free logics"), chapter **5.2** (retitled "Non-indicative and counterfactual conditionals") and **5.9** (Counterpart theory). All this material has been rewritten from scratch. Elsewhere I have made many changes (probably in most of the sentences) to remove blemishes of style and content. The structure of the book remains much the same (the only significant changes to the table of contents concern chapter 3), which I hope will encourage those instructors who have used it in their classes to continue to do so.

The layout of the book has been greatly improved: exercises and footnotes are now on the page with the related text. For this, and for allowing me a second edition, I would like to thank Blackwell Publishers.

In this edition I have not starred sections as optional. A beginner wanting the shortest possible route through the topic might read chapters **1**, **2**, **4.1–4.4**, **5.1–5.5**, **5.8–5.9** and **6**.

I am very grateful to Dorothy Edgington and Scott Sturgeon who gave me detailed comments on the new material on conditionals. I wish I had been able to do better justice to them. I would also like to thank others who have pointed out mistakes small or large, in person or in print, in particular: Roderick Batchelor, Stewart Candlish, Max Kölbel, David Lewis, Stephen Neale, Alex Oliver, Thomas Patton and Yannis Stephanou. To King's College London I owe thanks for a semester's sabbatical leave, Autumn 1999.

Mark Sainsbury
London

Introduction

Some kind of knowledge of logical forms, though with most people it is not explicit, is involved in all understanding of discourse. It is the business of philosophical logic to extract this knowledge from its concrete integuments, and to render it explicit and pure.

Bertrand Russell, *Our Knowledge of the External World*

This book is an introduction to philosophical logic. It is primarily intended for people who have some acquaintance with deductive methods in elementary formal logic, but who have yet to study associated philosophical problems. However, I do not presuppose knowledge of deductive methods, so the book could be used as a way of embarking on philosophical logic from scratch.

Russell coined the phrase "philosophical logic" to describe a programme in philosophy: that of tackling philosophical problems by formalizing problematic sentences in what appeared to Russell to be *the* language of logic: the formal language of *Principia Mathematica*. My use of the term "philosophical logic" is close to Russell's. Most of this book is devoted to discussions of problems of formalizing English in formal logical languages.

I take validity to be the central concept in logic. In the first chapter I raise the question of why logicians study this property in connection with artificial languages, which no one speaks, rather than in connection with some natural language like English. In chapters 2–5 I indicate some of the possibilities and problems for formalizing English in three artificial logical languages: that of propositional logic (chapter 2), of first order quantificational logic (chapter 4) and of modal logic (chapter 5). The final chapter takes up the purely philosophical discussion, and, using what has been learned on the way, addresses such questions as whether there was any point in those efforts at formalizing, what can be meant by *the* logical form of an English sentence, what is the domain of logic, and what is a logical constant.

In this approach, one inevitably encounters not only questions in the philosophy of logic, but also questions in the philosophy of language, as when one considers how best to formalize English sentences containing empty names, or definite descriptions, or adverbs, or verbs of propositional attitude.

My own preference in teaching logic is to begin with the elementary formal part, keeping the students' eyes blinkered to philosophical questions, which are dauntingly hard. In introducing the philosophical issues later on, I am conscious of the width of the gap I expect students to leap, from the drill of truth tables and proofs, to discussions of the semantics of names or conditionals. This is the gap this book is designed to fill.

Logic raises a host of problems that call for philosophical discussion, like the nature of truth, the relation between logical rules and psychological processes, the nature of logical knowledge, the question of what exists. I have turned aside from as many of these as I could, limiting myself to the single theme of the nature of logical form, together with whatever tributaries I seemed absolutely compelled to navigate.

There are exercises throughout, and these are required for completeness. Various issues that could well be regarded as essential to the main business are relegated to exercises, so that the reader can, in effect, contribute to the development of the argument. For example, expert readers of the section on Russell's theory of descriptions might be aghast to find that the body of the text makes no mention of Russell's claim that "The present King of France is not bald" is ambiguous, and in a way which resolves a certain puzzle. The reader is invited to discover the ambiguity, and its relation to the puzzle, in Ex. 4.18. There are bibliographical notes at the end of each chapter which suggest further reading and provide some information about the sources of some of the points in the text.

The reader will find it useful to look at the section which immediately follows ("Organization"), which explains the system of numbering displayed material, the arrangement of glossary entries and related matters.

The book is progressive. The first chapter is written with the complete novice in mind; the last chapter addresses a considerably more knowledgeable audience. My hope is that the intervening material will help beginners progress from the one state to the other.

Organization

Most displayed material is numbered by a bold numeral followed by a right parenthesis, thus:

1) This is how displayed material appears.

Subsequent references to displayed material use the numeral enclosed in parentheses, so the above is referred to as (**1**). Numbering begins afresh with each section, and a reference like "(**1**)" refers to the displayed sentence labelled "**1)**" within the current section.

Sections and chapters are also numbered. A reference like "(**2.3**)" refers to the displayed sentence labelled "**3)**" in the second section within the current chapter. A reference like "(**4.2.3**)" refers to the displayed sentence labelled "**3)**" in the second section of chapter **4**.

Reference to a section within the current chapter takes the form: §3. Reference to a section outside the current chapter takes the form: chapter **3.2** (that is, chapter **3**, section **2**).

Left-hand page headings give the current chapter number and title. Right-hand page headings give the current section number and title.

A full understanding of the matters discussed requires further reading. Suggestions will be found in the bibliographical notes at the end of each of chapters **2–6**. Works are referred to by the author's name followed by a date in square brackets. The bibliography (p. 406) must be used to obtain full publication details.

There is a glossary (p. 392) which lists definitions of technical terms alphabetically. Glossary entries include one or two set-theoretical terms (e.g. "sequence") that are used though not explained in the text.

There is a list of symbols (p. 403), with a brief definition of each. They are arranged in order of occurrence, with a page indicating where they were first introduced.

I have not been very particular about my use of inverted commas. In many cases, strict accuracy requires corner quotes (cf. Quine [1940], §6), but I decided to dispense with this additional complexity, in the belief that no confusion is likely to result.

1

Validity

1 What is logic about?

The philosophy of logic gives an account of what logic is, of the concepts that it uses, and of how it relates to other disciplines and to our ordinary thought and talk.

Logic is about reasons and reasoning. There are reasons for *acting*: wanting to keep thin is a reason for avoiding fatty foods. There are reasons for *believing*: that the potatoes have been boiling for twenty minutes is a reason for believing that they are ready to eat. Historically, logic has primarily concerned itself with reasons for believing.

We may give a reason for believing when we answer the question "Why does so-and-so believe such-and-such?". But such a question can be answered in two different kinds of way.

Suppose we ask of an orthodox Hindu: why does he suppose that one should not eat meat? One kind of answer is: this belief was instilled in him by his family at an early age, and has been sustained by a variety of social and personal pressures. This kind of answer may *explain* the origin of the belief. But it does not give a *reason* for the belief, in the sense of "reason" in which logic is concerned with reasons. Explanations of this kind belong to psychology or sociology. They are quite foreign to logic.

Suppose we answer the question in a different way, saying: the Hindu believes that killing, and everything which requires killing, is wrong; and that eating meat requires killing. This answer *may* explain the origin of the belief. But it also does, or purports to do, something else: it *justifies* the belief. Understood in this way, as attempting to provide a justification, the answer shows a concern with reasons in the sense in which logic can be said to be the study of *reasons*.

Logic is a *normative* discipline. It aims to say what reasons are *good* reasons. It does not merely describe the reasons that in fact move people. It lays down standards. It says what reasons *ought* to move one. Even so, the starting-point has to be what we generally think of as good reasons. Logic starts with an intuitive commonsensical and pretheoretical distinction between good and bad reasons, a distinction made by people pursuing their ordinary daily concerns. From this the logician hopes to fashion an articulate and defensible distinction between good reasons and bad. One would expect a large measure of agreement between the logician's technical distinction and the commonsensical one. But one should not turn one's back on the possibility of a divergence: common sense may need to be corrected in the light of reflection.

Here are some examples, of an everyday kind, of the commonsensical distinction at work. Most people would agree that

> James is a banker and all bankers are rich

is a good reason for

> James is rich.

By contrast, most people would agree that

> Henry is a playwright and some playwrights are poor

is not a good reason for

> Henry is poor.

There is (or until recently was) no general agreement about whether

> James and Henry lead pretty similar lives except that James is a non-smoker and Henry smokes twenty cigarettes a day

is or is not a good reason for

> Henry is more likely to die of heart disease than James.

We can regard a reason as a good reason without having to believe it ourselves. We do not have to believe that all bankers are rich to see that "James is a banker and all bankers are rich" constitutes a good reason for "James is rich". Traditionally, the logician has supposed that one can investigate whether one thing is or is not a good reason for another without having to form a view about whether the supposed reason, or what it is a reason for, is true.

When we talk about reasons, we do not have to talk about particular people and what they believe. Even if no one had ever had the beliefs we attributed to the Hindu, we could still say that

> Killing, and everything which requires killing, is wrong; eating meat requires killing

together form good reasons for

> One should not eat meat.

What one means can be partly understood in this way: if anyone were to believe that killing, and everything which requires killing, is wrong, and were also to believe that eating meat requires killing, he would thereby be right (rational, reasonable, logical, justified) also to believe that one should not eat meat.

When we want to consider something like

> Killing is wrong

in abstraction from whether anyone believes it or not, we shall call it a *proposition*. A proposition is the sort of thing that *can* be believed, or asserted, or denied, but it does not have to be: it can be disbelieved, or merely entertained, or not even thought of at all. Perhaps no one believes, or had even until just this moment supposed or entertained, the proposition that Julius Caesar built New York single-handed in a day. We shall, nevertheless, say that there is such a proposition.

The most general question which confronts the logician can now be expressed as follows: what makes one proposition (or collection of propositions) a good reason for a proposition?

We shall call the propositions offered as reasons *premises*, and the proposition which the premises are supposed to support the *conclusion*.

When some premises and a conclusion are assembled together, we shall call the result an *argument*. The technical use of these terms, as just introduced, differs in some ways from the ordinary use. In particular, as used in logic, the term "argument" does not imply any kind of disagreement or dispute.

We have already considered various arguments:

1) *Premises*: Killing, and anything which requires killing, is wrong; eating meat requires killing.
Conclusion: One should not eat meat.

2) *Premise*: James is a banker and all bankers are rich.
Conclusion: James is rich.

3) *Premise*: Henry is a playwright and some playwrights are poor.
Conclusion: Henry is poor.

4) *Premise*: James and Henry lead pretty similar lives except that James is a non-smoker and Henry smokes twenty cigarettes a day.
Conclusion: Henry is more likely than James to die of heart disease.

Common sense pronounces that in (**1**) and (**2**) the premises constitute good reasons for the conclusion, that this is definitely not so for (**3**), and that (**4**) is a debatable case.

In (**1**) and (**2**) the conclusion *follows from* the premise(s). The branch of logic with which we shall be mainly concerned – *deductive logic* – investigates the contrast between arguments in which the conclusion follows from the premises, and those in which it does not. One way in which premises can give good reason for a conclusion is for the conclusion to follow from the premises.

We ask whether an argument's premises constitute good reason for the conclusion, or whether the conclusion follows from the premises. In either case we are asking about a *relation* between two lots of propositions: on the one hand the premises, on the other hand the conclusion. In the case of arguments like (**1**) and (**2**), in which the conclusion follows from the premises, the point is not that the premises make the conclusion true, or even likely to be true. Perhaps the premises themselves are false, or likely to be false. Rather, *if* the premises were true, so would be the conclusion. That is why it would be rational to believe the conclusion if one believed the premises. Whether or not the relation obtains is something that can often, though not always, be detected *apriori*, that is, without any appeal to experience or experiment.[1] Our knowledge that the premises in (**1**) and (**2**) constitute good reasons for their respective conclusions is quite independent of knowing whether bankers are rich, or whether James is a banker (who is James, anyway?), or whether killing is wrong, or whether eating meat requires killing. Traditionally, logic has been held to be wholly apriori, and this has been used as a mark to distinguish it from other disciplines. This is a controversial view. But it would not be tenable even for a moment if the logician was called upon to pronounce on whether or not the premises of arguments are true, or probable. The view owes its attraction to the thought that the *relational* fact – whether the premises are so related to the conclusion that they constitute good reason for it – can be known apriori, even if the propositions involved in the relation cannot be known apriori.

Ex. 1.1 Suppose that Henry is indeed poor. How would you show that (**1.3**) is not a good argument, that is, that its premise does not constitute a good reason for its conclusion? Would you appeal to any facts about the actual income of playwrights, or to any further facts about Henry?

2 Inductive versus deductive logic

An old tradition has it that there are two branches of logic: deductive logic and inductive logic. More recently, the differences between these disciplines have become so marked that most people nowadays use "logic" to mean deductive logic, reserving terms like "confirmation theory" for at least some of what used to be called inductive logic. I shall follow the more recent practice, and shall construe "philosophy of logic" as "philosophy of deductive logic". In this section, I try to set out the differences between the two disciplines, and to indicate very briefly why some people think that inductive logic is not logic at all.

[1] A slightly fuller account of "apriori" is given in the glossary, which also glosses some terms which are not explained in the text.

In §**1**, we saw that one way in which an argument's premises can constitute good reasons for its conclusion is for the conclusion to follow from the premises. Let us say that any argument whose conclusion follows from its premises is *valid*. An initial test for validity is this. We ask: is it possible for the premises to be true, yet the conclusion false? In the case of (**1.3**), about poverty and playwrights, the answer is "Yes". Even if *some* playwrights are poor, it is possible that others, perhaps even the vast majority, are rich, and that Henry is among the rich ones. In general, an argument is valid just on condition that it is impossible for the premises all to be true yet the conclusion false. Could one hope to distinguish deductive from inductive logic by saying that the former, but not the latter, is concerned with validity?

Consider two arguments which occur in hundreds of text-books:

1) All men are mortal. Socrates is a man. So Socrates is mortal.

2) The sun has risen every morning so far; so (probably) it will rise tomorrow.

The first is a standard example of an argument classified as valid by deductive logic. The second is an argument which is not classified as valid by deductive logic. However, the inductive logician is supposed to accord it some reasonably favourable status. Certainly, the reasons which the premises of (**2**) give for its conclusion are better by far than those given by the same premises for the opposite conclusion:

3) The sun has risen every morning so far; so (probably) it will *not* rise tomorrow.

This may seem a silly argument, but apparently something quite like it moves some gamblers. The "Monte Carlo fallacy" consists in the belief that if there has been a long run of reds on the roulette wheel, it is more likely to come up black next time.

Ex. 1.2 What is the statistical truth, of which the Monte Carlo fallacy is a distortion? Can the truth be used to formulate a rational betting policy for roulette?

The deductive logician contrasts (**1**) and (2) by saying that the first but not the second is valid. The inductive logician will make a contrast between (2) and (**3**) – probably not by using the word "valid", but perhaps by saying that (2), unlike (3), is "inductively strong". The premises of (2), but not those of (3), provide strong reasons for the conclusion.

The premises of (1) also provide strong reasons for its conclusion. How are we to distinguish strong deductive reasons from strong inductive ones? We have a suggestion before us: the truth of the premises of a valid deductive argument makes the falsity of its conclusion impossible, but this is not so in the case of inductively strong arguments. Another way of putting this is: the reasons given by a deductively valid argument are *conclusive*: the truth of the premises guarantees the truth of the conclusion. This way of making the contrast fits (**1**) and (**2**). The truth of the premises of (**2**) may make the conclusion that the sun will rise tomorrow *probable*, but it does not guarantee it: it does not make it *certain*.

Inductive logic, as the terminology of inductive strength suggests, must be concerned with a relation which holds to a greater or lesser degree. Some non-conclusive reasons are stronger than others. So unlike deductive logic, which will make a sharp dichotomy between valid and invalid arguments, inductive logic will discern a continuum of cases, along which (**2**), perhaps, registers fairly high, whereas (**3**) registers very low indeed.

Deductive validity is, as logicians say, *monotonic*. That is, if you start with a deductively valid argument, then, no matter what you *add* to the premises, you will end up with a deductively valid argument. Inductive strength is not monotonic: adding premises to an inductively strong argument can turn it into an inductively weak one. Consider (**2**), which is supposed to be a paradigm of inductive strength. Suppose we add the premises: there is a very large meteor travelling towards us; by tonight it will have entered the solar system and will be in stable orbit around the sun; it will lie between the sun and the earth, so that the earth will be in permanent shadow. When these premises are added, the resulting argument is far from strong. (I have assumed one particular interpretation of what it is for the sun to "rise". However one interprets this phrase, it is easy enough to find premises adding which would weaken the argument.)

Much everyday reasoning is non-monotonic, and there are endless much simpler and more realistic cases than the one just given. At the

Table 1.1

	Valid deductive reasoning	*Strong inductive reasoning*
Truth of premises gives good reason for truth of conclusion	✓	✓
Truth of premises makes falsity of conclusion impossible	✓	×
Premises are conclusive reasons	✓	×
Monotonic	✓	×
Degrees of goodness of reasons	×	✓

start of the investigation, Robinson's confessing to the crime gives you a powerful reason for believing him guilty. But you may rightly change your mind about his guilt, without changing your mind about whether he confessed, when a dozen independent witnesses testify to his being a hundred miles away at the time of the crime. This is a typical case in which adding information can weaken reasons which, on their own, were strong.

Table 1.1 summarizes the differences between deductive and inductive logic that we have so far mentioned.

I said that not everyone would agree that there is any such thing as inductive logic. A famous proponent of an extreme version of this view is Karl Popper ([1959], ch. 1, §1). He has argued that the only sort of good reason is a deductiyely valid one. A consequence of his view is that there is nothing to choose between (**2**) and (**3**), considered simply as arguments: both are equally bad, being alike deductively invalid. He would therefore reject the ticks in the first and last rows of the right-hand column of table 1.1. For Popper, there is no such subject matter as the one I have tried to demarcate by the phrase "inductive logic"; no inductive argument gives a good reason; and there is no difference of degree among the goodness of "inductive reasons", all being equally bad. Accordingly, Popper sees the main activity of science not as a search for supporting evidence for hypotheses, but as an effort to weed out false hypotheses by showing by experiment that they have false deductive consequences.

A less radical sceptic about inductive logic may allow that there are good reasons which are not deductively valid, but deny that there is any systematic discipline worthy of the name "inductive logic". Reflection on the role of background knowledge in what are called inductively strong arguments, like (**2**), may ground such a scepticism. Inductive strength, as we have seen, is non-monotonic. Hence an argument cannot be assessed as inductively strong *absolutely*: some possible background information would greatly weaken the conclusion. This means that every assessment of inductive strength must be relativized to a body of background knowledge. It is far from obvious how the project of inductive logic should attempt to accommodate this point, for it is quite unclear how the background knowledge could be specified in a way which is neither question-begging (for example, saying that such-and-such an argument is inductively strong relative to any bodies of background knowledge not containing any information which would weaken the conclusion), nor quite unsystematic (for example, listing various bodies of background knowledge). There is thus a genuine (I do not say decisive) ground for doubting whether inductive logic could aspire to the kind of system and generality attained by deductive logic.

A still less radical scepticism about the possibility of inductive logic takes the form: there is such a subject matter, but it does not deserve to be called logic. Here is one reason why a person might hold this view. It may be said that anything worthy of the name of logic must be formal: the property of arguments with which it is concerned must arise wholly from the form or pattern or structure of the propositions involved. Whatever exactly "formal" means (see below, §**10**), it certainly seems to be the case that no formal question is at issue between those who do, and those who do not, think that the evidence shows that smoking increases the risk of heart disease.[2]

Another form of this kind of scepticism is as follows. Logic is apriori, but inductive "logic" is not, so it is not really logic. Consider the assessment of (**1.4**), the argument about smoking and heart disease. No doubt the interpretation of statistical evidence would be important, and *perhaps* there is an apriori discipline of statistics. But even conceding

[2] The view that inductive "logic" is not a formal discipline has been given impetus by a famous discussion by Goodman [1955], Part II. For a general introduction to problems of induction, see Skyrms [1966], esp. chs 1–3.

Table 1.2

	Deductive logic	*Inductive "logic"*
Truth of premises gives good reason for truth of conclusion	✓	?
Systematic	✓	?
Formal	✓	?
Apriori	✓	?

this, it seems at least arguable that some non-apriori material is involved. If so, if, that is, it is not a purely apriori matter whether or not some argument is inductively strong, then inductive "logic" would not be an apriori discipline, and this would make it very unlike deductive logic.

Ex. 1.3 What is likely to be at issue between one who does and one who does not think that the evidence shows that smoking increases the risk of heart disease?

Table 1.2 summarizes the various kinds of scepticism about the possibility of inductive logic.

I offer no assessment of the sceptical claims. However, from now on I shall discuss only deductive logic – logic, for short – and deductive validity – validity, for short.

3 Possibility: logical and physical

Consider the following two arguments:

1) This young tomato plant has all the moisture, nutrients, warmth and light that it needs; so it will grow good tomatoes.

2) This person is an adult male and has never married; so he is a bachelor.

Tradition has it that the first is invalid (i.e. not valid) and the second is valid. We suggested in §2 that a valid argument is one whose premises cannot be true without the conclusion being true also. But is there not a sense of "impossible" in which it is impossible for the premises of (**1**) to be true without the conclusion also being true?

Perhaps. But there is also a sense of "possible" in which this does not hold. The plant might be attacked by wireworms or destroyed by a meteorite before the tomatoes grow, even though it has all the moisture etc., that it needs.

The following two claims will help us distinguish two kinds of possibility: physical possibility and logical possibility.[3]

3) It is impossible for an internal combustion engine, used on level roads on the earth's surface, to return 5,000 mpg.

4) It is impossible for there to be a car which, at a given time, both has exactly three and exactly four wheels.

(**3**) is probably true, if the kind of impossibility involved is physical: the laws of nature being what they are, no ICE could be as efficient as that. But it is not true if the kind of impossibility involved is logical. What is logically impossible involves some kind of contradiction or inconsistency, as illustrated in (**4**). Logical impossibility typically issues from the very nature of the concepts involved, and is not beholden to the laws of nature. It is logically possible for the laws of nature to be very different from what they actually are.

Ex. 1.4 (a) Why should a difference in natural law not be a physical, rather than a logical, possibility?

(b) Give three examples, not in the text, of states of affairs which are logically possible but physically impossible.

(c) Give three examples, not in the text, of states of affairs which are logically impossible.

(d) Are any states of affairs logically impossible but physically possible?

(e) Are the following logically impossible, physically impossible, or neither? ("Don't know" may be the best response in some cases!):

[3] For a good discussion, see Plantinga [1974]. By logical possibility I mean approximately what he calls "broadly logical possibility" (p. 3). (For a qualification, see chapter **6.5** below.)

(i) Mr Stamina has at last perfected his perpetual motion machine.
(ii) Mary has precisely twice as many children as Jane: she has the twin girls and little William.
(iii) Jock ("the Flash") McVite ran the mile in under 3.5 minutes.
(iv) The notorious swindler, Siva Malgavany, who died in 1880 in the arms of one of his many mistresses, was reincarnated in 1997 as a small Irish terrier, owned by Mrs Fortescue-Brown of Egham, Surrey.
(v) On 15 March 1998, Dr Chronowski stepped into his time machine, and stepped out again to find himself present at the battle of Waterloo.

A definition of validity needs to draw upon the notion of logical rather than physical possibility, if it is to give a correct account of the logician's usage. Consider the following example:

5) This creature has the form of a finch; so it will not discourse intelligently about Virginia Woolf.

As the word "valid" is standardly used in logic, this is not a valid argument. This classification might be challenged: intelligent discourse requires suitable musculature and thorax, and suitable complexity of brain; but it is *impossible* that a creature having the form of a finch should have such a thorax etc., and a sufficiently large brain. So it is impossible for the premise of the argument to be true without the conclusion also being true. So the argument is valid.

This objection uses the notion of physical rather than logical possibility. The laws of nature that we actually have rule out there being a brain of sufficient complexity for discourse in a finch-sized skull. So it is physically impossible for the premise of (**5**) to be true yet its conclusion false. But it is not logically impossible. We *might* have had different laws of nature. There is no *logical* guarantee that discourse requires a larger than finch-sized brain. So it is logically possible for the premise of (**5**) to be true yet the conclusion false. So the argument is not valid.

We can now set out our preliminary definition of validity:

Ex. 1.5 (a) Is (**3.5**) nevertheless inductively strong? Compare its inductive strength with that of (**3.1**), justifying your comparison.
(b) Would it make a difference to the validity of the argument in (**3.5**) if the first premise were "This creature *is* a finch"?

An argument is VALID if and only if it is logically impossible for all the premises to be true yet the conclusion false.

Ex. 1.6 (a) In each of the following cases, say whether the argument is deductively valid, inductively strong, or neither. Justify your answers:
(i) I told you to add more pepper if you thought the stew was too bland. You added more pepper even though you didn't think the stew was too bland. So you disobeyed me.
(ii) You must know how to make a sling, if you are really a qualified nurse. But evidently you don't know how to make a sling. So you are not a qualified nurse.
(iii) He has coughed up arterial blood, so his lungs must be extensively damaged.
(iv) My father was born in London, yours in Paris. So we can't be brothers.
(v) This plant grew from a seed produced by a tomato plant. So it, too, is a tomato plant.
(vi) Frank Whittle died in poverty. Therefore, the inventor of the jet engine died in poverty. [Historical note: Frank Whittle invented the jet engine.]
(vii) The satellite picture shows a cold front sweeping in from the Atlantic, and it should reach our shores by morning. So it will probably rain tomorrow.
(viii) The explanation of any fact is in terms of other facts. So if any facts can be explained, there are facts which cannot be explained.
(ix) If you know anything, you can't be mistaken. So any proposition that could be false is one which you do not know.
(x) The proposition "All propositions are false" is false. For if it were true it would be false. [In ordinary English, the conclusion of an argument may be presented before the premises.]
(xi) Geraniums are not frost hardy. There are frequent frosts in Iceland. So geraniums are not native to Iceland.
(xii) If either the CIA agent or the KGB agent killed the President, he used a gun. But the President died of cyanide poisoning. So neither the CIA agent nor the KGB agent killed him.

(b) Think of a stack of four cards with numbers on one side and letters on the other. Someone tells you that all cards which have an even number on one side have an "E" on the other. Say which of the following cards you need to turn over in order to tell whether what he said is true:

6	3	E	A

Explain how this example relates to the validity of various arguments involving what the person said as one of its premises.

(c) We can use cards in the manner of the previous example to specify situations. Thus each card might represent a person in a bar, one side indicating whether he is an adult or a minor, the other side indicating whether he is drinking alcohol or a soft drink. Suppose there are again four cards having visible sides as follows:

Alcohol	Soft	Adult	Minor

Which cards do you need to turn over in order to check whether the bar is violating the rule that only adults may drink alcohol? Explain how your answer relates to the validity of arguments having as a premise "Only adults drink alcohol".

This definition has some merits. For one thing, it suggests an answer to why we should use valid arguments: valid arguments are necessarily truth-preserving. So long as you start out with truth, you will never depart from the truth if you keep to valid arguments. Moreover, it is rather surprising how much can be extracted about the nature of validity from even this preliminary definition (see §**6**).

However, the definition has many defects. We characterized logic as the study of validity. But now, in defining validity, we have used the notion of *logical* impossibility. If we fully understood what *logical* impossibility is, presumably we would already know what *logic* itself is. So our characterizations have run in a circle.

We mentioned a connection between logical impossibility and inconsistency and contradictoriness. But these terms themselves were left unexplained.

This unsatisfactory state of affairs will have to persist for some time. One feature of definitions of validity for the formal languages to be

considered in later chapters is that they can entirely avoid such notions as logical possibility and inconsistency (in the ordinary sense). For the moment, we shall see how far the ordinary notions can take us.

4 Validity, inconsistency and negation

A collection of propositions is *inconsistent* if and only if it is logically impossible for all of them to be true. Here logical impossibility is used to explain inconsistency, whereas in §**3** inconsistency was used to explain logical impossibility. This shows that the two notions are closely related. It also shows that we could reasonably hope for a further elucidation of both notions, one which takes us out of the circle. For the moment, we shall simply take for granted the notion of logical impossibility.

Consider the propositions:

1) The earth is spheroid.

2) The earth is not spheroid.

It is logically impossible for both these propositions to be true and logically impossible for both of them to be false. In short, (**1**) and (**2**) are *contradictories*: each is a contradictory of the other. Moreover, (**2**) is the *negation* of (**1**). You get the negation of a proposition if you insert "not" (or some equivalent expression) into it in such a way as to form a contradictory of it.

Ex. 1.7 Without using "not" (or any equivalent word or phrase) form a contradictory of each of the following:

(i) Tom has no children.

(ii) Richard Nixon died in 1981.

(iii) London is north of Paris or south of Paris.

Being the negation of a proposition is one way, but not the only way, of being a contradictory of it. Being contradictories is one way, but not the only way, for two propositions to be *inconsistent*. I shall amplify these points, and then connect the notions of contradictory, inconsistency and negation with that of validity.

If one proposition is the negation of another, if follows trivially from the definition that the two propositions are contradictories. The converse does not hold. Two propositions can be contradictories without either being the negation of the other. For example:

3) John is more than six feet tall

and

4) John is either exactly six feet tall or else less than six feet tall

are contradictories, but neither is the negation of the other. Negation is one way, but not the only way, of forming a contradictory.

Inserting "not" into a proposition does not always yield the negation of it, for inserting "not" does not always yield a contradictory. Consider:

5) Some men are happy.

6) Some men are not happy.

The second results from the first by inserting a "not", but the two propositions are not contradictories, since both could be – and presumably actually are – true. So **(6)** is not the negation of **(5)**.

Similarly,

7) Reagan believes that Shakespeare was a genius

8) Reagan believes that Shakespeare was not a genius

are not contradictories, since both could be false. They would be false if Reagan had no view one way or the other about Shakespeare's qualities. Hence **(8)** is not the negation of **(7)**.

Ex. 1.8 Which of the following pairs consists of a proposition and its negation?

(i) I hope you will come.
I hope you will not come.

(ii) I hope you will come.
I do not hope you will come.

(iii) I am pleased with your progress.
I am not at all pleased with your progress.

(iv) Everyone who goes to see Mick the Fix is satisfied with his handling of their problem.
Everyone who goes to see Mick the Fix is not at all satisfied with his handling of their problem.

(v) You must not walk on the grass.
You must walk on the grass.

Any collection of propositions containing a contradictory pair is inconsistent. It is impossible for both of two contradictory propositions to be true, so it is impossible for all the propositions in a collection containing a contradictory pair to be true. The converse does not hold: there are inconsistent collections containing no contradictory pair. For example:

9) John is over six feet tall. John is under six feet tall

is an inconsistent collection, for it cannot be that both propositions are true. Since both could be false (and would be, if John were exactly six feet tall), they are not contradictories.

Ex. 1.9 Give three examples (not in the text) of collections of propositions which are inconsistent, but where no member of the collection is a contradictory of any other member.

Figure 1.1 summarizes the relationships mentioned. All pairs of propositions of which one is a negation of the other are contradictories, and all contradictories are inconsistent. However, there are inconsistencies which are not contradictories, and contradictories of which neither is a negation of the other.

Part of the link between validity and inconsistency, mediated by the notion of contradictoriness, consists in the following fact:

10) If an argument is valid, a collection of propositions consisting of its premises together with a contradictory of its conclusion is inconsistent.

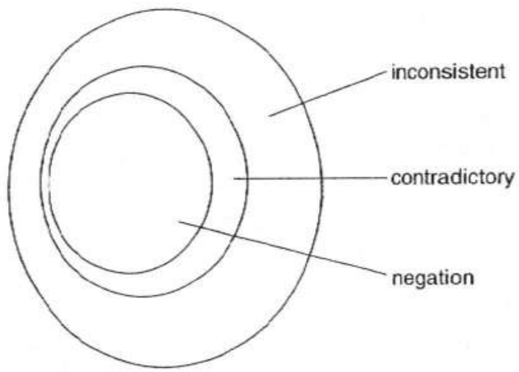

Figure 1.1

To illustrate, consider the following argument:

11) Anyone who drives a car risks death. Anyone who risks death is brave. So all car drivers are brave.

Ex. 1.10 Is the argument in (**4.11**) valid? Is the conclusion true?

The following collection contains the premises of (**11**) as (i) and (ii) and the negation of its conclusion as (iii):

12) (i) Anyone who drives a car risks death.
(ii) Anyone who risks death is brave.
(iii) Not all car drivers are brave.

Take any two of these propositions. We can see that, if these two are true, then the third cannot be. So the collection is inconsistent.

We can argue quite generally for (**10**), using the definition of validity given in §**3**. Take any valid argument. By definition, it is logically impossible for its premises to be true yet its conclusion false. In other words, it is logically necessary that if all the premises are true, so is the conclusion. But if the conclusion is true, then, necessarily, any contradictory of it is false. So, necessarily, if the premises are true, any contradictory of the conclusion is false. So it is logically impossible for the premises and a contradictory of the conclusion all to be true. So this collection is inconsistent.

The link between validity and inconsistency also runs in the other direction:

13) If a collection of propositions is inconsistent, any argument whose premises consist of all but one of the collection, and whose conclusion is a contradictory of the remaining proposition, is valid.

The argument for this is rather like the one just given for (**10**). Taking (**10**) and (**13**) together shows that we could have defined validity in terms of inconsistency, rather than in terms of logical impossibility.

Ex. 1.11 Provide an argument which justifies (**4.13**).

Ex. 1.12 Give a definition of validity in terms of inconsistency.

5 Arguments and argument-claims

We use "argument" to refer to any collection of propositions, one of which is singled out as the conclusion. It is useful to have a standard pattern for writing out arguments. We adopt the convention that where there is no contrary indication, the conclusion of an argument is the last proposition in a list, and is marked off from its predecessors by being preceded by a semicolon. Thus if an argument has two premises, *A* and *B*, and a conclusion, *C*, we shall write it:

A, B; C.

More generally, where the argument has *n* premises, and its conclusion is *C*, we write:

$A_1, \ldots A_n; C.$

(n may be equal to or greater than zero. For the case in which $n = 0$ (there are no premises) see §**6**.)

A useful abbreviation is "$\models$", short for "is valid". It works like this:

1) $A_1, \ldots A_n \models C$ abbreviates "$A_1, \ldots A_n; C$" is valid.
$A_1, \ldots A_n \not\models C$ abbreviates "$A_1, \ldots A_n; C$" is not valid.

The symbol "$\models$" is pronounced "(double) *turnstile*".

An argument ($A_1, \ldots A_n; C$) must be distinguished from what I shall call an *argument-claim*: $A_1, \ldots A_n \models C$, or $A_1, \ldots A_n \not\models C$. The component propositions in an argument are true or false, but the argument itself cannot significantly be said to be true or false. One correct dimension of assessment for an argument is whether it is valid or not; another is whether it is persuasive or not; but truth and falsehood do not provide a proper dimension of assessment. By contrast, an argument-claim *is* true or false: true if it is a positive argument-claim ($\models$) and the argument in question is valid, or if it is a negative argument-claim ($\not\models$) and the argument in question is not valid; and otherwise false.

In an argument-claim, "$\models$" appears in the very place in which, in an expression of the argument in ordinary English, one would find a word like "so", "therefore" or "hence": a word used to show that one has come to the conclusion which is being drawn from the previous propositions. This gives rise to a tendency to confuse the role of "$\models$" with that of conjunctions like "therefore". But the roles are really very different.

First, "$\models$" and "therefore" belong to different grammatical categories. "$\models$" is a predicate, the sort of expression which can be used to attribute a property to something. "Therefore" is not a predicate, but rather a word used to join sentences together. To see the force of this point, consider the fact that we can meaningfully say "Some arguments are valid but others are not" (bearing in mind that "$\models$" abbreviates "is valid"), though we cannot meaningfully say "Some arguments are therefore, but some are not".

Secondly, something of the form "$A_1, \ldots A_n$, therefore C" is an argument, about which the question arises whether or not it is valid. By contrast, something like "$A_1, \ldots A_n \models C$" is not itself an argument, but rather a claim *about* an argument, the claim that it is valid.

Thirdly, in ordinary circumstances, one who propounds an argument, $A_1, \ldots A_n$, therefore C, is thereby committing himself to the truth of all of $A_1, \ldots A_n$. But one who makes the claim that $A_1, \ldots A_n \models C$ makes no such commitment, since there are valid arguments whose premises are not true.[4]

Ex. 1.13 (a) What is the relation between ";", as used here in setting out arguments, and "hence" or "therefore"?

(b) For each of the following pieces of discourse, state what argument, if any, you think the speaker is intending to propound. If you think he is intending to propound one, say what its premises and conclusion are:

(i) My opinion, for what it's worth, is that you should take to robbing banks. This is a demanding career, requiring a high degree of responsibility, but offering greater than average financial rewards. Moreover, it gives you the opportunity to be your own boss and to develop your talents of leadership and initiative. If all else fails, it provides you with free board, lodging and protection at the expense of Her Majesty's Government.

(ii) If John and Mary are godparents of the same child, as indeed they are, they cannot be married according to Christian rites.

(iii) Since John and Mary are godparents of the same child, they cannot be married according to Christian rites.

(iv) John and Mary are godparents of the same child. Therefore, they cannot be married according to Christian rites.

(v) Treason doth never prosper. What's the reason?
Why, if it doth, then none dare call it treason.

(vi) No contemporary politician understands the French as well as de Gaulle did. Mitterrand is at home with the intellectual Left, and d'Estaing knows the aristocracy and the *haute bourgeoisie*. But de Gaulle knew something about the traditional peasant class, which is what sets him apart from our contemporaries.

(vii) Authors of sentimental novels were not being so ridiculous as they appeared to later generations when they described so many abductions, ladders, musketeers and escapes from convents. All these exciting events actually took place, and frequently, too. It was part of the current vogue for aggressive Spanish manners.

(viii) It was late and I was tired. So I took a taxi.

[4] The contrast between whether a speaker is propounding an argument or not, and thus the contrast between whether he has said something appropriately assessed for validity or not, is not as clear as the text would suggest. This should be apparent from reflection on Ex. 1.13. Cf. also van Dijk [1977].

6 Some important properties of validity

Although our definition of validity in §3 is not as illuminating as one might wish, it none the less enables us to discover some important general features of validity.

The key property of validity is that it logically guarantees the preservation of truth. If you start with truth and argue validly then you are bound to end up with truth. That is why it is a good thing to argue validly. But validity does not always generate truth (see **1)**), nor does truth always generate validity (see **3)**).

Ex. 1.14 (a) Are there any valid arguments with false premises and a true conclusion? If so, give an example. (Here and elsewhere, select examples of propositions which are well known to be true, or false, as the case may be.)

(b) Are there any valid arguments with true premises and a false conclusion? If so, give an example. If not, say why not.

(c) If an argument is invalid and at least one of its premises is false, what can be inferred about whether or not its conclusion is false?

(d) If an argument is valid and has a false conclusion, what can be inferred about whether or not its premises are true?

1) There are valid arguments with false conclusions.

Example:

2) All heavenly bodies revolve around the earth. The sun is a heavenly body. Therefore the sun revolves around the earth.

Moreover:

3) There are invalid arguments with true premises and true conclusions.

Example:

4) Petroleum can be used as a fuel. More people live in Paris than in Boston. Therefore, the first man on the moon was an American.

We have already seen that deductive validity is monotonic. Using the notation of §5, this can be expressed:

5) If $[A_1, \ldots A_n \models C]$,
then $[A_1, \ldots A_n, B \models C]$, whatever B may be.[5]

In other words, you cannot turn a valid argument into an invalid one by adding to the premises. This elaborates what is meant by saying that deductive logic aims to pick out arguments in which the premises give *conclusive* reasons for the conclusion.

Another important property of validity, as classically conceived, and as defined in §3, is a kind of *transitivity*. Chaining arguments together will preserve validity:

6) If $[A_1, \ldots A_n \models C]$ and $[B_1, \ldots B_k, C \models D]$,
then $[A_1, \ldots A_n, B_1, \ldots B_k \models D]$.

The intermediate conclusion, C, can be cut out, since the premises which establish it can establish anything it can establish.

Ex. 1.15 Using the displayed definition of validity near the end of §3, show that **(5.6)** is true.

Validity has a property akin to *reflexivity*:

7) If C is among the $A_1, \ldots A_n$, then $[A_1, \ldots A_n \models C]$.

Ex. 1.16 Using the displayed definition of validity near the end of §3, show that **(5.7)** is true.

This shows that circular arguments are valid. (Of course, they are not normally *useful*: see **(7.4)** below.)

[5] Square brackets [,], are used for greater legibility.

A new piece of terminology: we shall express the claim that a collection of propositions $(A_1, \ldots A_n)$ is inconsistent by writing:

$(A_1, \ldots A_n) \models$.

The terminology is justified by the fact that if an argument's premises are inconsistent, it is valid; and this is suggested by the blank after the turnstile, accepting any completion. More formally:

8) If $[(A_1, \ldots A_n) \models]$, then $[A_1, \ldots A_n \models B]$, whatever B may be.

Ex. 1.17 Use this notation to express the claim: "If a collection of propositions is inconsistent, it remains inconsistent whatever proposition is added to it." Now argue for the truth of this claim.

Like all the other properties of validity described in this section, this one follows from the definition given in §**3**. If premises are inconsistent, they cannot all be true. If premises cannot all be true, then, in particular, the following case cannot arise: that all the premises are true, and also some arbitrary proposition, B, is false. So it is impossible for all the premises to be true yet B false. So an argument with inconsistent premises is valid, whatever its conclusion.

(**8**) should not be read as saying that you can infer anything from an inconsistency. As we normally speak of inference, one can *infer* something from one or more premises only if those premises are true. A detective can draw inferences, correct or incorrect, from the footprints, but if there were no footprints, and he were merely hallucinating, we wouldn't normally allow that he could no any inferring, because we would not normally allow that there was anything from which to infer. Given this conception of inference, (**8**) does not license inferences from inconsistencies, for no inconsistencies are true.

A further piece of terminology. Let us write

$\models A$

to abbreviate: "it is logically impossible for A to be false". The terminology is justified by the fact that if an argument's conclusion cannot

be false, then it is valid; and this is suggested by the blank before the turnstile, accepting any completion. This claim can be expressed as follows:

9) If $[\models A]$, then $[B_1, \ldots B_n \models A]$, whatever $B_1, \ldots B_n$ may be.

This shows how we can extend the notion of an argument to include the case in which there are zero premises. This does not reflect anything in ordinary usage, but it is convenient for logic.

Ex. 1.18 (a) Show that:

If $[\models B]$, then $[A_1, \ldots A_n \models C]$ if and only if $[A_1, \ldots A_n, B \models C]$.

This shows that necessary truths are redundant as premises. You may now wish to think back to your answer to Ex. 6a(vi).

(b) Show that:

If $[(B) \models]$, then $[A_1, \ldots A_n, B \models C]$, whatever $A_1, \ldots A_n, C$ may be.

Could "if" here be strengthened to "if and only if"? Justify your answer.

The properties of validity mentioned in this section are properties of the traditional notion. In various ways, some of which we shall discuss in §**7**, the traditional notion may seem to fall short of what we want. This has prompted the development of various "non-classical" conceptions of validity. Our concern is confined to the classical notion.

Ex. 1.19 (a) Say whether or not you think that:

$[A_1, \ldots A_n \models C]$ if and only if $[\models$ If A_1 and $\ldots$ and A_n then $C]$.

Briefly justify your view.

(b) Show that:

$[(A) \models]$ if and only if $[\models$ It is not the case that $A]$.

(c) Is it true that

$[(A) \models]$ if and only if $[\not\models A]$?

Justify your answer.

7 Validity and usefulness: "sound", "relevant", "persuasive"

Even if an argument is valid, it may not be *useful*: it may not be a good one to use, either to discover what is true or to persuade an audience of something. For example, consider:

1) Some circles are square. Therefore there will be no third world war.

Since it is logically impossible for any circles to be square, (**1**) is valid (its validity follows from (**6.8**)). But the argument would not be a good one to use for any purpose, and certainly not to convince someone that there will be no third world war. Normally, a good argument is not merely valid. In addition, it has true premises. An argument which has true premises and is valid is called *sound*.

The last remark is qualified by "normally" since there is at least one circumstance in which it is useful to propound a valid argument with a false premise. This is when one hopes that one's hearer will recognize that the conclusion is false and that the argument is valid, and so will be persuaded that at least one premise is false. This mode of argument is called *reductio ad absurdum*.

Suppose your hearer believes that Harry is a merchant seaman, but you disagree. Suppose also that you both know that Harry's arms are not tattooed. Then you might say:

2) Suppose Harry is a merchant seaman. All merchant seamen have tattoos on their arms. So Harry must have tattoos on his arms.

One intends one's hearer to recognize the validity of the argument, and, persisting in his belief that the conclusion is false, to come to infer that at least one premise is false. One has to hope that he will be more firmly persuaded of the truth of "All merchant seamen have tattoos on their arms" than of "Harry is a merchant seaman", so that he will retain the former belief and abandon the latter.

A sound argument may fail to persuade an audience if the audience does not realize that the premises are true, or does not realize that the argument is valid. Here the fault lies with the audience, not with the argument. But a sound argument can still be defective, in that it may not be useful. Consider:

3) Washington is the capital of the USA. Therefore all dogs are dogs.

Since it is logically impossible for "All dogs are dogs" to be false, (**3**) is valid (its validity follows from (**6.9**)). Since the premise is true, it is also sound. But the argument is not useful. Part of the reason is that no argument is needed in order to persuade someone of something so trivial as the conclusion of (**3**). Another part of the reason is that the premise has no proper relevance to the conclusion. For an argument to be useful, it must, normally, be sound, and must, always, be relevant. Logicians have tried to devise special logics to reflect the concept of relevance. But this is one more topic we shall not pursue (see Anderson and Belnap [1975], Read [1988]).

Consider:

4) The whale will become extinct unless active measures are taken to protect it. Therefore the whale will become extinct unless active measures are taken to preserve it.

This is valid (its validity follows from (**6.7**)). It is sound, since its premise is true (if you disagree, select your own example). It is intuitively relevant, for whatever precise account we give of this notion it appears that nothing could be more relevant to whether a proposition is true than whether *that very proposition* is true. But the argument is plainly useless. It could not persuade anyone of anything, and it could not help in the discovery of truth.

For an argument to be *persuasive* for a person he must be willing to accept each of the premises but, before the argument is propounded to him, be unwilling to accept the conclusion. When the premise *is* the conclusion, he cannot be in this state. This is the general reason for the uselessness of circular arguments.

Ex. 1.20 (a) Could there be invalid arguments which are:
(i) sound?
(ii) relevant?
(iii) persuasive?
If so, give examples. If not, say why not.
(b) An argument can be sound and relevant, yet fail to be persuasive through being too elliptical. Give an example.

How could a valid argument ever be persuasive? It is possible because we do not always acknowledge or take explicit note of all the logical consequences of our beliefs. If we did explicitly hold before our minds all the logical consequences of our beliefs, seeing them *as* consequences, we would already have accepted the conclusion of any valid argument whose premises we have accepted. Hence no valid argument could be persuasive. This is how things would be with a perfectly rational being. The utility of valid arguments is a monument to our frailty: to the fact that we are not completely rational beings.

To sum up this section: validity is not the only desirable property in an argument. But it is the only one which normally concerns logicians.

8 Sentences and propositions

An argument consists of *propositions*. A proposition is what is believed, asserted, denied, and so on. This section elaborates this idea.

We can start with the relatively straightforward idea of a meaningful sentence. A sentence is a series of words arranged in accordance with the grammatical rules of the language in question in such a way that it can be used to say something (or to ask something or order something). "The cat sat on the mat" is a sentence which can be used to say that the cat is on the mat, but "cat sat mat on the the", though composed of just the same words, is not a sentence. Properly speaking, we should say: not an *English* sentence, for a sentence is defined relative to a language. The same series of words could be a sentence in one language but not in another. Thus:

1) Plus Robert court, plus Juliette change

is not an English sentence, despite being composed only of English words. But it is (or so I am told) a French sentence.

It is disputed both whether every grammatical sentence is meaningful, and whether every meaningful sentence is grammatical. On the latter point, most people occasionally speak ungrammatically, yet they are understood; and if the sentence they use can be understood, it is hard to justify counting it as not meaningful. On the first point, a standard example of grammatical sentences which are not meaningful is:

2) Green ideas sleep furiously together.

Arguably, this conforms with the rules of grammar, yet is not meaningful. The notion that will be important in this book is that of being meaningful, usable to say something (or to ask something or to order something). To the extent that being grammatical is not a useful guide to being meaningful, it should be set aside.

A preliminary definition of a proposition might run as follows:

3) A proposition is what is expressed, in a given context, by a meaningful, declarative, indicative sentence.

Various aspects of this definition require comment. A *declarative sentence* is one that could be used to make an assertion, to affirm that something is or is not the case. Thus:

4) The King is in his counting house

is a declarative sentence. By contrast

5) Is the King in his counting house?

is not a declarative sentence, but rather an interrogative one. Its typical use is not to *affirm* how things are, but to *ask* how things are.

6) Put the King in his counting house

is not a declarative sentence, but rather an imperative one. Its typical use is not to affirm that things are so-and-so, but to order that they be so-and-so.

An *indicative sentence*, one in what grammarians call the indicative mood, contrasts with a subjunctive sentence. Corresponding to the indicative (**4**) is the subjunctive:

7) Were the King in his counting house.

Subjunctive sentences are not used by themselves to affirm anything, but they may occur in sentences usable to affirm things. One common use is in subjunctive conditionals, for example:

8) Were the King in his counting house, the Queen would be content.

Two distinct sentences can express the same proposition. The English sentence

9) Snow is white

expresses the same proposition as the French sentence

10) La neige est blanche.

Even within a single language, distinct sentences can express the same proposition if they have the same meaning.

11) Blackie is a puppy

expresses the same proposition as

12) Blackie is a young dog.

Two sentences with different meanings can express the same proposition if they are used in different contexts, which is why the definition (3) mentions contexts. Suppose you are my only audience, and I address the following remark to you:

13) You are hungry.

Suppose that you then utter the sentence:

14) I am hungry.

We both express the same proposition. The first sentence, in the context of being *directed at* you, expresses the same proposition as the second, in the context of being *uttered by* you. Had you uttered the same sentence as me, (13), you would not have expressed the same proposition. This shows that the same sentence can, with respect to different contexts, express different propositions.

Ex. 1.21 (a) Suppose Tom utters the following sentence, directed at you: "I like kissing you". Give an example of a sentence you could utter and thereby express the same proposition as Tom expressed.

(b) Suppose Tom yesterday in London uttered the sentence "It is raining here". Give an example of a sentence you could utter today in Paris and thereby express the same proposition as Tom expressed.

(c) Suppose Tom, walking in the woods yesterday, dimly perceived what was in fact a boa constrictor (though he did not know it was), and, referring to that object, uttered: "That is the strangest looking rabbit I ever saw". Give an example of a sentence that you could utter today, in the comfort and security of your own home, and thereby express the same proposition as Tom expressed.

There is another way in which this can occur. A sentence may be ambiguous. For example, "There's that crane again" may refer to a lifting-device or a bird. There is no such thing in general as *the* proposition such a sentence expresses.

I have simply stipulated certain features of propositions, and their relations to sentences. What makes this an appropriate definition to adopt in logic? A standard answer is this: validity is defined in terms of truth conditions, and so one should identify a proposition by truth conditions. This answer relies on ideas that will be introduced in §9. However, we can see at once something of the motivation for the notion of a proposition, as used by logicians.

Consider the argument:

15) I am hungry; therefore I am hungry.

Intuitively this should count as valid. But suppose we thought of the components of arguments as sentences, and suppose we imagine the context shifting between the utterance of the premise and the utterance of the conclusion. Suppose you are hungry and utter the premise, and I am not hungry and utter the conclusion. Then we would have a true premise and a false conclusion, so the argument could not be valid. Clearly we need to avoid such problems, and introducing the notion of a proposition, in the style of this section, is one way of doing so.

We still *could* have defined an argument as a collection of sentences, but we would have had to say something about the context being held constant over all the sentences of an argument. The upshot would have been the same. On some occasions, it is easiest to think of arguments as composed of propositions, on others it is easiest to think of them as composed of sentences, with a background assumption of constancy of context. We will help ourselves to both notions, as convenient.

9 Validity and truth conditions

A sentence like "Snow is white" is true in some but not all logically possible circumstances. There are logically possible circumstances, including those which actually obtain, in which the sentence is true, and logically possible circumstances (in which, say, snow is black) in which the sentence is false. A circumstance is one *in which* a sentence is true just on condition that if the circumstance actually obtained, then the sentence would be true.

Some sentences, for example, "Snow is snow", are true in all logically possible circumstances. Some sentences, for example "7 is less than 5", are true in no logically possible circumstances.

We shall say that a sentence's *truth conditions* are the circumstances in which it is true. We can think of these circumstances as bundled together in a collection or set – a set with no members, in the case of sentences like "7 is less than 5". Using this notion, we can give yet another definition of validity:

1) $[A_1, \ldots A_n \models C]$ if and only if the truth conditions of C include those of $[A_1, \ldots A_n]$.

We could put it another way: every circumstance in which all of A_1, $\ldots A_n$ is true is one in which C is true. Equivalently: the truth conditions of $A_1, \ldots A_n$ are included in those of C. (**1**) and these variants define the same notion as that defined in §3.

Ex. 1.22 Show that an argument is valid according to (**10.1**) iff it is valid according to the definition of validity displayed near the end of §3.

The importance of this definition is that it shows that only the truth conditions of a sentence matter to the validity or otherwise of any argument in whose expression it occurs. A consequence is that if a sentence occurs in the expression of an argument and you replace it by one having the same truth conditions, the argument will remain valid, if it was valid before, or invalid, if it was invalid before. This will bear importantly on questions of formalization, to be considered later.

10 Formal validity

Logic, or at any rate formal logic, is not primarily concerned merely with the very general notion of validity which we have so far discussed. It is concerned with a particular species: *formal validity*. Formal validity, being a kind of validity, has all the properties of validity; but it has some additional distinctive features.

We could try to define formal validity by saying that an argument is formally valid if and only if it is valid in virtue of its form or pattern. This captures part of what is intended, though it will need supplementation. Here is a pair of arguments which are said to have the same form:

1) Frank will marry Mary only if she loves him. But Mary does not love Frank. So he will not marry her.

2) The whale will be saved from extinction only if active measures are taken. But active measures will not be taken. So the whale will not be saved from extinction.

These have a common form or pattern, which we could distil out as follows:

3) . . . only if——. It is not the case that——. So it is not the case that . . .

The dots are meant to be filled on both occurrences by the same sentence, and likewise the dash. It is more convenient to use letters, rather than dots and dashes, thus:

4) A only if B. It is not the case that B. So it is not the case the A.

This can be called an *argument-form*. It is an argument-form of each of (**1**) and (**2**), since these both result from it by making suitable replacements for A and B.

Ex. 1.23 Give two further instances (not in the text) of (**10.4**). Can you find any invalid instances?

The logician wants to say that (**1**) and (**2**) are valid in virtue of their pattern or form, the same in each case. This represents among other things an attempt to attain generality. It would be hopeless to try going through each argument in turn, picking out the valid ones. But if we are granted the idea of an argument-form we can say: not only is this specific argument valid; so are all of the same form.

One way to elaborate this a little is to define a notion of validity for argument-forms:

5) An argument-form is valid if and only if, necessarily, each of its instances is valid.

Ex. 1.24 Why is the qualification "necessarily" needed in (**10.5**)?

So (**4**) is an example of a valid argument-form.[6] An argument is valid *in virtue of* its form just on condition that it is an instance of a valid argument-form.

This goes some way towards saying what formal validity is, and we can reinforce the idea with two more examples before showing that an important ingredient is missing.

Consider:

6) All camels are herbivores. All herbivores are pacific. Therefore all camels are pacific.

This is an instance of the argument-form: "All . . . are——, all——are ***; therefore all . . . are ***." "Camel" and "herbivore" are traditionally classified as nouns, "pacific" as an adjective. We shall call both of these, as they occur in (**6**), *predicates*. We shall also include among predicates relational expressions like "loves" and "is bigger by . . . than". To mark the sort of gap that a predicate can fill, we shall use capital letters starting with "F", so we shall write out the argument-form:

7) All F are G. All G are H. Therefore, all F are H.

Assuming that this is a valid argument-form, it follows that (**6**) is not merely valid, but also formally valid.

Ex. 1.25 Give a further instance (not in the text) of (**10.7**). Can you find any invalid instances?

8) Ian is Scottish. All Scots are prudent. So Ian is prudent.

This is an instance of the argument-form:

9) α is F. All F are G. So α is G.

Here we use "α" (and if needed, "β", "γ", . . .) to mark the position which a name can occupy. Assuming that (**9**) is valid, (**8**) is formally valid.

Ex. 1.26 (a) Give a further instance (not in the text) of (**10.9**). Can you find any invalid instances?

(b) Specify valid argument-forms for any of the following which are formally valid:

(i) Some politicians are liars and all liars are charming. So some politicians are charming.
(ii) Some politicians are liars. Veredici is a politician. So Veredici is a liar.
(iii) If you are going to die by drowning, then there is no point in learning to swim. If you are not going to die by drowning, then there is no point in learning to swim. Therefore, there is no point in learning to swim.
(iv) Everyone who has studied logic is able to spot an invalid argument when he sees one. You haven't studied logic. So you are not able to spot an invalid argument when you see one.
(v) 7 is prime. $7 = 5 + 2$. So $5 + 2$ is prime.

[6] One could show this by applying the propositional calculus, if one were willing to make a certain assumption about the relationship between that language and English. See chapter **2.10**.

(vi) The battle of Marengo occurred before the French invasion of Moscow, and the battle of Waterloo came after that. So the battle of Marengo occurred before the battle of Waterloo.
(vii) He will die unless he is given a blood transfusion. But he will not be given a blood transfusion. So he will die.

This discussion has already begun to include some controversial elements. Highlighting these must wait for later chapters. For the moment, I want to bring out what is missing from the account so far: a gap which makes it inadequate as a presentation of the traditional idea of formal validity.

The idea was formal validity should be a special kind of validity. But as presented so far, nothing has been said to rule out formal validity coinciding with validity. For nothing has been said which prevents an argument itself counting as an argument-form. This is no accident of the particular way in which I have presented the idea. The general problem is this: what is the difference between pattern or form, and what fills it: substance or content? In (4), for example, the remaining English words correspond to the pattern or form, the letters *A* and *B* to the places where one could insert content or substance to yield a genuine argument. But what is the basis for this distinction?

To bring out the problem, consider the sort of example that would standardly be given in a logic text of an argument which is valid but is not formally valid:

10) Tom is a bachelor. Therefore, Tom is unmarried.

This is certainly valid (reading "bachelor" in a familiar way). A case for saying that it is not formally valid might start by pointing out that (**10**) is an instance of the invalid argument-form.

11) α is *F*. Therefore α is *G*.

Ex. 1.27 (a) Justify the claim that (**10.11**) is invalid by finding an invalid instance of it. (To make the invalidity plain, pick an example with an obviously true premise and obviously false conclusion.)

(b) Show by example that each of (**10.1**), (**10.2**) and (**10.6**) are instances of invalid argument-forms.

(c) Show that *every* argument is an instance of at least one invalid argument-form.

How could the case be pressed further? What is needed for formal validity is that there be *some* valid argument-form of which the argument is an instance. So to establish failure of formal validity, it is not enough to cite one invalid argument-form of which the argument is an instance (cf. Ex. 1.27). You have to show that it is not an instance of *any* valid argument-form. But who is to say that (**10**) itself is not an argument-form? If it is, then, since (**10**) itself is its only instance, it is an instance of a valid argument-form, and so formally valid, contrary to the intention.

We might try to block this difficulty by stipulating that every argument-form must have some gaps (marked by dots and dashes, or letters). So every argument-form will have more than one instance. But (**10**) would still come out as formally valid, in virtue of being an instance of the valid

12) α is a bachelor. Therefore, α is unmarried.

If the concept of formal validity is to be narrower than that of validity, as the logician intends, the concept of form will have to be made more restrictive. The logician will stipulate that in an argument-form the only expressions we may use, other than the dots or dashes or letters which mark the gaps for the "content", are the *logical constants*. "Bachelor" and "unmarried", which occur in (**12**), are not logical constants, so (**12**) is not an argument-form, and so does not establish the formal validity of (**10**).

11 The logical constants

I shall begin by giving a list of expressions that are generally held to be *logical constants*:

1) it is not the case that
and
or
if . . . then . . .
if and only if
some
a
everything
all
is
are
is the same as
[plus any expression definable just in terms of the above].

It is a matter for philosophical debate whether the list should be extended to include, for example, "necessarily" or "is a member of" (as used in set theory). The debates arise from the fact that there is no general agreement about what *makes* an expression a logical constant. A list like the above fails to speak to this issue.

Historically, logical constants are so-called because they are given a constant interpretation within a logical system. Corresponding to words like "Socrates" and "man", by contrast, there are variable interpretations. This gives a clear guide to what expressions are treated as logical constants by a logical system, but it does not explain or justify the selection of the logical constants: *any* expression could, in principle, be given a constant interpretation.

A widely held view, which certainly captures part of the truth, is that an essential feature of a logical constant is that it introduces no special subject matter. It should be "topic-neutral". This is because, in logic, we are concerned with reasoning in general, and not with this or that special area of knowledge. It is all very well for an anthropologist concerned with kinship to take a particular interest in what is signified by such words as "bachelor" and "married". The logician aims at greater generality. He will concern himself only with expressions which can occur in an argument on any subject whatsoever. The expressions in the list, but not expressions like "bachelor", satisfy this requirement.

Bearing in mind the suggestion that argument-forms should contain, apart from devices for marking gaps, only logical constants, we can

verify that (**10.4**), (**10.7**), (**10.9**) are argument-forms, but (**10.12**) is not one. This is consistent with the formal validity of (**10.1**), (**10.2**), (**10.6**) and (**10.8**), but is hostile to the formal validity of (**10.10**). If our list of logical constants can be assumed to be complete, or at least to exclude expressions like "bachelor" and "unmarried", then it seems that no argument-forms for (**10.10**) would be more apt to reveal it as formally valid than either the invalid (**10.11**) or the equally invalid

2) *A*; *B*.

If so, (**10.10**) is not formally valid.

In chapter **6.5** we will ask whether there is any illuminating and general account of logical constancy. If there is, then we can use it to give an account of formal validity: it will amount to validity in virtue of the meaning of the logical constants, and in abstraction from other than structural features of premises and conclusion. By *structural features* I mean facts about the recurrence of certain non-logical elements, for example the fact that "camels" occurs in the two places it does in (**10.6**). For the moment, we will make do with a relativized notion: given some list of constants, we will say that an argument is formally valid if and only if it is valid in virtue of the meanings of the constants on the list, and in abstraction from other than structural features of premises and conclusion.

The investigation of formal validity has in practice proceeded by turning away from ordinary English and studying artificial, "logical" languages like the *language of the propositional calculus*, and the *language of the predicate calculus*. What is the rationale for introducing these unfamiliar languages? The standard answer is in terms of their "clarity", but it is not clear that ordinary English is other than clear. In the following section, I consider some supposed defects of natural languages like English, considered from a logical point of view.

12 The project of formalization

If logicians really aim to study validity, as it occurs in our everyday thought and talk, why do they study artificial languages, which no one speaks? Why not stick to English, or French, or some other natural language?

An argument is a collection of propositions. But, according to a traditional justification of the turn towards formal languages, the *sentences* of natural languages like English do not adequately reflect the logical properties of the *propositions* they express. Formal logic *is* concerned with the very arguments we use in daily life, but it has to express these arguments in a different way.

This introduces the crucial idea of the *logical form* of a sentence. A sentence's logical form is supposed to lay bare the logical features of the proposition which it expresses. This logical form, it is said, is often hidden by ordinary language.

A traditional hope is that logic should provide a mechanical means of testing for validity.[7] But how could you present a machine with arguments? If arguments are composed of propositions, then you cannot present a machine with arguments in any direct way, for propositions are too abstract. What you would have to feed into the machine are sentences. If the machine is to test the validity of the argument the sentences express, every logically relevant feature of the propositions must be correlated with some property of the physical make-up of the sentences.

It has been held that such a correlation does not obtain, or at least does not obtain in any readily statable fashion, between sentences of natural languages and the propositions they express. Hence the need for artificial languages. The idea is that these will supply the logical forms of sentences in natural languages. By translating a natural sentence into an artificial one, the hidden logical features of the proposition expressed will be brought to the surface.

Let us now consider some ways in which natural sentences may be supposed inadequate for logical purposes: inadequate as vehicles for bringing out the logical features of arguments.

(1) Lexical ambiguity

As a special case of (**6.7**) it holds quite generally that:

$$C \models C.$$

[7] The hope has a long history, going back at least to Leibniz. If it were made precise in terms of the technical notion of *decidability*, it is a hope which will be disappointed for any interesting logic: see e.g. Delong [1970], pp. 132ff., or Kirwan [1978], pp. 169ff.

This holds whatever proposition C may be. But, as we saw in connection with (**8.15**), it does not hold for arbitrary sentences of natural languages. One reason is that many sentences are ambiguous: they have more than one meaning, express more than one proposition. When this is due to the sentence containing a word with more than one meaning, we shall call the ambiguity *lexical*.

1) John cut the painter. Therefore John cut the painter

is not valid if we interpret "painter" in the first sentence to mean an artist, and interpret this word in the second sentence to mean a rope used to secure a boat; for then the first sentence may be true while the second is false.

An obvious way to deal with this problem is to distinguish two words, say "painter_1" and "painter_2", one for each of the meanings, and throw away the ambiguous word. Then the proposed interpretation of (**1**) would look like this:

2) John cut the painter_1. Therefore John cut the painter_2

and no one would be particularly tempted to think that this expressed an argument of the form: C; C. (No doubt one would also have to distinguish "cut_1" and "cut_2".) The strategy of eliminating ambiguous words already involves departing from natural languages, in which ambiguous words are rife. But the proponent of artificial languages envisages altogether more radical departures.

(2) Structural ambiguity

Some sentences are ambiguous, yet the ambiguity cannot be attributed to the ambiguity of one or more words in the sentence: the ambiguity is not lexical but *structural*. Here are some examples of allegedly structurally ambiguous sentences, with alternative interpretations added in brackets.

3) Harry is a dirty window cleaner. [(a) Harry is a dirty cleaner of windows; (b) Harry is a cleaner of dirty windows.]

4) Tom and Mary are visiting friends. [(a) Tom and Mary are visiting some people, and they are friends with these people;

(b) Tom and Mary are friends with one another, and they are visiting some people; (c) Tom and Mary are visiting some people, and these people are friends with one another.]

5) Receipts from this source are not liable to income tax under section iv, paragraph 19. [(a) Section iv, paragraph 19, exempts receipts from this source from liability to income tax; (b) section iv, paragraph 19, does not impose a liability to income tax on receipts from this source.]

6) I thought you were someone else. [(a) I thought you were someone other than the person you in fact are; (b) I thought you were someone who is not identical to himself.]

7) *First speaker*: "I ought to send flowers."
Second speaker: "No you ought not."
[(a) You are not under an obligation to send flowers; (b) you are under an obligation not to send flowers.]

8) Nicholas has written a book about everything. [(a) Nicholas has written a book, and it treats every subject; (b) for every subject, Nicholas has devoted at least one whole book exclusively to it.]

In none of these cases can the alleged ambiguity be attributed to any word. One way to test for this is to verify that each of the words can occur in a variety of sentences lacking the corresponding ambiguities, and this is not to be expected if the words are ambiguous.

Ex.1.28 Consider the following objection to the test mentioned in the text:

> Despite the fact that "cut" is ambiguous, "I cut the string with my pocket knife" is not. Hence the fact that a word can occur in unambiguous sentences does not show that is unambiguous.

The existence of structural ambiguity shows that the elimination of lexical ambiguity is not enough. Some more radical approach is required.

Structural ambiguity seems to affect even logical constants, for example "not" in (**7**) and "a" and "everything" in (**8**). On one reading of (**7**) we think of "not" as dominating the sentence, to form its negation, a reading which we might write as "Not: you ought to send flowers". On the other reading, we think of "not" as governing just the description of the action, a reading which we might write as "You ought to do this: not send flowers". On one reading of (**8**) we think of "a" as dominating the sentence, a reading which we might write as: "A book by Nicholas is like this: it is about everything". In the other reading, we think of "everything" as dominating the sentence, a reading which we might write as: "Everything has this property: Nicholas has written a book about it". The logical constants determine formal validity (see §**10** and §**11**). If structural ambiguity can affect the logical constants, then the hope of giving a general characterization of formal validity for English as it stands is undermined. Consider:

9) Logic, epistemology and metaphysics are all the philosophical subjects there are. Nicholas has written a book about logic. Nicholas has written a book about epistemology. Nicholas has written a book about metaphysics. Therefore, Nicholas has written a book about every philosophical subject.

If (**9**) is valid, the standard view is that it is formally valid. But there is no straightforward answer to the question whether it is valid. It all depends on how we understand the conclusion, which is structurally ambiguous in the same fashion as (**8**).

The problem for the account of formal validity is as follows. We said that a formally valid argument is valid in virtue of its form, and that this in turn is a matter of it being an instance of a form all of whose instances are valid. However, (**9**), read as invalid, is an instance of every argument-form of which (**9**), read as valid, is an instance. Hence (**9**) is not an instance of a valid argument-form. The problem of structural ambiguity threatens to deprive even the apparently formally valid reading of (**9**) of its formal validity.

Ex.1.29 If any of the following is ambiguous, give unambiguous paraphrases of the alternative interpretations:

(i) I am going to buy a book.
(ii) Everyone has a problem.
(iii) "In the whole wide beautiful world, Aldo Cassidy was the only person who knew where he was." (Le Carré [1971], p. 8)
(iv) "Most of all I would like to thank my students, who have taught me more than they know." (E. Bach [1974], p. vi)

It is theoretically possible that structural ambiguity could be filtered out of natural languages. In (3)–(8) unambiguous paraphrases in English were given; perhaps structurally unambiguous paraphrases are always available. But it is unclear whether precise rules can be given which would effect this filtering. One can see why logicians might prefer artificial languages: they are constructed from the ground up in such a way that structural ambiguity is impossible.

(3) Syntactic irregularity

The *syntax* or *grammar* of a language is a set of rules which determine how sentences are constructed from the language's vocabulary. A syntactic distinction is one which we have to make in order to devise such rules.

As we have seen, there are two possible answers to the question: what ought to be picked out as sentences? One answer is: just the series of words which constitute *grammatical* sentences, where it is supposed that we have some antecedent grasp on what it is for a series of words to be a grammatical sentence. The other answer is: just the series of words which constitute *meaningful* sentences. (Cf. (**8.2**) and related discussion.) Without prejudice to this debate, I shall in this section mean by syntactic rules ones which determine the class of meaningful sentences.

We have already been obliged to make various syntactic distinctions in English, for example, between sentences, predicates and names. We used $A_1, \ldots, A_n$, B, C, as letters marking the sort of position that can be occupied by a sentence; F, G, H as letters marking the sort of position that can be occupied by a predicate; and α as a letter marking the sort of position that can be occupied by a name. We have attempted no definition of these categories. Rather, we have simply picked out examples, and gestured towards the category as a whole.

The gesture is supposed to determine the category from the example in the following way. Anything belongs to the category if it can replace the example (at least in the context under consideration) without turning sense into nonsense. Given that "Tom" belongs to the category of names, we can infer that "Harry" does too, since it can replace "Tom" without turning sense into nonsense. But neither "herbivores" nor "2 + 2 = 4" are names, by this test, since replacing "Tom" in "Tom is a bachelor" yields the nonsensical "Herbivores is a bachelor" and "2 + 2 = 4 is a bachelor". I call this way of determining syntactic categories the *naive syntactic test*. The taxonomy the test produces is inadequate for the study of validity. It places expressions with similar logical powers in different categories; and it places expressions with dissimilar logical powers in the same categories.

The expressions "Mount Everest" and "Ronald Reagan" have logically similar powers. Each serves to pick out an object. Yet it is at least arguable that replacing the latter by the former in "Ronald Reagan is thinking of Vienna" turns sense into nonsense. If so, these names fall into different categories, according to the naive syntactic test. The uncertainty reveals the vagueness of the distinction between sense and nonsense.

Here are some examples which suggest that the taxonomy produced by the naive syntactic test places expressions with dissimilar logical powers in the same categories. By the naive syntactic test, it would seem that the category of names would contain not only expressions like "Clinton" and "Harry" but also what logicians call *quantifier phrases*, like "everyone", "no one", "someone". For example, the results of replacing "Clinton" in "Clinton is a bachelor" by any of "everyone", "no one" or "someone" make perfectly good sense. But the logical powers of "Clinton" and "no one" are very different, as is brought out by the fact that (**10**) is valid but (**11**) is not:

10) Clinton is a bachelor. So someone is a bachelor.

11) No one is a bachelor. So someone is a bachelor.

Ex.1.30 Although as they occur in sentences like (**10**) and (**11**), "Clinton" and "no one" can be interchanged without turning sense into nonsense, it does not follow that they belong to the same syntactic category by the naive syntactic test: this requires that they be everywhere substitutable. Doubts are raised by examples like "No one ever complains" (since "Clinton ever complains" isn't acceptable in contemporary English). Can you think of other, perhaps more decisive, cases? (Cf. Oliver [1999], pp. 253–4.)

The contrast is exploited by Lewis Carroll [1872]:

12) "Who did you pass on the road?" the King went on, holding out his hand to the Messenger for some more hay.

"Nobody", said the Messenger.

"Quite right", said the King: "this young lady saw him too. So of course Nobody walks slower than you."

"I do my best", the Messenger said in a sullen tone. "I'm sure nobody walks much faster than I do!"

"He can't do that", said the King, "or else he'd have been here first." (pp. 143–4)

The King pretends to treat "Nobody" as a name rather than as a quantifier phrase.

Despite its vagueness, the naive syntactic test at least doesn't definitely rule out counting both the expressions "is sensitive to pain" and "is evenly distributed over the earth's surface" as predicates. Since "Harry is sensitive to pain" is clearly sense, this means allowing that "Harry is evenly distributed over the earth's surface" is sense too. (Presumably the sentence is false, and so meaningful. Perhaps we could imagine it also being true, if Harry were chopped into small pieces, which were then dropped at regular intervals from an aeroplane.) If both expressions belong to the same category, however, we run into a problem.

13) Human beings are sensitive to pain. Harry is a human being. So Harry is sensitive to pain

would standardly be said to be formally valid. Does not its validity turn only on the logical constants it contains? It is an instance of the argument-form

14) F are G. α is an F. So α is G.

It is tempting to believe that (**14**) is valid, and explains the formal validity of (**13**). But the temptation must be resisted, as the invalidity of (at least one reading of) the following shows:

15) Human beings are evenly distributed over the earth's surface. Harry is a human being. So Harry is evenly distributed over the earth's surface.

Ex. 1.31 What, if anything, is the reading of **(12.15)** upon which it is valid?

The invalidity of (**15**) establishes the invalidity of (**14**). We need a more refined notion of a predicate, if we are to attain interesting generalizations about valid forms of argument.

Ex. 1.32 Give two further invalid instances of **(12.14)**.

The following argument would generally be considered formally valid:

16) Every candidate is a clever or industrious person. Every clever or industrious person is worthy of praise. So every candidate is worthy of praise.

But the argument-form we would reach for to sustain this judgement is invalid:

17) Every F is a G. Every G is H. So every F is H.

For the following instance of (**17**) is invalid:

18) Every number is a number or its successor. Every number or its successor is even. So every number is even.[8]

Expressions which look similar, at least to the naive eye, can contribute in very different ways to the meanings of sentences in which they occur. This is the phenomenon I refer to by the phrase *syntactic irregularity*. A closely allied phenomenon is that in natural languages it seems to be impossible, or at least difficult, to characterize properties which are of logical importance in the way which would make mechanical testing possible: that is, on the basis of the physical make-up of sentences.

This can be illustrated by the relation of *negation*, which is clearly important to logic, as its connection with inconsistency and validity has already made plain. We saw in (**4.6**) and (**4.8**) that there are many sentences in which one can insert a negative particle, for example "not", without forming the negation of the sentence. This means that

[8] The example is from Geach [1972], pp. 492–3.

it will be hard to formulate a general rule picking out just those pairs of sentences one of which is the negation of the other. Might one at least give an infallible rule for *one* way of forming the negation, for example the rule that, for any sentence, *S*, "it is not the case that *S*" is its negation? The rule works well in many cases. For example, it says, correctly, that "It is not the case that the earth is flat" is the negation of "The earth is flat". But it does not work for all. For example, prefixing:

19) I will marry you, if you change your religion

with "It is not the case that" yields

20) It is not the case that I will marry you, if you change your religion.

This is at best ambiguous between the negation of (**19**) and something equivalent to "If you change your religion, I will not marry you".

We need some kind of bracketing device. We might write: "It is not the case that (I will marry you, if you change your religion)". In spoken English, a similar effect can be achieved by inflection, for example one which includes a small pause after "that".

The introduction of such special devices is typical of the formal logician's approach. One point of the devices is that they facilitate the characterization of relations which are of logical importance (like negation) purely in terms of the physical make-up of sentences.

The question remains open whether such a result could be achieved merely by tinkering with a natural language, or whether it requires starting from scratch. The idea of starting from scratch, constructing an artificial language constrained only by the demands of logic, has inspired a philosophical tradition (though one whose merits are nowadays being questioned). Russell, for example, coined the expression "philosophical logic" to represent his view that the workings of natural language, and of our thought, could be adequately represented only by an artificial language, the language of his *Principia Mathematica*.

With this approach comes a problem. How are whatever results are obtained for the artificial language to be applied to natural language and to our everyday thoughts? A project opens up, which I call the project of formalization. The idea is to pair each natural sentence with an artificial one. The latter is, or reveals, the logical form of the former. Thanks to the pairing, the results about validity which we have been able to obtain, with relative ease, for the artificial language can be transferred to the natural one. To put it in another idiom: the results about validity which we have obtained by expressing arguments in an artificial language become relevant only if these arguments are, or are specially related to, those we use in our everyday thought and talk. One demonstrates the relevance by showing how to pair natural language sentences with artificial language sentences in such a way that the propositions expressed by the former are the very same as, or specially related to, the arguments expressed by the letter.

Within this tradition, the first question to ask about an argument expressed in a natural language is: what is its logical form? The answer is to be given by translating the argument into some artificial language: by, as it is called, *formalizing* the argument. In the next four chapters, we examine in detail how the project of formalization proceeds.

2

Truth functionality

This chapter begins (in §§**1–2**) by introducing an artificial language: the language of propositional truth functional logic, here called **P**. Readers already familiar with this language should merely skim these sections, to check on the terminology and symbolism used here. (Of particular importance is a grasp of the precise notion of *interpretation*, in terms of which an appropriate notion of validity – here called **P**-validity – is defined.) In the later sections, the following question is discussed: what can the validity of arguments expressed in this artificial language tell us about the validity of arguments expressed in English? A crucial prior question will be whether the logical constants of **P** adequately translate the English expressions to which they are taken to correspond.

1 The classical propositional language

The two main features of **P** are these: (i) the only logical constants it recognizes are *sentence connectives*; (ii) all its sentence connectives are *truth functional*.

Sentences of **P** are composed of two kinds of symbol: the *letters* of **P**; and the *sentence connectives* of **P**. The letters of **P** are p, q, r, p' etc. (we here envisage an endless supply), and they are used to formalize sentences which express propositions.

The ***P**-logical constants* are the following sentence connectives:

$\neg$ (corresponding to "it is not the case that"; the symbol is called "tilde");

& (corresponding to "and"; called "ampersand");

$\vee$ (corresponding to "or"; called "vel");

$\rightarrow$ (corresponding to "if . . . then . . ."; called "arrow");

$\leftrightarrow$ (corresponding to "if and only if"; called "double arrow").[1]

The **P**-sentences are the following:

1) a letter, standing alone. (That is, the letters themselves also count as sentences.)
2) (**a**) any sentence preceded by "$\neg$". We can write this more economically as follows: if X is a sentence, so is $\neg X$.

   (**b**) if X and Y are sentences, so are
   $(X \& Y)$;
   $(X \vee Y)$;
   $(X \rightarrow Y)$;
   $(X \leftrightarrow Y)$.

Examples: (i) "$\neg p$" is a sentence; for "p" is a sentence by (**1**), and "$\neg p$" results from it by preceding it by "$\neg$"; so, by (**2a**), it is a sentence. (ii) "$(\neg p \& (r \leftrightarrow s))$" is a sentence, since "$(r \leftrightarrow s)$" is a sentence, by (**1**) and (**2b**), and so is "$\neg p$", which establishes, by a further application of (**2b**), that "$(\neg p \& (r \leftrightarrow s))$" is a sentence.

The above gives what is called the *syntax* of the language **P**: rules which determine what is to count as a sentence of **P**. (We sometimes omit outer parentheses around **P**-sentences, provided that there is no danger of confusion.) We now turn to the *semantics* of **P**. These are rules which in some sense specify the meanings of sentences of **P**.

(1) Truth values

"Coal is white" is false, but "Snow is white" is true. We shall record this information by saying that "Coal is white" *has the truth value false* and "Snow is white" *has the truth value true*. We are thus thinking of the truth values, truth and falsity, the true and the false, as kinds of object. True sentences stand in the special relation of *having* to *the true*;

1 "–" and "~" (the last being a tilde properly so-called) are sometimes used in place of "$\neg$"; ".", "$\wedge$" or simple juxtaposition in place of "&", "$\supset$" instead of "$\rightarrow$", and "$\equiv$" instead of "$\leftrightarrow$". For a fuller list of variants, see Kirwan [1978], p. 280.

false sentences stand in that very relation to *the false*. This way of putting things has become standard, having been proposed, for subtle philosophical reasons, by Frege ([1892b], esp. p. 47). If you are reluctant to posit these abstract objects, true and false, you are in good company (see Dummett [1973], pp. 401–27). However, truth values, conceived of as objects (and so as capable of being "assigned" to sentences), are standardly assumed in logic; let us go along with this assumption.

(2) Interpretations

> **3)** An interpretation of **P** assigns exactly one of the truth values, true or false, to each sentence-letter in **P**.

For any one letter of **P**, there are two interpretations, one which assigns it the true, and one which assigns it the false. For any two letters, there are four interpretations, one assigning the true to both, one assigning the false to both, and two assigning different values. We cannot properly speak of a **P**-letter as being true or false, without qualification. Rather, a letter can be spoken of as *true (or false) upon an interpretation.* This means that **P** is the shell or structure of a language, rather than a real language which can be used to say things true or false (cf. Kirwan [1978], pp. 3–8, 32–41, Smiley [1982]): the structure requires completion by an interpretation.

Interpretations are supposed to be a contribution to semantics, and semantic theory is supposed in some sense to specify meaning. It is obvious that the meaning of a sentence cannot consist in its truth value. So how can an interpretation, in the sense of (**3**), be counted as part of semantics? The hope is that interpretations will specify *enough* meaning for the logical purpose in hand: enough for the study of validity to the extent that this arises from the presence of **P**-logical constants. For this, the meaning of these constants will be critical, and arguably it can be specified just in terms of interpretations.

Ex. 2.1 Give a simple argument for the conclusion that the meaning of a sentence cannot consist in its truth value.

(**3**) is not a complete specification of what an interpretation is, for it applies only to **P**-letters, whereas there are (infinitely many) **P**-sentences which are not **P**-letters. In extending the account of an inter-

Table 2.1

X	Y	¬X	X∨Y	X&Y	X→Y	X↔Y
T	T	F	T	T	T	T
T	F	F	T	F	F	F
F	T	T	T	F	T	F
F	F	T	F	F	T	T

pretation to other **P**-sentences, one arguably fixes the meaning of the **P**-logical constants. This extension is effected by interpretation rules which fix the ways in which truth values are transmitted upwards, from **P**-letters to more complex **P**-sentences.

(3) P-interpretation rules

Henceforth "if and only if" will be abbreviated "*iff*".

> 4) For any interpretation of **P**, say *i*,
> ¬ *X* is true upon *i* iff *X* is false upon *i*;
> (*X* & *Y*) is true upon *i* iff *X* is true upon *i* and *Y* is true upon *i*;
> (*X* ∨ *Y*) is true upon *i* iff *X* is true upon *i* or *Y* is true upon *i*;
> (*X* → *Y*) is true upon *i* iff *X* is false upon *i* or *Y* is true upon *i*;
> (*X* ↔ *Y*) is true upon *i* iff either both *X* and *Y* are true upon *i* or both *X* and *Y* are false upon *i*.

One standard way to codify this information is by means of *truth tables*. In table 2.1 a "T" ("F") below an expression indicates that it is true (false) upon an interpretation which assigns the values to *X* and *Y* which are indicated at the left of the row. Thus, for example, the third row of the table for the "→" column says that an interpretation upon which *X* is false and *Y* is true is one upon which *X* → *Y* is true.

Neither the interpretation rules nor the truth tables could be adequately represented by using just **P**-sentences. Suppose the rule for "¬" were written:

> "¬*p*" is true on an interpretation iff "*p*" is false on that interpretation.

The displayed condition tells us what the truth value of "p" preceded by a tilde is, relative to an assignment of a truth value to "p", but does not tell us what truth value will be accorded, relative to an assignment of a truth value to "p", to the result of prefixing a tilde to "q", nor what truth value will be accorded, relative to an assignment of a truth value to "p", to the result of prefixing a tilde to a complex sentence, not a sentence-letter, for example, to the result of prefixing "$\neg p$" by a tilde. We thus use "X" and "Y" to stand for arbitrary **P**-sentences; they are called metalinguistic variables relative to **P**. We could have attained the required generality without using "X" and "Y" by, for example, writing the rule for "$\neg$":

> A **P**-sentence consisting of a tilde followed by a **P**-sentence is true upon an interpretation iff the latter sentence is false upon the interpretation.

In the light of the rules, an interpretation will determine a truth value for every **P**-sentence. There are just as many interpretations as there are ways of assigning truth values to the **P**-letters; each assignment to the letters determines, via (**4**), a unique truth value for every **P**-sentence.

(4) Nomenclature

$\neg X$ is called the **P**-negation of X;
$(X \& Y)$ is called the **P**-*conjunction* of the **P**-*conjuncts* X and Y;
$(X \vee Y)$ is called the **P**-*disjunction* of the **P**-*disjuncts* X and Y;
$(X \rightarrow Y)$ is called the (*material*) **P**-*conditional* with **P**-*antecedent* X and **P**-*consequent* Y;
$(X \leftrightarrow Y)$ is called the (*material*) **P**-*biconditional* of X and Y.

(5) P-validity

5) An argument in **P**, $X_1, \ldots X_n$; Y, is ***P**-valid* iff every interpretation upon which all the premises are true is one upon which the conclusion is true.
Abbreviation: $X_1, \ldots X_n \models_P Y$.

The "**P**-" prefix and subscript will be dropped when there is no danger of ambiguity. The point of it is to facilitate the formulation of various comparisons, for example with English sentences. Thus we have already defined negation for English. A question to be asked later is: is **P**-negation essentially the same thing as negation? Equally, we have given a definition of validity for English. A question to be asked later is: is **P**-validity (represented by "$\models_P$") essentially the same thing as validity in English (represented by plain, unsubscripted, "$\models$")?

The sentence connectives of **P** are so called because they *take* one or more sentences to *make* a fresh sentence. Languages other than **P** contain sentence connectives. For example, "It is not the case that" is a sentence connective in English: it takes *one* sentence, say "John is happy", to form a fresh sentence, "It is not the case that John is happy". Typically (setting aside the kinds of irregularity discussed at (**1.12.20**)), the sentence thus formed is the negation of the sentence from which it was formed. "And" is a sentence connective which takes *two* sentences to make a sentence. For example, it can make "John is happy and Mary is sad" from the two sentences "John is happy" and "Mary is sad". Let's call the sentence or sentences a sentence connective takes to make a fresh sentence the *component(s)* of the fresh sentence; and let's call the fresh sentence itself the *resultant* sentence.

(6) Scope

6) The *scope* of an occurrence of a sentence connective is the shortest **P**-sentence in which it occurs.

Thus the scope of "&" in

7) $\neg(p \& q)$

is "$(p \& q)$", whereas in

8) $(\neg p \& q)$

its scope is the whole of (**8**). An occurrence of a sentence connective is said to *dominate* the sentence which is its scope. Nomenclature of **P**-sentences is determined by their *dominant* connective.

2 Truth functional sentence connectives

This section gives an account of truth functionality. It can be skipped by those already familiar with the notion.

Standard ("classical") propositional logic deals with sentence connectives having a special property: they are *truth functional*. The terminology derives from the mathematical notion of a function, and one can use this to give a mathematically precise definition of truth functionality.[2] Alternatively, we can define a truth functional sentence connective in more informal terms:

1) *A sentence connective is truth functional* iff whether or not any resultant sentence it forms is true or false is determined completely by, and only by, whether its components are true or false.

Take, for example, "it is not the case that . . .", as this is used to form the negation of what follows. Suppose what fills the dots is true. Then the resultant is false. Suppose what fills the dots is false. Then the resultant is true. So "it is not the case that" is truth functional, according to the definition.

Ex. 2.2 Show that the claim that "it is not the case that . . ." is truth functional with respect to what fills the dots needs qualification. (Compare (**1.12.20**).)

All the sentence connectives of **P** are truth functional, according to the definition, as can be seen from (**1.4**) above.

Not every expression which we are inclined to classify as a sentence connective is a truth functional one. For example, we might naturally think of "Napoleon knew that" as a sentence connective. It can take the sentence "St Helena is in the Atlantic Ocean" to form the sentence: "Napoleon knew that St Helena is in the Atlantic Ocean". However, it is not truth functional. The component is true, but this does not, in and of itself, determine whether or not the resultant sentence is true. The mere fact that "St Helena is in the Atlantic Ocean" is true does not settle whether Napoleon knew this or not.

Truth functionality is subject to the *substitution test*. If a sentence connective is truth functional, then the truth or falsehood of every resultant sentence which it forms depends only on the truth or falsehood of the components. So replacing a true component by another true one will make no difference to whether the resultant is true or false, and likewise for replacing a false component by another false one. Let us apply the substitution test to "Napoleon knew that St Helena is in the Atlantic Ocean". We assume that this resultant sentence is true. (Anyone who disagrees can choose their own example of a truth of the form "Napoleon knew that . . .".) The component "St Helena is in the Atlantic Ocean" is also true. If "Napoleon knew that" were a truth functional sentence connective, it would form only sentences which pass the substitution test. But it does not. Consider any truth which Napoleon did not know, for example, "Quarks come in four colours". "Napoleon knew that quarks come in four colours" is false, whereas "Napoleon knew that St Helena is in the Atlantic Ocean" is true. So "Napoleon knew that" fails the substitution test. Substituting one true component for another *does* sometimes yield resultants which differ in truth value.

Ex. 2.3 (a) Can substituting one *false* component for another in the context "Napoleon knew that . . ." lead to different truth values for the resultants?
(b) On the assumption that "necessarily", in, for example, "necessarily no even prime number is greater than 2", is a sentence connective, use the substitution test to determine that it is not a truth functional one.

Now suppose we are wondering whether "and" is truth functional. If it is, then it is clear that it will express the same truth function as "&", as specified by (**1.4**): that is, "*A* and *B*" will be true iff "*A*" is true and "*B*" is true. "Paris is west of Berlin" and "London is north of Paris" are both true. If "and" is truth functional, then the conjunction "Paris is west of Berlin and London is north of Paris" is true. Moreover *any* true conjuncts will yield a true conjunction. So we could replace, say, "Paris is west of Berlin", in the conjunction, by any other truth whatsoever, for example, "5 + 7 = 12", and the new conjunction ("5 + 7 = 12 and London is north of Paris") will be true. The results of this application of the substitution test are consistent with "and" being truth

[2] Where a sentence connective ¢ takes some number *n* of sentences to make a sentence, let's call ¢ an *n-ary* sentence connective. We can write an arbitrary sentence resulting from applying ¢ to the appropriate number of sentential components, $x_1 \ldots x_n$, as ¢$(x_1, \ldots x_n)$. An *n*-ary truth function is a function from *n*-ary sequences of truth values to a truth value. An *n*-ary sentence connective, ¢, *expresses* an *n*-ary truth function *f*, iff *f* is an *n*-ary truth function and for every sentence ¢$(x_1, \ldots x_n)$, where Σ is the sequence of truth-values possessed by $x_1, \ldots x_n$, $f(\Sigma)$ is the truth value of ¢$(x_1, \ldots x_n)$.

functional. The results do not establish the truth functionality of "and". For that, it would be necessary and sufficient to show that *whatever* sentences composed a conjunction, *any* replacement of either component by *any* like truth-valued sentence fails to affect the truth or falsity of the conjunction.

Instead of saying that a connective is truth functional, I shall sometimes say, equivalently, that it *expresses a truth function*.

P can represent only truth functional connections between sentences. English can, it seems, also represent non-truth functional connections between sentences. How extensive is this dissimilarity? And to what extent does it undermine the project of using **P**-validity to understand validity in English? This question is addressed later (especially in §**9** and §**10** below). The next section considers specific cases of formalizing English arguments in **P**, and using **P**-validity, where possible, to say something about validity.

3 Formalizing English in P

Does **P**-validity give a partial characterization of validity, or formal validity? The question divides: can we be sure that if a rendering in **P** of an English argument is **P**-valid, the English argument is valid (or formally valid)? And can we be sure that if a rendering in **P** of an English argument is not **P**-valid, the English argument is not valid (or not formally valid)? The examples which follow suggest some detailed considerations which bear on these questions.

Following a standard terminological practice, we shall call rendering an English sentence or argument in **P** *formalizing* it (more fully, **P**-formalizing it). Consider how one might formalize the following:

1) The battery is flat. If the battery is flat the car will not start. So the car will not start.

First, we stipulate a *correspondence scheme* between English sentences and **P**-letters. For (**1**) we might choose:

2) Let "p" correspond to "The battery is flat" and "q" to "The car will not start".

Using this correspondence scheme, (**1**) is standardly formalized by

3) $p, (p \rightarrow q); q.$

This is **P**-valid (in the sense of (**1.5**) above). Any interpretation, i, upon which all the premises are true is one upon which "p" is true. By the rule for "$\rightarrow$" (specified in (**1.4**) above), if "$p \rightarrow q$" is true upon i, then either "p" is false upon i or "q" is true upon i. So if both "p" and "$p \rightarrow q$" are true upon i, then "q" is also true upon i. Hence every interpretation upon which all the premises are true is one upon which the conclusion is true. Standardly, one infers from the **P**-validity of (**3**), together with the correspondences of (**2**), to the validity of (**1**). The correctness of this inference is a main theme of this chapter. The examples of this section contribute to the answer, and an explicit statement is defended in §**10**.

A further straightforward example (compare (**1.10.1**)):

4) You can buy a ticket only if you have the exact fare. You haven't got the exact fare. So you cannot buy a ticket.

Let "p" correspond to "You can buy a ticket", "q" to "You have got the exact fare". Then the following formalizes (**4**) and is **P**-valid:

5) $(p \rightarrow q), \neg q; \neg p.$

If the premises are to be true upon an arbitrary interpretation, i, then "q" must be false upon i, and so, by the "$\rightarrow$" rule, "p" must also be false upon i, so "$\neg p$" must be true upon i. So all interpretations upon which the premises are true are ones upon which the conclusion is also true.

We obviously cannot expect to make any inference from the quality of a **P**-argument to the quality of an English argument supposedly formalized by it unless the formalization meets some standard of adequacy. We stipulate that if a formalization is to be *adequate*, the associated correspondence scheme should be such that if we replace the **P**-letters by the corresponding English sentences, and then replace the **P**-connectives by the corresponding English connectives, the result is a sentence (argument) that says the same as the original English. (Which English expressions correspond to which **P**-connectives was stipulated in §1.) Let us call the result of applying the correspondences to an argument the *recovered* argument. The proposed standard of adequacy is that the argument to be formalized should say the same as the recovered argument.

A related test of a formalization's adequacy can be given in terms of the notion of an *intended interpretation*. An intended interpretation is one which assigns to the relevant sentence-letters the same truth values as the ones the corresponding English sentences possess. A necessary condition for adequacy is that every sentence in the formalization should be true (false) on an intended interpretation iff the corresponding English sentence is true (false). In other words, an intended interpretation, by assigning the "right" truth values to the sentence-*letters*, must thereby assign the right truth values to the complex sentences (those which are not letters). We do not always know what the truth values of the English sentences are, so we cannot always apply this test for adequacy.

Ex. 2.4 Why would it be incorrect to speak of *the* intended interpretation?

The adequacy of a formalization is relative to a correspondence scheme. Hence, presenting a correspondence scheme is an essential part of presenting a formalization.

There may be more than one adequate formalization. Consider the argument:

6) If the figure is closed and has sides of equal length, then it is square or rhomboid. The figure is closed. The figure has sides of equal length. So it is square or rhomboid.

We could set up the following correspondence scheme:

7) "p" corresponds to "if the figure is closed and has sides of equal length, then it is square or rhomboid", "q" to "The figure is closed", "r" to "The figure has sides of equal length" and "s" to "The figure is square or rhomboid".[3]

[3] Officially, "s" is not a **P**-letter, but let it be regarded as an abbreviation of "p'"; similarly for "t", "u", etc.

Relative to the correspondences of (**7**), the **P**-formalization of (**6**) is:

8) $p, q, r, s.$

According to the definition (**1.5**) of **P**-validity, (**8**) is not **P**-valid, for the premises are all true and the conclusion false upon an interpretation which assigns truth to "p", "q" and "r" and falsehood to "s". Intuitively, (**6**) is valid, yet by our stipulations (**8**), given (**7**), is adequate.

An alternative correspondence scheme is as follows:

9) "p" corresponds to "the figure is closed", "q" to "the figure has sides of equal length", "r" to "the figure is square" and "s" to "the figure is rhomboid".

Relative to (**9**), an adequate formalization is:

10) $(p \,\&\, q) \rightarrow (r \vee s), p, q; (r \vee s).$

(I omit parentheses where no confusion can result.) (**10**) is **P**-valid. For consider an interpretation, say i, upon which all the premises are true. Then "p" and "q" are true upon i. So, by the rule for "&", "$(p \,\&\, q)$" is true upon i. The rule for "$\rightarrow$", together with the supposition that the premise "$(p \,\&\, q) \rightarrow (r \vee s)$" is true upon i, ensures that "$(r \vee s)$" is also true upon i. So any interpretation upon which all the premises are true is one upon which the conclusion is true. For short:

11) $(p \,\&\, q) \rightarrow (r \vee s), p, q \models_{\mathbf{P}} (r \vee s).$

So (**10**), relative to (**9**), is a candidate for demonstrating the validity of the English.

We could have used the following correspondence scheme:

12) "p" corresponds to "the figure is closed", "q" to "the figure has sides of equal length" and "r" to "the figure is square or rhomboid".

Relative to (**12**), the formalization is:

13) $(p \,\&\, q) \rightarrow r, p, q; r.$

This is **P**-valid. So (**13**) is also a candidate for demonstrating the validity of the English.

Ex. 2.5 Show that (**3.13**) is **P**-valid either by informal reasoning using the definitions of the connectives and **P**-validity given in §1, or, if you know one, by a formal method.

All of (**8**), (**10**) and (**13**), coupled with their correspondence schemes, count as adequate formalizations of (**6**): intuitively, they are faithful to the meaning of the English, or at least to those aspects of the meaning of the English that are relevant to propositional logic. But as only (**10**) and (**13**) are **P**-valid, only they could purport to demonstrate the validity of (**6**).

(**8**), (**10**) and (**13**) differ in how much of the structure of (**6**) they make manifest. (**8**) captures the least, (**10**) the most. (**10**) captures more than is necessary to demonstrate the validity of (**6**), (**8**) captures less than is necessary. How much of a given sentence's structure needs to be reflected in its **P**-formalization will vary, depending on specific facts about the argument in which the sentence occurs. I shall say that the more structure an adequate formalization captures, the *deeper* it is. When people speak of "the" logical form of a sentence, they have in mind a formalization which goes as deep as possible.

If a formalization is invalid, can we infer that the argument formalized is also invalid? Consider the following example:

14) If you are good at mathematics, you will find logic easy. But you are not good at mathematics. So you'll find logic hard.

Let "p" correspond to "You are good at mathematics", and "q" to "You will find logic easy". Then a candidate for a formalization of (**14**) is:

15) $(p \rightarrow q), \neg p; \neg q$.

This is not **P**-valid. For consider an interpretation, i, which assigns truth to "q" and falsehood to "p". The conclusion is false upon i. But both premises are true upon i, by the rules for "$\neg$" and "$\rightarrow$". Is (**15**) an adequate formalization of (**14**)?

To obtain the recovered argument from (**15**), replace the letters by the corresponding English sentences as specified in the correspondence scheme, "$\rightarrow$" by "if . . . then . . ." and "$\neg$" by "it is not the case that". The result is:

16) If you are good at mathematics then you will find logic easy. It is not the case that you are good at mathematics. So it is not the case that you will find logic easy.

The sentence from (**14**) "You'll find logic hard" is replaced by "It is not the case that you will find logic easy". You can fail to find something easy without finding it hard. So (**16**) does not say the same as (**14**) and so (**15**) is an inadequate formalization of it. Does this show that there is no inference from the invalidity of (**15**), together with the associated correspondences, to the invalidity of (**14**)?

Consider a different correspondence scheme. Let "p" correspond to "You are good at mathematics", "q" to "You will find logic easy", "r" to "You'll find logic hard". Then the **P**-formalization is:

17) $(p \rightarrow q), \neg p; r$.

This is, obviously, **P**-invalid. Moreover it does not fail the condition for adequacy. Yet (**17**) is, intuitively, in some respect a less good formalization of (**14**) than is (**15**). Can we make sense of this intuition?

Ex. 2.6 Specify an interpretation which demonstrates the **P**-invalidity of (**3.17**), by saying what truth values the interpretation assigns to its **P**-letters.

(**14**) is an argument which someone might actually propound in good faith, believing it to be valid. (**17**) gives no hint about how this mistake is possible, whereas (**15**) does. (**17**) is miles from anything that looks valid. But (**15**) might for a moment look valid, because of its passing resemblance to the **P**-valid

18) $(p \rightarrow q), \neg q; \neg p$.

One of the advantages of formalization is that it makes logical mistakes easier to spot. It is intelligible that one who is reasoning in English should think that the validity that attaches to an argument having the logical form of (**18**) attaches to (**14**). This explanation is abetted by (**15**), but not by (**17**).

If a formalization that counts as inadequate by the present standards can none the less be useful in explaining how someone might wrongly think an argument valid, would it not be better to revise the standards of adequacy? Consider again the first correspondence scheme, and the relationship between (**16**) and (**14**). It is logically impossible for someone to find logic hard, yet not fail to find it easy. As we shall say, "You will find logic hard" *entails* "You won't find logic easy". In general, A entails B iff it is logically impossible for A to be true yet B not true. If A entails B but B does not entail A, we say that A *is stronger than* B and that B is *weaker than* A. The argument that one recovers from the formalization (**15**) by using the correspondences thus has a weaker conclusion than the conclusion in (**14**). However, if some premises are inadequate to establish even a weaker conclusion, they are inadequate to establish a stronger one. So if we allow that the invalidity of (**15**) establishes the invalidity of (**16**), then it should also be allowed to establish the invalidity of (**14**).

This suggests the following emendation of the account of adequacy: we allow that an invalid formalization may be adequate to an invalid argument even if, though the premises of the recovered argument say the same as the original premises, the recovered conclusion is weaker. A parallel relaxation, a case of which will be considered shortly (see (**21**) and (**22**)), would allow that a valid formalization may be adequate to a valid argument even if the recovered argument has weaker premises.

To be able to use these relaxations, however, we would need already to be in a position to recognize the validity of arguments in English, whereas we were investigating whether we could use **P** to test for this property. A test which required us to recognize the target property before we could apply the test would not be much help. So I shall not allow the relaxed standards of adequacy.

Is there any way, using **P**-methods, that we can establish the invalidity of (**14**)? One possibility we would need to rule out is that (**14**) has a deeper formalization which is valid. Think how mistaken it would have been to have concluded that (**6**) was invalid on the basis of (**8**). A necessary condition for there being no deeper formalization is that every truth functional sentence connective in English be somehow reflected by sentence connectives in the formalization. This condition is clearly met by (**15**).

However, the fact that the deepest **P**-formalization is invalid is still not sufficient to establish the invalidity of the English. This is because there are valid arguments whose validity cannot be represented in **P**, for example:

19) All football supporters are interested in sport. Some football supporters are hooligans. So some hooligans are interested in sport.

The following correspondence scheme will ensure we go as deep as we can: let "p" correspond to "All football supporters are interested in sport", "q" to "Some football supporters are hooligans" and "r" to "Some hooligans are interested in sport". This correspondence goes as deep as possible, for there are no words in the sentences corresponding to **P**-letters which correspond to any of the **P**-connectives. Hence no further complexity in the formalization could be justified. Yet the resulting formalization is the **P**-invalid

20) $p, q; r$.

The conclusion is that we cannot hope that all validity should be reflected as **P**-validity.

Sometimes an argument's premises are stronger than required for the conclusion. A putative example is:

21) If the mare dies, the farmer will go bankrupt, and then he will not cultivate the ground. The result will be that the wheat will fail, and this, in turn, will lead to local food shortages. Then the revolutionary spirit of the people will become inflamed, and they will man the barricades. So if the mare dies, the people will man the barricades.

An appropriate correspondence scheme is:

"p" for "The mare dies";
"q" for "The farmer will go bankrupt";
"r" for "The farmer will cultivate the ground";
"s" for "The wheat will fail";
"t" for "There will be local food shortages";

"u" for "The revolutionary spirit of the people will become inflamed";
"v" for "The people will man the barricades".

A possible candidate formalization is:

22) $(p \rightarrow q)$ & $(q \rightarrow \neg r)$, $(\neg r \rightarrow s)$ & $(s \rightarrow t)$, $t \rightarrow (u \;\&\; v)$; $p \rightarrow v$.

The formalization corresponds only approximately to the original. There is a syntactic aspect: for example, the sentence "This will lead to local food shortages" is formalized as a conditional, whereas it is not one. In the project of formalization, this kind of reorganization is standard (though it is hard to find, or devise, an explicit justification for this practice). There is a semantic aspect: explicitly causal idioms like "the result will be" and "this will lead to" have been formalized by weaker **P**-conditionals, e.g. "$\neg r \rightarrow s$". What the correspondences recover from the latter is entailed by "If the ground is not cultivated, then, as a causal consequence, the wheat will fail", but does not entail it.

However, (**22**) is **P**-valid. So, provided we have no qualms about the adequacy of "$p \rightarrow v$" as a formalization of the conclusion, and provided we have the general assurance that **P**-validity can establish validity, we can use (**22**) to establish the validity of (**21**), despite the fact that the former does less thân full justice to the strength of the latter's premises. The idea is that (**21**) would still have been valid, even if the premises had been weaker, as weak as the premises of the argument recoverable from (**22**).

This approach is not available when the **P**-conclusion (more exactly, the conclusion recoverable from the **P**-formalization) is weaker than the English conclusion. For example, consider

23) Putting garlic in the salad will make Richard think that we care nothing for his preferences, and if he thinks that he will be upset. So putting garlic in the salad will make Richard upset.

A suitable correspondence scheme is:

"p" for "Garlic will be put in the salad";
"q" for "Richard will think we care nothing for his preferences";
"r" for "Richard will be upset".

The obvious formalization is:

24) $(p \rightarrow q)$, $(q \rightarrow r)$; $(p \rightarrow r)$

and this, plainly, is **P**-valid. However, "If garlic is put in the salad then Richard will be upset" is weaker than "Putting garlic in the salad will make Richard upset". The falsehood of "Garlic is put in the salad" is enough for the truth of "If garlic is put in the salad, Richard will be upset" on the assumption that "$\rightarrow$" correctly translates the "if" that occurs here. (To check this, consult the rule for "$\rightarrow$" in §**1**.) However, the falsehood of "Garlic is put in the salad" is not enough for the truth of "Putting garlic in the salad will make Richard upset". So the **P**-validity of (**24**) does not show that the premises of (**23**) establish the stronger conclusion: "Putting garlic in the salad will make Richard upset".

I think that (**23**) *is* valid. The trouble is that, like (**19**), its validity cannot be shown by **P**-formalization. As we shall say: it is not *valid in virtue of its **P**-logical form*.

The next example shows how **P**-formalization may be used to resolve ambiguities:

25) John will choose the colour for his new bathroom and will paint it with his own hands only if his wife approves. But his wife doesn't approve. So he will not paint it with his own hands.

A suitable correspondence scheme is:

"p" for "John will choose the colour for his new bathroom";
"q" for "John will paint his new bathroom with his own hands";
"r" for "John's wife approves".

Two formalizations are possible, depending on how we understand the organization of the first sentence:

26) $((p \;\&\; q) \rightarrow r)$, $\neg r$; $\neg q$.
27) $(p \;\&\; (q \rightarrow r))$, $\neg r$; $\neg q$.

(**27**) is **P**-valid, but (**26**) is not. One cannot speak of the validity or invalidity of (**25**). Rather, one must speak of it as valid on one reading (the one corresponding to (**27**)), and invalid on another (the one corresponding to (**26**)). The formalizations treat the ambiguity of the first premise of (**25**) as a matter of scope. In (**26**), "$\rightarrow$" dominates the first premise: it has *wide scope* relative to "&". Analogously, we shall say that (**26**) treats "if", in the first premise of (**25**), as having wide scope relative to "and". In (**27**), "&" dominates the first premise: it has wide scope relative to "$\rightarrow$". Analogously, we shall say that (**27**) treats "and", in the first premise of (**25**), as having wide scope relative to "if".

Ex. 2.7 Give assignments to the letters of (**3.26**) to establish its **P**-invalidity. Show (by informal reasoning, or by a formal method) that (**3.27**) is **P**-valid.

Ex. 2.8 Give English versions of (**3.26**) and (**3.27**) which make plain the contrast between them.

The final example shows how finding the **P**-logical form of an argument may lead one to the view that it is valid, even though one was not clear whether it was valid or not when one looked at the English version:

28) If common sense is correct, then physics is true. If physics is true, then common sense is incorrect. Therefore common sense is incorrect.

Using obvious correspondences, this formalizes to the **P**-valid

29) $p \rightarrow q, q \rightarrow \neg p \models_{\mathbf{P}} \neg p.$

Ex. 2.9 (a) Demonstrate the **P**-validity of (**3.29**).

(b) Formalize each of the following arguments in **P**, showing your correspondence scheme. Determine the **P**-validity of the formalizations, and say what you think can be inferred about the validity of the English:

(i) Peter or Quentin killed Richard. If it was Peter, then the motive was jealousy. If it was Quentin, the motive was greed. But in fact the motive was not greed. So Peter killed Richard.

(ii) Peter will win the election unless Quentin does. Quentin won't win unless he buys all the electors drinks, and that is something he won't do. So Peter will win.

(iii) Neither Peter nor Richard will win the election unless Quentin doesn't stand. But Quentin's not standing would ensure Richard's success. So whatever happens, Peter will not win.

(iv) Although protective legislation has been enacted in most communities, the number of sperm whales is continuing to decline. The cause is either illegal whaling by nationals of countries participating in the protective legislation, or whaling by nationals of non-participating countries, which, under the circumstances, cannot be regarded as illegal. If the first alternative is ruled out, then the decline in numbers will be halted by bringing pressure on the non-participating governments. No doubt they will yield in exchange for subsidized loans. So if the decline in numbers of the sperm whale is not caused by illegal whaling, halting the decline requires making subsidized loans to non-participating governments.

(v) We can infer that the local plonk is good. For if it is not good, some people will not be drinking it. But some people *are* drinking it.

(vi) Either, if Tokyo is the capital of Japan, the EEC will collapse before 2007, or else, if the EEC will collapse before 2007, then Tokyo is the capital of Japan.

Two main questions have been raised in this section: can the validity of an argument be inferred from the **P**-validity of its **P**-logical form? And can the invalidity of an argument be inferred from the **P**-invalidity of its **P**-logical form? The second question has been answered negatively. The first question has not been explicitly answered, though the formalizations of examples like (**1**), (**4**) and (**6**) may encourage optimism. I return to these questions more systematically in §**10** below.

4 Comparison of P-connectives and English

In §**3**, we took for granted that the connectives of **P**, though given their official definitions by the interpretation rules, correspond closely to their English counterparts. It is now time to examine that assumption.

First, we must try to state what sort of correspondence we require. The simplest idea is that when we recover an argument from a logical form by applying the correspondences, the result should be the original argument. But this standard is unduly restrictive. The tradition allows for a certain amount of reorganization, allows one, at the very least, to match "unless" with "∨". A more relaxed standard is that the recovered argument should have all the validity-relevant features of the original.

Ex. 2.10 Show that there is a case for thinking that "He will die unless he is given penicillin" is equivalent to "He will be given penicillin or die".

If this standard is to be met, the **P**-connectives must have all the validity-relevant features of the English expressions they are matched with in formalizing. Since validity is definable in terms of truth conditions, the requirement can be put: the **P**-connectives must make the same contribution to truth conditions as the English connectives with which they are matched. The contribution to truth conditions of a **P**-connective is given by the appropriate interpretation rule. The standard is thus that the English expressions should express the same truth function as the **P**-connectives with which they are matched. A precondition for an expression's expressing a truth function is that it be a sentence connective. This is because it is sentences that are true or false (possess truth values), and a truth function fixes a truth value from (a sequence of) truth values. (For example, the truth function expressed by "&" fixes the value false for a conjunction on an interpretation iff one of the conjuncts is false on the interpretation.) The discussions thus tend to fall into two parts: putative cases in which the English expression corresponding to a **P**-connective is not a sentence connective, and putative cases in which, though a sentence connective, it is not truth functional.

The important notion of *implicature* which is mentioned here is discussed in more detail in §**6**.

(1) "¬" and "not"

The **P**-connective "¬" corresponds closely to the English word "not", and similar phrases like "it is not the case that", as these are used to form negations. If "John is here" is true (false), then "John is not here" and "It is not the case that John is here" are false (respectively, true). So it appears that on at least some occurrences, "not" is a sentence connective and expresses the same truth function as "¬".

There are cases in which "not" does not form a negation. For example, if "Some cows eat grass" is true, it does not follow that "Some cows do not eat grass" is false. Here "not" does not seem to function as a sentence connective, let alone as a sentence connective expressing the same truth function as "not". Rather, it seems to form a new predicate, "do not eat grass".

There are other cases in which "not" seems not to function as a sentence connective. For example:

1) Not Abraham, but George, chopped down the cherry tree.

Here, on the face of it, "not" attaches to a name rather than a sentence. However, it could be argued that (**1**) is merely a fanciful way of expressing

2) Abraham did not chop down the cherry tree but George did

and here "not" does, after all, form a sentence from a sentence (from "Abraham did chop down the cherry tree"), and in a way that expresses the truth function expressed by "¬".

There are no cases in which "not" appears to function as a sentence connective, but a non-truth functional one.

Ex. 2.11 (a) Suppose "*p*" corresponds to "Some cows eat grass". Why does "¬*p*" not formalize "Some cows do not eat grass"? (Check the definition of "¬" in §2.)

(b) Suppose "¬*q*" formalizes "Some cows do not eat grass". To what does "*q*" correspond?

(c) Are there any difficulties in interpreting "not" as it occurs in the following sentences as expressing the same truth function as "¬"? If there are none, provide a **P**-formalization (showing your correspondence scheme). If there are any difficulties, explain.

(i) I am not very optimistic about the upshot of the talks.
(ii) The world will end not with a bang but with a whimper.
(iii) You ought not to smoke.
(iv) All the passengers who have not got tickets will wait in line.

(d) Suppose someone affirms:

This book isn't bad it's very bad.

Should we see "not" as non-truth functional? How else might we understand such an utterance?

(2) "&" and "and"

The **P**-connective "&" is fairly closely matched by the English "and". "Two is an even number and three is an odd number" is indeed true iff "Two is an even number" is true and so is "Three is an odd number". But there are, arguably, some discrepancies.

(a) Cases in which "and" appears not to function as a sentence connective, in which case it certainly could not be translated by "&", which *is* a sentence connective:

3) Tom and Mary came to dinner.

On the face of it, far from taking two sentences to make a fresh sentence, "and" in (**3**) takes two *names* to make a complex subject expression "Tom and Mary". In this sort of case you might argue that appearances are superficial, and that (**3**) abbreviates

4) Tom came to dinner, and Mary came to dinner.

Even if the suggestion works for some cases in which "and" is, superficially, a name connective rather than a sentence connective, it may not work for all. Consider:

5) Tom and Mary lifted the piano.

Arguably this is not equivalent to

6) Tom lifted the piano and Mary lifted the piano.

(**5**) suggests a joint effort, which (**6**) does not. (One could add "together" at the end of (**5**), to make the collaboration entirely plain.) In such cases, it seems we have to regard "and" as forming a complex name out of two simple names. For different reasons, we seem to have to say the same for examples like the following:

7) Tom and Mary are compatriots.

This is not equivalent to

8) Tom is a compatriot and Mary is one too.

We seem to have another kind of case in which "and" joins names rather than sentences.

Yet another putative example of "and" occurring as something other than a sentence connective is the following:

9) Some girls are pretty and flirtatious.

Ex. 2.12 There is an analysis of (**4.7**) according to which "and" functions in (**4.7**) the same way as it does in (**4.9**). Can you discover it?

Here "and", at least superficially, joins two adjectives, "pretty" and "flirtatious", to form a complex adjectival expression. Perhaps (**9**) is an abbreviation of a sentence in which "and" is genuinely a sentence connective. But what sentence? Not

10) Some girls are pretty and some girls are flirtatious.

For (**10**) does not have the same truth conditions as (**9**): (**10**) could be true, yet not (**9**), if it were the case that ugly girls, and only ugly girls, flirt.

11) John washed immediately and thoroughly.

Here "and", at least superficially, joins two adverbs, "immediately" and "thoroughly", to form a complex adverbial expression. Perhaps (**11**) is an abbreviation of a sentence in which "and" is genuinely a sentence connective. But what sentence? Not

12) John washed immediately and John washed thoroughly.

For (**12**) does not have the same truth conditions as (**11**): (**12**) would be true yet not (**11**), if John washed twice, once immediately but not thoroughly, and once thoroughly but not immediately.

(b) Cases in which "and" is a sentence connective but is, allegedly, not truth functional, and so not equivalent to "&".

A standard kind of example is:

13) Mary got married and had a baby.
14) Jane Austen died in 1817 and was buried at Winchester.

Here there is no question about "and"'s claim to be a sentence connective. (There is a slight element of abbreviation: the second component elides the name in each case.) But it is argued that "and" cannot be translated by "&". If it could be, then the truth or falsehood of (**13**) and (**14**) would depend on nothing more than the truth or falsehood of the components. The objection is that this is not so: (**13**) and (**14**) require for their truth that the event reported in the second component occur *after* that reported in the first.

From the interpretation rule for "&", we know that "X & Y" is true iff "X" is true and "Y" is true. So "X & Y" is true iff "Y & X" is true: as we shall say, "X & Y" and "Y & X" are *equivalent*. This fact is reflected in the general truth about **P**-validity:

15) $X \& Y \models_{P} Y \& X$.

This is to be read as follows: if X and Y are any **P**-sentences, the result of forming their conjunction in that order serves as a premise to a **P**-valid argument whose conclusion is the conjunction of X and Y in the other order.

If "&" expressed the same truth function as "and" in (**13**) and (**14**), the following arguments would be valid:

16) Mary got married and had a baby. So Mary had a baby and got married.
17) Jane Austen died in 1817 and was buried at Winchester. So Jane Austen was buried at Winchester and died in 1817.

Those who think that the premises do not entail the conclusion will hold that (**16**) and (**17**) are not valid, so that in these cases "and" does not express the truth function that "&" expresses.

The standard response is to say that (**16**) and (**17**) are valid, the contrary appearance being created by the fact that (**13**) *implicates*, but does not entail, that Mary got married before having the baby, and similarly for (**14**).

Ex. 2.13 (a) Are there any difficulties in interpreting "and" as it occurs in the following sentences as expressing the same truth function as "&"? If there are none, provide a **P**-formalization (showing your correspondence scheme). If there are any difficulties, explain.
(i) John and Mary bought a boat.
(ii) All elephants have short ears and long tails.
(iii) Some elephants have short ears and long tails.
(iv) You and I are the only people who matter.
(b) Evaluate the following claim:
"&" is binary, but "and" is not. For example, in
John is happy, Mary is tall and Sarah is weak
"and" makes a sentence out of three sentences.
Cf. McCawley [1981], pp. 49–54.

(3) "∨" and "or"

The nearest English equivalents to "∨" are "or" and "either . . . or . . .". These are certainly sometimes sentence connectives, as in:

18) You're a fool or you're a rascal.
19) Either you're a fool or you're a rascal.

It certainly seems that (**18**) and (**19**) are true iff at least one of "You're a fool" and "You're a rascal" is true. So there is a case for saying that "or" expresses the same truth function as "∨".

Another class of cases in which "∨" seems close to "or" is provided by a certain kind of game. "I'll give you a clue: either William hid the silver or Tom hid the gold . . . Which box contains the silver, which the gold, which the lead?"

As with "and", I divide the alleged discrepancies into those cases in which "or" (or "either . . . or . . .") is supposedly not a sentence connective, and those in which it supposedly occurs as a sentence connective which does not express the same truth function as "∨".

(a) Cases in which "or" appears not to function as a sentence connective.

20) Tom or Mary could help you.

Here "or", at least superficially, joins not two sentences to form a sentence, but two names to form a complex subject expression. (**20**) cannot be regarded as an abbreviation of

21) Tom could help you or Mary could help you,

for most people hear (**20**), but not (**21**), as meaning that Tom could help you and so could Mary.

Ex. 2.14 (4.20) does not seem to entail that Tom and Mary could both help you (simultaneously), for it could be true in a situation making it impossible for there to be more than one helper. Give an example to show this. Does this suggest a way of avoiding the unattractive suggestion that "or" in (4.20) really means "and"? Provide details.

22) Every number is odd or even

should be compared with (**9**). At least superficially, "or" here joins two adjectives, "odd" and "even", to form a complex adjectival expression. Could (**22**) be an abbreviation of a sentence in which "or" is genuinely a sentence connective? What sentence? Not

23) Every number is odd or every number is even

for this plainly means something different from (**22**). Indeed, (**22**) is true, (**23**) false.

A problematic case is:

24) He asked whether John would win or not.

We cannot understand this as meaning

25) He asked whether the following is true: John will win or John will not win.

Everyone knows that "John will win or John will not win" is true: *this* cannot have been what the questioner wanted to know. But it is not

easy to see how "or" as a sentence connective could be used to express what is being asked.

(b) Cases in which "or" is a sentence connective but is, allegedly, not equivalent to "∨".

These cases are of two kinds: (i) those in which it is agreed that "or" expresses some truth function, and the disagreement is over whether it expresses the same one as "∨"; (ii) those in which it is contended that "or" does not express a truth function.

(i) "∨" expresses what is standardly called *inclusive disjunction*. If "X" and "Y" are both true, so is "$X \vee Y$". It is sometimes claimed that "or", and, more especially "either . . . or . . .", express *exclusive disjunction*. The exclusive disjunction of X with Y is true just on condition that *exactly* one of "X" and "Y" is true. We can, of course, easily *define* a **P**-connective which expresses this function. And if "or" does, sometimes or always, express exclusive disjunction, we need, sometimes or always, to avoid matching it simply with "∨".

Ex. 2.15 Show how the exclusive disjunction of any **P**-sentences X and Y can be expressed using just the **P**-connectives already defined (in §2).

26) This number is odd or this number is even

might be offered as a candidate example of exclusive disjunction. However, the case is inconclusive. We must admit that the truth of both disjuncts is excluded, but it remains to be shown that the excluding is done by "or" rather than by the particular senses of the disjuncts, which already preclude their joint truth. This phenomenon is entirely consistent with "or" expressing inclusive disjunction.

A better example is:

27) You are welcome to come to dinner on Monday or Tuesday.

If "or" is a sentence connective here, it presumably connects "you come to dinner on Monday" and "you come to dinner on Tuesday", so that (**27**) is an abbreviation of something like:

28) I would welcome your making it true that: you come to dinner on Monday or you come to dinner on Tuesday.

But many people hear (**27**), and so, presumably, (**28**), as constituting an invitation for just one dinner, not two.[4] If this is right, there is a case for thinking that "or" sometimes expresses exclusive disjunction, and so on such occasions should not be matched simply with "∨".

Ex. 2.16 A restaurateur who puts on the menu "dessert or fruit" commits himself only to allowing you one of these, but he does not falsify his menu, or violate any undertaking to which his menu commits him, if he gives you both. Does this support the inclusive or the exclusive interpretation of "or"? Reflection on this case should, in my view, remove any appearance of exclusive disjunction in (**4.27**). Can you explain how this line of thought runs?

(ii) The most telling reason for thinking that "or" is not truth functional issues from such cases as:

29) Either the superpowers will abandon their arms race, or there will be a third world war.

The suggestion is that (**29**) requires for its truth not merely the truth of at least one (or exactly one) of its disjuncts, but in addition that there be some special connection, presumably in this case causal connection, between the falsehood of one disjunct and the truth of another. (**29**) asserts, it may be said, that the arms race will *lead to* war.

If this is right, then we ought to be able to discover failures of the substitution test (see §2). The test is rather hard to apply in such a case, because there is likely to be disagreement about whether the disjuncts are true or false. Suppose we think that (**29**) is true, but that the first disjunct is false (i.e. it is false that the superpowers will abandon their arms race). Then, if the "or" it contains is truth functional, we ought to find the following true, despite the fact that no one could for a moment suppose that there is any causal connection between the truth of the first disjunct and that of the second:

30) Either 2 + 2 = 22 or there will be a third world war.

[4] The best putative examples of exclusive disjunction tend to involve rather complex and poorly understood constructions, like the one in (**27**). It may well be that it is these constructions, rather than "or", which are responsible for any exclusivity. For an illuminating account, see Higginbotham [1988], esp. pp. 226–7.

Holding our suppositions firmly in mind (the truth of (**29**), the falsehood of its first disjunct), it actually seems rather unlikely that one should find (**30**) false. The suppositions entail that there will be a third world war; and this seems to entail (**30**). So it seems as if the result of replacing the first disjunct of (**29**) (which we are supposing to be false) by an arbitrary falsehood results in a sentence with the same truth value as (**29**). This is a partial fulfilment of the substitution test.

Ex. 2.17 What else needs to be done to complete the substitution test? Try at least one further substitution in (**4.29**), and report whether you think it consistent or inconsistent with the truth functionality of "or".

Completing the substitution test would be hard work, so one might try to find a more general argument either for or against the view that (**29**) would fail it. One might suggest that (**29**) is equivalent to

31) If the superpowers will not abandon their arms race, there will be a third world war.

If this equivalence holds generally, the question of the truth functionality of "or" reduces to the question of the truth functionality of "if".

(4) "→" and "if"

The nearest English equivalents of "→" are: "If . . . then . . ."; ". . . if . . ."; and ". . . only if . . .".[5] I will concentrate on the first of the idioms, calling "if . . . then . . ." sentences *conditionals*, the sentence filling the first blank the *antecedent*, the sentence filling the second blank the *consequent*. A common view is that conditionals cannot be adequately formalized by "→".

One ground for this view is that a sentence "if *A* then *B*" requires for its truth some special connection between what would make "*A*" true and what would make "*B*" true, a causal connection, for example. No such connection is required for the truth of "$p \rightarrow q$". Consider a volume of water which in fact will not be heated to 90° at any time in the coming year, so that the sentence "this volume of water is

[5] For a justified protest at the standard treatment of "*A* only if *B*" as an idiomatic variant of "if *A* then *B*", see McCawley [1981], pp. 49–54.

heated to 90° at some time during the coming year" is false. Let this correspond to the **P**-letter "p". Let "q" correspond to "this volume of water will turn to ice". It follows that

32) $p \rightarrow q$

is true upon an intended interpretation, whereas, so the claim runs,

33) If this volume of water is heated to 90° at some time during the coming year, this volume of water will turn to ice

is false. (**32**) is an inadequate formalization, since "if" does not express the same truth function as "$\rightarrow$".

Ex. 2.18 Why is (**4.32**) true? (Refer back to the definition of "$\rightarrow$" at (**1.4**).)

Ex. 2.19 It might be suggested that the following example shows that "if" sometimes expresses a truth function, but not that of "$\rightarrow$":

If you need bandages, there are some in the first aid box.

Here, according to the suggestion, the conditional is true iff its consequent is. Discuss the suggestion.

The question of the relation between "$\rightarrow$" and "if" has been very widely discussed. I shall divide up the issues as follows. In the remaining part of this section, I will try to demarcate the area of the controversy: it concerns "indicative" rather than "subjunctive" conditionals, but it is not easy to give adequate criteria for this distinction. In subsequent sections, I present the case against the truth functional interpretation (§**5**), the outline of an implicature defence against this case (§§**6, 7**), and finally (§**8**) some general arguments in favour of the truth functional interpretation. The issue is not resolved in these discussions, and the question is taken up again in chapter **3**, and again in chapter **5.2**.

Consider the following two "if" sentences:

34) If Oswald didn't shoot Kennedy, someone else did.

35) If Oswald hadn't shot Kennedy, someone else would have.

It would be perfectly reasonable to regard one of these sentences as true and the other false. For example, you might reasonably believe (**34**), simply on the ground that someone did shoot Kennedy (he wasn't poisoned, etc.). Yet you might not believe (**35**), and would not if you thought that Oswald was a maniac working alone, and that no one else would have wanted Kennedy dead. The fact that (**34**) and (**35**) may diverge in truth value encourages the view that we cannot have a uniform account of the two kinds of sentence: we will need one account of the way in which "if" works in *indicative conditionals*, like (**34**), another of the way in which it works in *subjunctive conditionals* like (**35**) (see chapter **5.2**). The target of our discussion is the claim that indicative (rather than subjunctive) conditionals can be adequately formalized by material conditionals, those dominated by "$\rightarrow$".

The attempt to formalize (**35**) as a material conditional leads to an instructive kind of nonsense:

36) (Oswald hadn't shot Kennedy) $\rightarrow$ (someone else would have shot Kennedy).

We can extend the definitions of the **P**-connectives in the obvious way to allow them to stand between English sentences, A and B: "$A \rightarrow B$" is true iff either "A" is false or "B" is true. A sentence connective connects only sentences which are capable of having truth values; if it is a truth functional sentence connective, these are the values that are inputs to the truth function, and the sentences must possess them independently of their role in some putatively truth functional context. The components of (**36**) do not have this property. "Someone else would have shot Kennedy" arguably has no self-standing use at all, and so we cannot talk of its truth value. "Oswald hadn't shot Kennedy" does have a self-standing use, for example in a chronicle of Dallas: "At that time, Dallas was a peaceful city. The strains of unregulated growth had not become apparent, and Oswald hadn't shot Kennedy." This use is plainly not what is at issue in (**36**), which does not in that way carry an implicit reference to a past time. "Oswald hadn't shot Kennedy", as used in (**36**), cannot be used self-standingly. This means that (**36**) is not really intelligible, and that there can be no question of treating (**35**) on the pattern of (**36**).

There are other ways in which, even where the mood is not explicitly subjunctive, the English conditional does not link two sentences that are capable of being true or false when standing alone, so that "if" is not functioning as a sentence connective. One kind of example will be familiar from (**9**) and (**22**):

37) If someone is in debt, he should curb his expenditure.

Here the consequent "he should curb his expenditure" is not a complete sentence, since the referent of "he" is not determined. Hence the consequent is not capable, on its own, of being true or false. Hence we cannot translate (**37**) as

38) (Someone is in debt) → (he should curb his expenditure).

Ex. 2.20 What, if anything, would be wrong with the translation: (someone is in debt) → (someone should curb his expenditure)?

There are also conditionals which it is not obviously correct to classify as indicative conditionals, even though their components are in the indicative mood, and appear capable of being used self-standingly. For example, someone contemplating the future might truly affirm:

39) If John dies before Joan, she will inherit the lot.

The components are capable of being evaluated for truth and falsehood as self-standing sentences. However, it also seems that (**39**) is equivalent to

40) If John should die before Joan, she would inherit the lot.

This means that there is a case for saying that (**39**), despite being grammatically indicative, is best classified as a subjunctive rather than as an indicative conditional (cf. chapter **5.2**).

There are other uses of "if" which we would not wish even to try to formalize using "→", for example

41) John wonders if his life is meaningful.

It is literally true that each of "John wonders" and "John's life is meaningful" are capable of being evaluated for truth and falsehood as self-standing sentences, but it should also be clear that their use as self-standing sentences differs from their use in (**41**). This could be brought out by sentences similar to (**41**), for example

42) John wanted to know if he had been born in wedlock.

It is not that this is a subjunctive conditional. It is not a conditional at all. Rather, "if" in (**42**) is being used to form an indirect question, and could be replaced by "whether".

5 The case against the material implication account of "if"

The view to be attacked in this section is that we can adequately formalize (indicative) conditionals by "→" -sentences; equivalently, that conditionals are **P**-material implications; equivalently, that "if" (as it occurs in the cases under discussion) expresses the same truth function as "→". As a preliminary, let us see if we can establish the following:

1) Any truth functional occurrence of "if . . . then . . ." expresses the same truth function as "→".

(The phrasing allows for the possibility of non-truth functional occurrences of "if . . . then . . .".) If we can establish (**1**), then the question whether conditionals are material conditionals becomes the question whether "if", as it occurs in English conditionals, is truth functional.

If "if . . . then . . ." were truth functional, and functioned in the same way on all its occurrences, there would be a simple way to establish (**1**). It would be enough to find "if . . . then . . ." sentences whose truth values, in point of their components and resultant, match the truth table for "→". For example, suppose we find a true "if . . . then . . ." sentence with false antecedent and true consequent, perhaps "if New York is south of Beirut, then New York is south of Paris". The supposition

Table 2.2

X	Y	X → Y
T	T	T
T	F	F
F	T	T
F	F	T

that "if . . . then . . ." is always truth functional would enable us to infer that *any* "if . . . then . . ." sentence with false antecedent and true consequent is true, which corresponds to the third line of the truth table for "→" (table 2.2). It is not hard to do the same for the remaining three lines of the table. The result would be a correct table in which "if A then B" could replace $X \rightarrow Y$.

Ex. 2.21 For each of the first two and the last line of table 2.2, find a conditional which has the truth value indicated in the last column, and whose antecedent and consequent are well known to have the truth values indicated in the first two columns.

This approach assumes that "if . . . then . . ." makes the same contribution on every occurrence. If it were ambiguous, having one meaning in one of our sample sentences, another in another, we would not have shown that on any of its meanings it expresses the "→" truth function. Suppose one of our sample sentences were "If you need bandages, there are some in the first aid box" (cf. Ex. 2.19). It might be argued that on this occurrence it expresses a different truth function from the one it more typically expresses, on the grounds that this sentence is true iff its consequent is.

Here is an argument which goes some way towards meeting this difficulty. Suppose some arbitrary sentence "if A then B" is truth functional with respect to A and B. Suppose also that:

2) Using "if . . . then . . ." as it occurs in "if A then B", every instance of "if A and B then A" is true.

3) "and" expresses the same truth function as "&".

4) Using "if . . . then . . ." as it occurs in "if A then B", there are falsehoods of this form.

We can argue as follows, starting from (**2**). Take an instance of "if A and B then A" in which A is false. Then "A & B" is also false (by (**3**)). So by the assumption of truth functionality, every instance of "if . . . then . . ." with false antecedent and false consequent is true, establishing the fourth line of a truth table for "if . . . then . . .". Now suppose that A is true. There are two subcases. On one of them, B is true, so "if A and B then A" is a truth with true antecedent and true consequent, which, given truth functionality, establishes the first line of the table. One the other subcase, B is false, so "if A and B then A" is a truth with false antecedent and true consequent, which, given truth functionality, establishes the third line of the table. The second line of the table is then established by (**4**). So given (**2**)–(**4**), we can establish (**1**). The upshot is that the question whether (some occurrence of) "if . . . then . . ." expresses the same truth function as "→" reduces to the question whether (that occurrence of) "if . . . then . . ." is truth functional, that is, expresses some truth function or other. In this section, I present the case for a negative answer to this question.

(a) The falsehood of X upon an interpretation is enough for the truth of $X \rightarrow Y$ upon that interpretation. But the falsehood of the antecedent is not enough for the truth of an indicative conditional formed with "if" or "if . . . then". Examples:

5) If ice is denser than water, then ice floats on water

6) If ice does not float on water, then ice floats on water

are usually held to be false. But, using "p" to correspond to "ice is denser than water", and "q" to correspond to "ice floats on water", "$p \rightarrow q$" and "$\neg q \rightarrow q$" are both true upon an intended interpretation, viz. one in which the truth values of the **P**-letters match those of their corresponding English sentences, viz. one which assigns false to "p" and true to "q".

This is connected with a fact about validity. In **P** we have:

7) $X \models_P [\neg X \rightarrow Y]$, whatever Y may be.

Ex. 2.22 Show that **(5.7)** in true.

This says that any **P**-sentence constitutes the premise of a **P**-valid argument for any **P**-conditional having the negation of that sentence as antecedent. But, it is claimed, the following is false for English:

8) $A \models$ [if not-A then B], whatever B may be.

That is, the fact corresponding to **(7)** does not obtain for English. You cannot always validly infer from an English sentence to an arbitrary indicative conditional having the negation of that sentence as its antecedent. If this contrast is correct, then "if" does not make the same contribution to validity as "→" does; hence it does not make the same contribution to truth conditions; in particular, it is not truth functional.

(b) An interpretation upon which Y is true is one upon which $X \rightarrow Y$ is true. But the truth of the consequent is not enough for the truth of an indicative conditional formed with "if" or "if . . . then". An example is:

9) If ice is as dense as lead, then ice floats on water.

A connected claim is that whereas

10) $Y \models_{\mathbf{P}} [X \rightarrow Y]$, whatever X may be,

it is not the case that

11) $B \models$ [if A then B], whatever A may be.

Ex. 2.23 (a) To serve as an example, **(5.9)** needs to be false. Put the case for the view that it is true (even if you don't accept this view!).
(b) Can you think of a convincing example to show the falsehood of **(5.11)**?
(c) Suppose you believe that there will be a third world war whether or not there is a summit meeting next spring. Suppose you also firmly believe that there will be no summit meeting next spring. Should you accept or reject the conditional "If there is a summit meeting next spring, there will be a third world war"? What bearing does your answer have upon whether **(5.11)** is true?

(c) Whereas

12) $\neg(X \rightarrow Y) \models_{\mathbf{P}} X$,

it is not the case that

13) [it is not the case that (if A then B)] $\models A$.

For example, it is claimed that the following is invalid:

14) It is not the case that if the number three is even then it is prime. So the number three is event.

Ex. 2.24 Which answer to Ex. 2.23(c) undermines the claim that **(5.14)** is invalid? Explain your answer.

(d) Whereas

15) $\neg(X \rightarrow Y) \models_{\mathbf{P}} \neg Y$,

it is not the case that

16) [it is not the case that (if A then B)] $\models$ it is not the case that B.

For example, it is claimed that the following is invalid:

17) It is not the case that, if I go to the party tonight, I shall get drunk tonight. So I shall not get drunk tonight.

(e) Whereas

18) $(X \rightarrow \neg Y), Y \models_{\mathbf{P}} \neg X$, it is not the case that

19) [if A then not-B, B] $\models$ not-A.

For example, it is claimed that the following is invalid:

20) If it rains, then it will not rain heavily. It will rain heavily. So it will not rain.

(f) "If" is not transitive, whereas "$\rightarrow$" is. By the *transitivity* of "$\rightarrow$" is meant the following:

21) $X \rightarrow Y, Y \rightarrow Z \models_P X \rightarrow Z$.

The non-transitivity of "if" can be analogously represented by the claim

22) if A then B, if B then $C \not\models$ if A then C.

This is said to be established by the invalidity of arguments like:

23) If Smith dies before the election, Jones will win.
If Jones wins, Smith will retire from public life after the election.
So, if Smith dies before the election, he will retire from public life after the election.

(g) Whereas

24) $(X \& Y) \rightarrow Z, X \models_P Y \rightarrow Z$,

the alleged invalidity of the following example purportedly shows that the analogous fact does not hold for English:

25) If this room is getting warmer and the mean kinetic energy of the molecules of its contents remains the same, then the scientific world will be astonished.
This room is getting warmer.
So if the mean kinetic energy of the molecules of this room's contents remains the same, then the scientific world will be astonished.

Ex. 2.25 (a) Explain how one who thinks that (**5.25**) is invalid could best justify his position.
(b) Explain how the following argument could be used to strengthen the case against the truth functional interpretation of "if".
If John is in Paris, then he is in France.
If he is in Istanbul, then he is in Turkey.
So if he is in Paris, he is in Turkey, or if he is in Istanbul, he is in France.

In other words, it is claimed that

26) [if (A and B) then C, A] $\not\models$ if B then C.

(h) Whereas

27) $(X \rightarrow Y) \models_P (X \& Z) \rightarrow Y$, whatever Z may be,

the alleged invalidity of the following example purportedly shows that the analogous fact does not hold of English:

28) If I put sugar in this cup of tea it will taste fine.
So, if I put sugar and also diesel oil in this cup of tea, it will taste fine.

In other words, the claim is that

29) [if A then B] $\not\models$ [if (A and C) then B], whatever C may be.

(i) Whereas

30) $\models_P (X \rightarrow Y) \vee (Y \rightarrow X)$,

it is not the case that

31) $\models$ Either (if A then B) or (if B then A).

Suppose that Peter says that there will be a third world war and Quentin denies this. Then it seems that both "If Peter is right, then so is Quentin" and also "If Quentin is right then so is Peter" are false, so that substituting these sentences in (**31**) leads to a falsehood (cf. Read [1988], pp. 23–6).

One who would defend the view that "if" or "if . . . then . . ." are truth functional against these charges will need some impressive resources. They will include the notion of implicature, mentioned, but not yet discussed.

I offer no separate discussion of the correspondence between "↔" and "if and only if".

6 Implicature

How is one to tell whether, for example

1) Jane Austen was buried at Winchester and died in 1817

is true or not (cf. **4.17**))? Clearly it would help to know where and when Jane Austen died. There can be no general answer to how we know things like this, but let's pretend we do know that (**1**) correctly specifies the place and date of Jane Austen's death. It is debatable whether this is enough to answer the question of whether (**1**) is true.

One way to proceed is to imagine someone addressing this remark to us. The thought irresistibly presents itself that there is something wrong. But does what is wrong consist in the remark failing to be true?

There are all sorts of ways in which a remark can sound "wrong" even if it is true; and sound "right" even if it is false. On being introduced to an ugly person, it would be wrong – morally wrong – to say

2) You are without question the ugliest person I have ever met,

even if (**2**), as uttered in the circumstances, would be a literal truth. Now suppose that someone enquires whether you know anything about birds, and you wish to convey, in as graphic a fashion as possible, that you know next to nothing. You say

3) I can't tell a crane from a canary.

This, we shall suppose, is not true. Still, it can be a perfectly proper thing to say – a permissible, and readily understood, exaggeration. No one would take you quite literally.

The conclusion is that the truth of a remark is neither necessary nor sufficient for it being a right and proper one.

It is possible to *convey* something without strictly and literally *saying* it. By uttering (**2**) you convey that you care nothing for the other person's feelings, though you do not say this. What is right about (**3**) is that it correctly conveys that you know next to nothing about birds.

The distinction is made even more graphically in a famous example. H asks S his opinion about Jones's qualities as a philosopher. (Jones is a student of S's, and H has every right to know S's opinion.) S replies:

4) Jones has beautiful handwriting

and says nothing further. S conveys that Jones is a bad philosopher, but he does not say this.

Ex. 2.26 Say which if any of the following could have been used by S to convey his unflattering opinion:
(i) Either he is a good philosopher or he isn't.
(ii) He has a beautiful wife.
(iii) He has appalling handwriting.
(iv) His handwriting is average.
In general, what are the necessary and sufficient conditions for an utterance to be usable by S, in this context, to convey his low opinion of Jones?

I once asked Gilbert Ryle whether he liked music. He replied

5) I can tell the difference between loud and soft.

Ryle conveyed that he did not like music; but this is not what he said.

In *Can You Forgive Her?* Mr Bott says to Alice Vavasor:

6) The frost was so uncommonly severe that any delicate person like Lady Glencowrer must have suffered in remaining out so long. (Trollope [1864], ch. 28)

Alice knew that Mr Bott knew that Alice had been out with Lady Glencora, yet Mr Bott had made no enquiry about Alice's health.

According to Trollope, Mr Bott thereby conveyed that Alice was not delicate (and so not upper class). But he did not say this. What is conveyed may be false even when what is said is true.

The nurse, beaming brightly, announces to the newly delivered mother

7) Congratulations! It's a baby.

The nurse conveys (presumably in jest) that there was some significant possibility of the mother having given birth to something other than a baby. But she does not say this.

Ex. 2.27 Can one say that *A* and thereby convey that ¬*A*? If so, give an example. If not, why not?

H. P. Grice coined the phrase "implicature" to apply to what, in such examples, is conveyed but not said.[6] A truth may implicate something false. (**7**) would normally be an example of this, but any of the other cases might be examples, in appropriately altered circumstances. This possibility could be exploited by someone wanting to defend a truth functional interpretation of (**1**). On this view, an utterer of (**1**) may implicate that the burial occurred before the death, and this may be false; but all that (**1**) strictly and literally says is that the two events occurred, so it is true. Hence it is not after all a counterexample to the truth functionality of "and".

Grice claimed that what he called "*cancellability*" is a mark of implicature, and will help us differentiate what belongs to implicature and what belongs to strict and literal saying.[7] If *A* entails *B*, then "*A*, but not *B*" will be contradictory. Thus "Tom is a bachelor" entails "Tom is

[6] Grice's phrase was "conversational implicature". He distinguished this from "conventional implicature". Arguably, some of the implicatures associated with the assertion of conditionals (though not, I think, conjunctions) should count as conventional rather than conversational. (See Jackson [1981].) To allow room for this alternative (which I do not explicitly discuss), I have mostly dropped the qualifier "conversational".

[7] In Grice, cancellability is a sign of conversational, but not of conventional, implicature. See previous note. For example, Grice suggested that "and" and "but" agree in expressing the same truth function, but differ in that "but" has some kind of contrastive *conventional* implicature (see also below, §9), which is, accordingly, not cancellable.

unmarried", and "Tom is a bachelor, but he's not unmarried" is contradictory. This shows that Tom has to be unmarried, if what is strictly and literally said by "Tom is a bachelor" is to be true.

By contrast, the following, uttered in the situation we envisaged for (**4**), is perfectly consistent:

8) Jones has beautiful handwriting – though I don't mean to suggest that he is other than an excellent philosopher.

This shows that Jones's being a bad philosopher, or being thought by *S* to be such, is not a necessary condition for the truth of (**4**), even as uttered in the special kind of context in question.

A defender of the truth functional interpretation of "and" can allow that in some sentences in which "and" occurs, something more like "and then" is implicated. But cancellability will show that it belongs to implicature only, and not to what is strictly and literally said. For example, in the circumstances envisaged for (**4.13**), one could quite consistently have said:

9) Mary got married and had a baby, but I'm not willing to pronounce on the correct order of these occurrences.

In place of (**1**) it would be quite consistent to say

10) Jane Austen died in 1817 and was buried at Winchester, but I'm not saying which happened first.

(A plausible conversational background for this case would be one in which the general topic of discussion is cases of people being buried alive. One wishes to make clear that one is carefully remaining neutral on the question of whether Jane Austen was an example.)

The phenomena we have discussed in connection with "and" can be reproduced merely by the use of separate sentences. There is little to choose between (**4.13**) and

11) Mary got married. She had a baby.

Putting things in this order might well implicate that the corresponding events occurred in that order. But it would be absurd to suggest that either of the sentences, or the utterance as a whole, would be false (as opposed to, say, misleading) if the birth preceded the wedding. That the same phenomenon can arise in the absence of "and", and in a context in which it cannot be attributed to the truth conditions of any expressions involved, suggests that it should not be attributed to the truth conditions of "and".

Let us now resume the essential features of an implicature defence of a truth functional interpretation, applying it to the interpretation of "and" as expressing the truth function expressed by "&".

The objector claims that there are cases in which a compound has a truth value inconsistent with the truth functional interpretation. In the present case, the allegation is that conjunctions can fail to be true, even when both conjuncts are.

The defence consists in saying that a sentence which is strictly and literally true may have false implicatures. The objector, it will be claimed, has mistaken the falsity of an implicature for the falsity of the sentence itself – that is, for the falsity of what the sentence strictly and literally says. So the examples allegedly of false conjunctions with true conjuncts are really examples of true conjunctions which have false implicatures.

This summary involves some oversimplifications, which would need to be addressed by a full-scale defence. As we have seen, it cannot be quite right to say that *sentences* have implicatures: one test for an implicature is its cancellability, which shows that what is implicated does not belong to the content of the sentence as such. Implicature is determined by conversational context, and so is not a stable feature of the sentence, but a feature of its use on a specific occasion, within a given dialogue. As Grice presents the notion of implicature, the kind of use liable to generate implicature is a use (typically assertive) of the whole sentence; a more cautious deployment of an implicature defence would have to take into account cases in which the sentence in question is not asserted, for example, when the sentence occurs within a larger compound. Granting that "Mary had a baby and got married", used assertively on a specific occasion, may be true yet generate a false implicature, a story needs to be told about how something similar can happen when the sentence is not used assertively. For example, the following pair might be claimed to differ in truth value:

12) If Mary had a baby and got married, her parents must have been ashamed.

13) If Mary got married and had a baby, her parents must have been ashamed.

In (**12**), "Mary had a baby and got married" is only a part of the larger sentence which is used in the whole speech-act: the conjunction itself is not used assertively, and so cannot generate implicature in the way Grice typically envisaged. Yet if (**12**) and (**13**) differ in truth value, the natural explanation is that their antecedents can differ in truth value. The implicature defence would need additional resources to deal with this argument for the claim that "and", used as a sentence connective, does not always express the "&" truth function.

Despite these difficulties, my own view is that "and", used as a sentence connective, does express the same truth function as "&". The relation between "if" and "→" is more complex.

7 "If": implicature in defence of the truth functional interpretation

I envisage two strategies at the disposal of the defender of the truth functional interpretation of "if". The first is to exploit the notion of implicature to try to defuse alleged counterexamples. The second strategy, deferred until the next section, is to provide direct arguments for the interpretation.

A defence of the truth functional interpretation of "if" requires a more systematic account of implicature. Consider an utterance which, though true, would give rise to a false implicature. For example, suppose that Jones really does have beautiful handwriting, and that you utter (**6.4**) assertively in the envisaged circumstances. It will be entirely reasonable for your hearer to infer that you think ill of Jones as a philosopher. Suppose that in fact you think highly of him. Then, in uttering (**6.4**) you have spoken truly, but misleadingly. You *should not* have asserted (**6.4**). In the context, (**6.4**), though true, had (as I shall put it) a *low degree of assertibility*. One goes some way to providing an account of implicature by giving principles which determine features of an utterance which raise or lower its degree of assertibility.

One relevant feature (suggested by Grice) is that an utterance is more assertible the more informative it is, relative to the needs of the conversation. For example, if you are asked where Tom is and you know he is in the library, you should not reply

1) He is either in the library or in a lecture

despite the fact that (**1**) is true. (**1**) is less assertible, because less informative, than the equally true

2) He is in the library.

The principle is:

3) The more informative a true utterance is (relative to the conversational needs in question), the more assertible it is.

Ex. 2.28 What would be wrong with (**7.3**) without the parenthetical qualification?

A consequence of this principle is that very uninformative true utterances will normally have very low assertibility.

Ex. 2.29 Can you suggest an example of something which is rather uninformative but highly assertible? If so, could (**7.3**) be amended to allow for your example? If not, is there a way of showing that there could be no such example?

Let us apply this to the problem for the truth functional account of "if" posed by (**5.10**) and (**5.11**). Consider the argument

4) Ice floats on water. So, if ice is denser than water, it floats on water,

and the allegation that this argument is invalid (having true premise but false conclusion), whereas formalizing "if" by "→" yields a **P**-valid argument. Then the defence of the truth functional interpretation using (**3**) claims that the conclusion is not really false, and the argument is not invalid. If we think the conclusion is false, it is because we imagine it being asserted in the context of the propounding of

an argument like (**4**). In such a context, the premise is much more informative than the conclusion, which is very uninformative. Anyone in a position to use the argument must be in a position to assert its premise. So (**3**) has it that, as conclusion of the argument, "if ice is denser than water, it floats on water" is highly unassertible. If anyone wrongly thinks it is false, it is because they confuse low assertibility with falsehood.

One problem with this account is that it is open to question whether (**3**) is true. But there is also a more internal problem. Consider

5) Ice floats on water. So, if ice does not float on water, it floats on water.

Here, again, we have an apparent counterexample to the truth functional interpretation of "if", for we have an argument which is intuitively invalid, having true premise but false conclusion, but which would be **P**-valid were "if" formalized by "→".

However, the same defence cannot be brought to bear on this case as on (**4**). For:

6) $\models_P X \leftrightarrow (\neg X \rightarrow X)$.

In other words, any sentence is equivalent to the material conditional having it as consequent and its negation as antecedent. Hence in one good sense of what it is for two sentences to be equally informative, a sense which the *proponent* of the truth functional interpretation of "if" cannot very well despise, the premise and conclusion of (**5**) are equally informative. Hence the defence of the truth functional interpretation which relies on (**3**) cannot be applied here: the conclusion of (**5**), if it is as informative as the premise, ought to be no less assertible. This casts doubt on whether the defence of (**4**) was legitimate. (**4**) and (**5**) appear to pose a common problem, requiring a single solution.

An alternative defence is based on the following principle:

7) The utterer of a conditional implicates that he has good grounds, of certain standard kinds, for his utterance.

Standardly, a conditional is used only when the truth or falsity of the components is not known. One standard kind of ground is that one has good evidence for a generalization of which the conditional is an instance. For example, one may well know, from general principles governing floating and density, that anything denser than water will sink in water. You may be ignorant of the density of some particular substance, say bakelite, relative to water. Still, you know from the general principle that if bakelite is denser than water, it will sink in water. The defence of the truth functional interpretation based on **(7)** has it that many of the apparent anomalies can be explained in terms of the falsity of implicatures relating to the grounds envisaged for the conditional.

For example, **(4.33)**, viz.:

> If this volume of water is heated to 90° at some time during the coming year, this volume of water will turn to ice

will be held to implicate that there is, or that the speaker believes there is, some general connection between heating water and it turning to ice. Given that there is no such connection, the implicature is false. When people allege that **(4.33)** is false, they are responding to the falsity of the implicature. On the assumption that this volume of water will not be heated to 90° at any time during the coming year, what **(4.33)** strictly and literally says is true, or so the defender of the truth functional interpretation will urge. The different approaches are marked by the fact that those who are confident of the falsity of **(4.33)**, unlike the defender of the truth functional interpretation, feel no need to find out whether or not its antecedent is true.

The style of defence has some plausibility for a number of the alleged counterexamples to truth functionality given in §**5**. For example, it says something relevant to **(5.14)**, viz.:

> It is not the case that if the number three is even then it is prime. So the number three is even.

To serve as a counterexample, it is necessary that the premise be true. For this, it is necessary that

8) if the number three is even then it is prime

be false. But, the defence goes, what we are responding to when we think **(8)** is false is merely the false implicature that being even is in general a sufficient condition for being prime. This is only an implicature: what **(8)** strictly and literally says is true.

There are counterexamples about which **(7)** has little to say. It is hard to see how it could even address the problem posed by **(5.23)**, viz.:

> If Smith dies before the election, Jones will win.
> If Jones wins, Smith will retire from public life after the election.
> So, if Smith dies before the election, he will retire from public life after the election.

There may well be appropriate general grounds for the premises: for the first, that it is a two-horse race; for the second, the honourable Smith's sincere declarations. It seems that only a determination to defend a truth functional interpretation come what may would lead one to accept that the conclusion is, despite appearances, true. Likewise, **(7)** would appear unable to address **(5.20)**, viz.:

> If it rains, then it will not rain heavily. It will rain heavily. So it will not rain.

And it has nothing to say about the sentence discussed in connection with **(5.30)**:

> Either, if Peter is right then so is Quentin, or, if Quentin is right then so is Peter.

Ex. 2.30 Evaluate the following argument in favour of the claim that "if Smith dies before the election, he will retire from public life after the election" is false:

> The truth of the antecedent would make the truth of the consequent impossible, and it is hard to see how more than this could be required for the conditional as a whole to be false.

Here is one further principle that might serve the cause of defending a truth functional interpretation of "if".

9) It is conveyed (but not said) that a conditional is "*robust with respect to its antecedent*".

One principal use of conditionals is in *modus ponens* arguments: ones having the form

10) If A, then B
A
Therefore B.

For this use to be possible, it must be possible to hold both to the conditional, "if A then B", and to its antecedent, "A". This means that it has not to be the case that were one to come to have evidence for "A", one's evidence for "if A then B' would thereby be undermined: in short, "if A then B" must be robust with respect to "A". If a conditional were not robust with respect to its antecedent, then one could never use it in a modus ponens argument, for one could not have a body of evidence which would support both of the needed premises. (Cf. Jackson [1979], [1987]; Lewis [1986b] – postscript to Lewis [1976].)

Here is an example of failure of robustness in a different connection. Suppose in January I am convinced that:

11) John will finish his book by April, or at any rate by May.

If I learn in June that the book is still unfinished, and hence infer

12) John did not finish his book by May,

I do not combine (**11**) and (**12**) and go on to infer

13) John will finish his book by April.

Rather, I abandon (**11**). We can express the phenomenon by saying that (**11**) is not robust with respect to the falsity of its second disjunct.

Let us apply (**9**) to the allegedly invalid pattern of argument (**5.8**):

$A \models$ [if not-A then B], whatever B may be.

If I used this to establish a conditional, it would not be robust with respect to the antecedent. For to use the argument in this way depends upon having good evidence for "A". Hence, subsequently acquiring evidence for the antecedent of the conditional would lead me to abandon the conditional, rather than use it to infer B. We are invited to conclude that arguments of this kind are valid, though as the conclusions will be very fragile indeed with respect to their antecedents, by principle (**9**) they will have very low assertibility. Confusing low assertibility with falsehood, we wrongly think the conclusions are false, and so think that the pattern of argument is invalid.

Ex. 2.31 Jackson has argued that one should sometimes assert something weaker in the interests of robustness. Suppose "A & B" is something of which you are sure and is relevant to the conversational needs in question, but that all you really care about is that your audience believe A. Show how the demands of robustness will conflict with the requirement of being maximally informative.

An interesting application of (**9**) is to (**5.20**). However, it is far from clear that (**9**) could deal with all the alleged counterexamples: (**5.23**) and (**5.30**) again appear to be resistant.

Ex. 2.32 Apply (7.9) to (5.20).

Perhaps what is needed is some combination of the principles we have discussed, supplemented with further principles at well. But you may be impatient: isn't all this rather ad hoc? And in any case, how convincing is the crucial claim that we mistake low assertibility for falsehood? Let us ask ourselves not whether we would be happy to *assert* the sentences having low assertibility but whether we would happily *believe* what they say. How could assertibility enter into belief? Yet we are inclined to suppose, for example, that even if we believe that A it does not follow that we are in error if we fail to believe that if not-A then B, with respect to every B. It is not obvious that the case against the truth functional interpretation essentially involves thinking of particular assertive utterances of conditional sentences, as opposed to thinking of how things are, and ought to be, with conditional beliefs.

It is hard to find neutral ground from which to assess these issues. If they can be treated adequately, it will be on the basis of comparing the truth functional account of indicative conditionals with alternatives. I turn now to the second strategy available to the defender of the truth functionality of "if".

Ex. 2.33 What is the best response a defender of the truth functional interpretation can make to the claims related to **(5.17)**, **(5.23)**, **(5.25)**, **(5.28)** and **(5.30)**?

8 "If": direct arguments for the truth functional interpretation

I shall consider two direct arguments for the truth functionality of "if". They both proceed by making claims about the validity of argument patterns involving "if", differing only in which argument patterns they select.

(1) The first argument

The following argument appears valid:

1) Either the butler or the gardener did it. Therefore if the gardener didn't do it, the butler did.

Suppose, as is natural, that the first premise is equivalent to

2) (the butler did it) $\vee$ (the gardener did it).

Then it is equivalent to

3) (the gardener didn't do it) $\rightarrow$ (the butler did it).

Replacing the premise of (**1**) by (**3**), the following ought also to be valid:

4) (the gardener didn't do it) $\rightarrow$ (the butler did it), therefore if the gardener didn't do it, the butler did.

Table 2.3

A	B	If A then B
T	T	
T	F	F
F	T	
F	F	

Assuming that this holds generally as a matter of form, and is independent of the particular subject matter of gardeners and butlers, then we can conclude

5) Any sentence of the form "$A \rightarrow B$" entails the corresponding sentence "if A then B".

It is generally accepted, and seems in any case obvious, that

6) Any sentence of the form "if A then B" entails the corresponding sentence "$A \rightarrow B$".

(**5**) and (**6**) together ensure that "$\rightarrow$" and "if" are equivalent, and thus they ensure the truth functionality of "if".

The most promising way to avert the conclusion of this argument is to deny the validity of (**1**) (cf. chapters **3** and **5.2** for views which would underwrite this denial).

(2) The second argument

First premise:

7) A, if A then $B \models B$.

This is the principle of modus ponens, and is generally uncontested. It ensures that a true antecedent and a false consequent is enough for a false conditional. If we imagine trying to construct a truth table for "if" on the lines of the truth table for "$\rightarrow$", this fact secures the correctness of the second line of the truth table (see table 2.3).

Second premise:

8) If $[A_1, \dots A_n, B \models C]$ then $[A_1, \dots A_n \models \text{if } B \text{ then } C]$.

This is a more controversial premise. It says, in effect, that if we have a valid argument for a conclusion, any one of the premises can be dropped, provided that it is made the antecedent of the conditional whose consequent is the original conclusion; and then the reduced premises will entail the new, conditional, conclusion. Let us take this principle, which might be called that of "*conditional proof*", on trust for the moment, and show how it would establish the remaining properties needed for the truth functionality of "if". Later we shall see how one might argue for the principle of conditional proof itself.

Let us recall two results from chapter **1**: (**1.6.5**), viz.:

If $[A_1, \dots A_n \models C]$ then $[A_1, \dots A_n, B \models C]$, whatever B may be

and (**1.6.7**), viz.:

If C is among the $A_1, \dots A_n$, then $[A_1, \dots A_n \models C]$.

As a special case of the latter, we have

9) $B \models B$.

So, using (**1.6.5**), we have

10) $B, A \models B$.

So, by (**8**), we have

11) $B \models$ if A then B.

This shows that a true consequent is sufficient for a true conditional, establishing the first and third lines of the table. Our table now becomes table 2.4.

Ex. 2.34 Explain how (**8.11**) can be used to establish that a true consequent is sufficient for a true conditional.

Table 2.4

A	B	If A then B
T	T	T
T	F	F
F	T	T
F	F	

Table 2.5

A	B	If A then B
T	T	T
T	F	F
F	T	T
F	F	T

Another result from chapter **1** is (**1.6.8**), viz.:

If $[(A_1, \dots A_n) \models]$, then $[A_1, \dots A_n \models B]$, whatever B may be.

This says that if the premises of an argument are inconsistent, then the argument is valid, no matter what its conclusion is. Hence we have:

12) B, not-$B \models A$.

So by (**8**):

13) $B \models$ if not-B then A.

This establishes that the falsity of the antecedent is enough for the truth of a conditional, and so establishes the third (again!) and fourth lines of the table. We now have the complete table for "if A then B" (table 2.5). We can conclude that (**7**) and (**8**) between them entail that English conditionals expressed by "if . . . then . . ." are truth functional, and express the same truth function as "$\rightarrow$" (cf. Hanson [1991]). The question now is whether (**7**) and (**8**) themselves can be justified.

(7), expressing modus ponens, is hardly ever questioned, so I shall assume that we can accept it without further ado. (8), the principle of conditional proof, is more controversial, so let us see if we can find an argument for it. It will be easier if we abbreviate "$A_1 \ldots A_n$" as "A", and so write (8) as follows:

14) if [A, $B \models C$], then [$A \models$ if B then C].

The antecedent of this conditional means (using an equivalent of the definition of "$\models$" in chapter **1.3**):

15) It is logically necessary that, if A and B are true, then so is C.

So:

16) It is logically necessary that, if A is true, then, if B is true, so is C.

So:

17) It is logically necessary that, if A is true, then if B then C is true.

But this means:

18) $A \models$ if B then C. QED.

The principles of reasoning involved seem to be modest. But they have been firmly resisted by theorists who contest the truth functional interpretation of "if". ~~Until~~ we examine alternative theories of "if", our conclusion has to be simply that such theories will have to provide good reasons for thinking that the reasoning **(14)** to **(17)** above is unsound.

Ex. 2.35 Which step in **(8.14)** to **(8.17)**, if any, do you regard as suspect, and why?

We can summarize the conclusions of §**5** and §**8** as follows: there are apparently compelling arguments for the conclusion that "if" is not truth functional, and apparently compelling arguments for the conclusion that it is. Appearances must deceive. Some attempts to say how they deceive are considered in chapter **3**.

In considering the use of **P**-validity in the investigation of validity in English, we will explore the consequences both of "if" being truth functional and of it not being truth functional.

9 Non-truth functionality in English

To deepen understanding of claims to the effect that some English expression is a truth functional sentence connective, let us examine claims to the effect that an expression is a sentence connective, but one which is not truth functional.

(1) John believes that

This expression seems to be syntactically like "not": it seems to be a *unary sentence connective*, taking a sentence like "the earth is flat" to form a new sentence, "John believes that the earth is flat". So the expression appears to function as a sentence connective (contrast chapter **4.16**). If it is a sentence connective, it is certainly not truth functional: John may have both true and false beliefs, so the truth value of "John believes that A" is not a function of the truth value of "A".

(2) Because

On at least some occurrences, "because" seems to function as a *binary sentence connective* (one which takes two sentences to make a sentence). For example

1) John shot Robert because Robert betrayed him.

If we take this appearance at face value, the straightforward response is to see "John shot Robert" (abbreviate to A) and "Robert betrayed John" (abbreviate to B) as the two components, in which case we are forced to regard "because", thus used, as non-truth functional. This view could be vindicated by attempting to construct a truth table in which we envisage various theoretically possible truth

Table 2.6

A	B	A because B
T	T	?
T	F	F
F	T	F
F	F	F

values for the components, and ask, with respect to each, whether (1) would then be true or false (table 2.6). A necessary condition for the truth of "because" sentences is the truth of both components. But this is not sufficient. If it were, the result of inserting "because" between arbitrary truths would always be a truth, and this is plainly not so.

Ex. 2.36 Give an example of a false "because" sentence whose components are both true.

Frege ([1892b], pp. 76–7) suggested that "because" sentences are truth functional, but are not a truth function of the components we have so far identified. Discussing the sentence

2) Because ice is less dense than water, it floats on water,

he suggested that there is a concealed component, in addition to "ice is less dense than water" and "ice floats on water", namely

3) Whatever is less dense than water floats on water.

The truth function is simply that of conjunction, applied to the *three* components.

A difficulty with this suggestion (and I think the difficulty is insuperable) is that there is no systematic way of eliciting the concealed component. What stands to (1) as (3) does to (2)? It would be absurd to suggest that (1) entails that whoever Robert betrays shoots him, or that whoever betrays John is shot by him, or that whoever betrays anyone is shot by the betrayed.

A more promising suggestion is that "because" is not really a sentence connective at all, but rather functions, somewhat analogously to "therefore", to mark an act of inference. We saw in chapter **1.5** that something like "*A* therefore *B*" cannot be evaluated as true or false. Rather, we have to ask whether the argument thereby presented is valid. "Therefore" is not a sentence connective, since what is produced by placing it between two sentences is not itself a *sentence*, something capable of having a truth value. Similarly, perhaps "because" forms not a sentence, but something more like an argument. (Very often, as in (1), one would be expected to evaluate the argument by inductive rather than deductive standards.)

Ex. 2.37 Explain how, contrary to what is said in the text, a Frege-style concealed component truth functional analysis might be applied to "*A* therefore *B*".

How could this suggestion be tested? One relevant point is that certain complexes containing "because" appear not to operate in the way one expects of genuine *sentences*. For example, if ¢ is a sentence connective, then "if *A*, then (*B* ¢ *C*)" is a conditional, and on every account of conditionals neither the truth of *B*, nor of *C*, nor of *B* ¢ *C* can be necessary for the truth of the conditional. But consider

4) If John comes to the party, then (Mary will leave because she has quarrelled with him).

It would seem that (**4**) is true only if Mary has quarrelled with John, and this would be hard to explain if "because" were a sentence connective.

Ex. 2.38 Could it be persuasively argued that, provided the scope-indicating brackets are held firmly in place, the truth of (**9.4**) does not require the truth of "Mary quarrelled with John"?

(3) But

I hold that "but" differs in meaning from both "and" and "&". This does not show that "but" does not express the conjunction truth function, for there can be differences of meaning which do not impinge on truth. When "but" functions as a sentence connective (which it does not always do) it expresses precisely the conjunction truth function – that expressed by "&". The only doubt is whether the truth of both *A* and *B* is enough for that of "*A* but *B*". This doubt is assuaged by the reflection that it is correct to infer from the falsehood of "*A* but *B*" that at least one of *A*, *B* is false.

Ex. 2.39 Give an example of an occurrence of "but" in which it is not a sentence connective.

The additional component in the meaning of "but", as opposed to "and", is reflected in the fact that one who asserts that *A* but *B* represents himself as supposing that in the context there is some contrast between the truth of *A* and that of *B*, something surprising, poignant, or worthy of special note or emphasis in the fact that both are true. Sometimes the element of surprise, or whatever, derives from *A* itself, as in the stock example "She was poor but honest". Sometimes it derives from something else, as in "The best smoked salmon comes from Scotland but unfortunately it's rather expensive" (cf. Jackson [1981], §3).

(4) When

Consider

5) When beggars die, there are no comets seen.

The following argument might be used to show that "when", on such an occurrence, is a non-truth functional sentence connective: "Beggars die" and "There are no comets seen" are both declarative, indicative sentences, capable of being used self-standingly, and each evaluable for truth and falsehood. So "when" is a sentence connective. But it is not truth functional, since the truth value of the two components leaves undetermined the truth value of the whole. Equally, it fails the substitution test. If **(5)** is true, it does not remain so when "Comets are seen" replaces "Beggars die", yet these components have the same truth value.

Ex. 2.40 Explain why the truth value of **(9.5)** is not a function of the truth values of "Beggars die" and "No comets are seen".

The argument is fallacious, for, as they occur in **(5)**, "Beggars die" and "There are no comets seen", are not genuine self-standing sentences. **(5)** means something like

6) Any time at which beggars die is not a time at which comets are seen.

Here there is no plausibility to the idea that there are two sentential components. Used self-standingly, "Beggars die" means something like

7) Every beggar dies sometime;

and "There are no comets seen" means something like

8) No comets are visible now/ever,

context determining the choice of "now" or "ever". It is plain that this is not what the expressions mean as they occur in **(5)**.

"When", as it occurs in **(5)**, is thus not an example of a non-truth functional sentence connective, since in that occurrence it is not an example of a sentence connective (cf. Frege [1892b], p. 72). This does not preclude there being other sentences in which it is a genuine sentence connective.

Ex. 2.41 (a) Give an example of a sentence in which "when" occurs as a genuine sentence connective. Is "when" truth functional, as it occurs in your example?

(b) For each of the following expressions, say whether it always, sometimes, or never occurs as a sentence connective. If always or sometimes, say whether it expresses a truth function on these occurrences, and, if so, which. Use examples to justify your answers.

(i) Iff
(ii) It is surprising that
(iii) It is true that
(iv) It is probable that

(v) Unfortunately
(vi) Fortunately
(vii) Science shows that
(viii) Although
(ix) Unless
(x) Hopefully
(xi) Even
(xii) Yet
(xiii) Not only . . . but also . . .
(xiv) Hastily (compare chapter **4.6**)
(xv) In the park
(xvi) Before (enthusiasts may wish to compare the discussion in Davidson [1970a], pp. 138–9)
(xvii) Possibly
(xviii) Despite the fact that.

10 From P-validity to validity

Logic is the study of reasoning, and of one vital feature that reasoning should possess, namely validity. There are two ways in which the language **P** may help us to understand validity in English: first, it may enable us to attain useful generalizations about validity in English, and offer ways of testing for validity in English in doubtful cases; secondly, the definition of **P**-validity may have virtues lacked by the original definition of validity in chapter **1**.

In §3 we discussed some ways of formalizing English arguments, but we left two questions with at best incomplete answers. One was: what can be inferred from the fact that an argument's logical form is **P**-valid? The other was: what can be inferred from the fact that an argument's logical form is not **P**-valid?

With some qualification, the answer to the first question is that one can infer the validity of the English. What we have to show is that if a **P**-argument ϕ is **P**-valid and is an *adequate* formalization of an English argument ψ, then ψ is valid.

The adequacy of the formalization ensures that, where ϕ' is the argument recovered from ϕ by applying the correspondences associated with the formalization, ϕ' says the same as ψ. The notion of "saying the same" that is required here is that the sentences of ϕ' should have the

same truth conditions as the corresponding sentences in ψ. Validity can be defined in terms of truth conditions, and if two arguments are related as ψ and ϕ' then, necessarily, both or neither is valid. So it will be enough if we can show that if ϕ is **P**-valid then ϕ' is valid.

A necessary condition for there being any adequate formalizations is that the **P**-connectives make the same contribution to truth conditions as the corresponding English expressions. We saw that this was doubtful in the case of "$\rightarrow$" and "if". Let us suspend this doubt for the moment. Then the following argument establishes what is needed:

1) (i) Suppose ϕ is **P**-valid.
(ii) Then every interpretation upon which the premises of ϕ are true is one upon which the conclusion is true.
(iii) Hence whatever may be the truth values of the components of ϕ', if the premises of ϕ' are true, so is the conclusion.
(iv) Hence, of logical necessity, all the conditions under which the premises of ϕ' are true are conditions under which its conclusion is true.
(v) Hence, the truth conditions of the premises of ϕ' are contained within those of the conclusion.

(v) is equivalent to the claim that ϕ' is valid, in the sense of $\models$.

One crucial step is from (ii) to (iii). This essentially requires that every English expression in ϕ' corresponding to a constant in ϕ expresses the same truth function as the **P**-constant. Those who dispute that "if" is truth functional will dispute this step.

Another crucial step is that from (iii) to (iv). The former makes no explicit mention of logical necessity, the latter does. How can the intrusion of this notion be justified?

Consider a very simple argument, like

2) Either John is happy or Mary is. But John is certainly not happy. Therefore Mary is happy.

No doubt there are all sorts of intrinsically different conditions under which the premises are true. For example, John's unhappiness might stem from frustrated ambition, misfortune in love, or whatever. (iii) appears not to allude to all these possibly different conditions, yet (iv) does: it speaks without qualification of *all* the conditions under which the premises are true.

Conditions, like anything else, can be classified either coarsely or finely. It is not that (iii), as applied to (**2**), fails to speak to some conditions under which the premises are true; rather, it classifies these conditions rather coarsely. In this example, they in effect fall into a single category: those in which "John is happy" is false (as required for the truth of the second premise), and in which "Mary is happy" is true (as required for the truth of the first premise, given what is required for the truth of the second). However the conditions in this category may vary, they all have the common property of verifying the conclusion. Hence all the conditions – all logically possible conditions – under which the premises are true are conditions under which the conclusion is true.

If "if" is not truth functional, (iii) does not follow from (ii). We will continue to assume that, even so, "if" entails "→", and that it is only the converse entailment that fails. In other words, an "if" sentence is stronger than the corresponding "→" sentence.

There are two cases, which have to be treated separately:

(a) The English argument contains "if" at most in the premises. In this case, the inference from **P**-validity to validity goes through. The reason is that the **P**-premises will be weaker than the English premises, and strengthening an argument's premises cannot invalidate it (cf. **1.6.5**).

(b) The English argument contains "if" in the conclusion. This divides into a number of subcases, depending on whether "if" is or is not the dominant connective. Let's just consider the case in which it is. Then the **P**-conclusion is weaker than the English conclusion. That the **P**-premises establish a weaker conclusion leaves open that they might not be strong enough to establish the stronger one. So there is no correct inference from **P**-validity to validity.

If a **P**-argument is invalid, there are three possibilities for the validity of the English argument:

3) It is valid in virtue of its **P**-logical form.

4) It is valid, but not valid in virtue of its **P**-logical form.

5) It is invalid.

I shall illustrate these possibilities in turn.

The first possibility is illustrated by (**3.8**): we saw that the formalization of an argument which is not only valid, but valid in virtue of its **P**-logical form, may be adequate yet invalid, through failing to be deep enough. In general, any argument with *n* premises can be formalized by an invalid **P**-argument which uses *n* sentence letters to correspond to the premises, and a further distinct letter to correspond to the conclusion.

The second possibility, (**4**), is illustrated by (**3.19**):

> All football supporters are interested in sport. Some football supporters are hooligans. So some hooligans are interested in sport.

The argument is valid, but its validity depends upon expressions which are not **P**-logical constants, "all" and "some".

The third possibility, (**5**), is illustrated by (**3.14**):

> If you are good at mathematics, you will find logic easy. But you are not good at mathematics. So you'll find logic hard.

P-logic offers us, naturally enough, no way of differentiating between cases (**4**) and (**5**), so the strongest conclusion we can reasonably draw from the **P**-invalidity of the deepest **P**-formalization we can find is that validity in virtue of **P**-logical form is a property which the formalized English argument lacks.

Ex. 2.42 Say what is wrong with the following reflections on some English argument, call it ϕ:

> The deepest **P**-formalization of ϕ I can find is **P**-invalid.
> Hence ϕ is not valid in virtue of its **P**-logical form.
> Hence, in virtue of its **P**-logical form, ϕ is invalid.

The definition of validity offered in chapter **1** used a notion which could do with elucidation: that of it being *logically impossible* for all the premises to be true, yet the conclusion false. The definition of **P**-validity dispenses with this notion, and employs instead the notion of interpretation.

Generalizing from the remarks adduced to support the move from (iii) to (iv) in (**1**), one way to classify the logical possibilities for the truth of a sentence dominated by a truth functional sentence connective is according to the truth values for the components which determine the sentence itself as true. *All* conditions for the truth of the sentence will be embraced by this classification, though there are finer classifications available which this one ignores. The notion of a **P**-interpretation thus gives a partial elucidation of the notion of logical impossibility (or logical necessity) used in the definition of validity in chapter **1**. Interpretations mirror logical possibilities. A sentence whose formalization is true on no interpretation represents a logical impossibility. A sentence whose formalization is true on all interpretations represents a logical necessity.

We can draw a stronger conclusion: if an English argument's formalization is valid, the argument itself is not merely valid, but formally so. This is because its validity depends only upon the meanings of the logical constants (the expressions corresponding to the **P**-connectives) and the pattern of occurrence of the sentences.

A traditional aim of logic has been to render mechanical the determination of the validity of arguments. **P** satisfies this aim. One can construct truth tables to check mechanically for the **P**-validity of **P**-arguments. There are also many computer programs which will do the same job.[8]

Finally, $\models_P$ has the general properties ascribed to $\models$ in chapter **1.6**.

Ex. 2.43 Using the definition of a P-interpretation, establish the following:

(i) If $[X_1, \ldots X_n \models_P Z]$ then $[X_1, \ldots X_n, Y \models_P Z]$, whatever Y may be. (Compare (**1.6.5**).)

(ii) If $[X_1, \ldots X_n \models_P Z]$ and $[Y_1, \ldots Y_k, Z \models_P W]$, then $[X_1, \ldots X_n, Y_1, \ldots Y_k \models_P W]$. (Compare (**1.6.6**).)

(iii) If Z is among the $X_1, \ldots X_n$, then $[X_1, \ldots X_n \models_P Z]$. (Compare (**1.6.7**).)

(iv) If $[(X_1, \ldots X_n) \models_P]$, then $[X_1, \ldots X_n \models_P Y]$, whatever Y may be. (Compare (**1.6.8**).)

(v) If $[\models_P X]$, then $[Y_1, \ldots Y_n \models_P X]$, whatever $Y_1, \ldots Y_n$ may be. (Compare (**1.6.9**).)

[8] **P** is decidable, that is, there exists a decision procedure for it (cf. e.g. Kirwan [1978], pp. 169 ff.).

Bibliographical notes

The history of logic goes back at least to Aristotle (384–322 BC). A classic account is Kneale and Kneale [1962]. Propositional logic was known to the Stoics (third century BC) but their work was lost (see Kneale and Kneale, ch. 3). George Boole developed a system equivalent to propositional logic in the mid-nineteenth century. The first wholly modern presentation is Frege [1879]. There are many good introductory texts, including Lemmon [1965], Hodges [1977] and Guttenplan [1992]. For a general discussion of the relationship between the **P**-connectives and English, see Strawson [1952], ch. 3, §II.

On conditionals, the contrast between (**4.34**) and (**4.35**) derives from Adams [1970], and his [1975] contains a number of arguments against the material conditional interpretation. I have drawn heavily on Jackson [1979]; see also his [1981] and [1987], and Lewis [1976], pp. 142–5 and postscript. A good recent overview is provided by Edgington [1995]. Further references on conditionals appear in the bibliographical notes to chapters **3** and **5**.

For Grice's work on implicature, see Grice [1961] (the first presentation of the notion, rather deeply embedded in a discussion of perception, but from which the famous (**6.4**) is drawn), [1975] (where he explicitly addresses an implicature defence of the truth functional interpretation of "if") and [1978]. See also McCawley [1981], ch. 8, esp. §3; Recanati [1991].

The butler and gardener example (**8.1**) comes from Stalnaker [1975], who offers a very sophisticated account of why we tend *wrongly* to suppose that it is valid. See also chapter **3.3**.

For the relation between the classical conception of validity (here represented by $\models$) and the material implication interpretation of conditionals, See Read [1988], ch. 2.

Frege's [1892b] defence of the truth functionality of natural language is contained in one of the most famous and important articles in philosophical logic.

3

Conditionals and probabilities

1 Degrees of confidence

We hold our beliefs with more or less confidence. We may believe with considerable confidence that we will not win the lottery; though if we bought a ticket we are presumably not absolutely certain that we will not win. We may believe we will not die within the week, but it would be foolish to be entirely confident of this, or perhaps even very confident. We may believe that some outsider will win the race, but evidently the bookmakers are much less sure.

We display different degrees of confidence towards conditionals. Suppose there are two urns, each containing a million marbles. Urn A's marbles are all white; urn B's are all white except for one black one. Suppose you know all this for certain, but do not know which of these urns is now before you. You reach into it and take a marble, and don't look to see what colour it is. You should be certain that

1) If this is urn A, I am not holding a black ball

and that

2) If I am holding a black ball, this is urn B.

But what about:

3) If this is urn B, I am not holding a black ball?

Perhaps one should not be certain of this, but it seems one ought to have a very high degree of confidence in it, for there is nearly a million

to one chance against "I am not holding a black ball" being false, given that "This is urn B" is true (cf. Appiah [1985], p. 172).

Facts (or apparent facts) about our degrees of confidence in conditionals can be used to provide a further argument against the view that English conditionals are properly formalized by "$\rightarrow$", that is, that they are material conditionals. Suppose I have no confidence in winning the lottery. It seems consistent to be, in addition, very unconfident in the following:

4) If I win the lottery I will not pay off the mortgage.

On the contrary, the first thing I would do with a lottery win is eliminate the mortgage. These degrees of confidence are hard to explain on the view that conditionals are material conditionals.

5) (I will win the lottery) $\rightarrow$ (I will not pay off the mortgage)

has an antecedent which I think is almost certainly false, and so I should think that **(5)** is almost certainly true (since the falsity of the antecedent of a material conditional is enough for the truth of the conditional itself). This conflicts with my apparently reasonable lack of confidence in **(4)** (cf. Edgington [1997], p. 106).

The systematic study of degrees of confidence began with an interest in betting, and a point similar to the one just made in connection with **(4)** and **(5)** can be made in this context. Betting lends itself to the assignment of numbers to measure degrees of confidence. How confident should you be that the die will land 6? There are just six possibilities, and nothing to choose between them, so the right answer appears to be 1/6. How confident should you be that

6) If the die lands with an even number, it will land 6?

There are three ways in which it can land with an even number, and just one of these three is 6, so the right answer seems to be 1/3. How confident should you be that

7) The die lands with an even number $\rightarrow$ it will land 6?

There are only two ways in which **(7)** can be false: by the die landing 2 or 4. On all four other possibilities **(7)** is true. So one should have a degree of confidence of 4/6 = 2/3. This is double the degree of confidence appropriate to (**6**), which would make it hard to explain how, as the material conditional thesis claims, (**6**) and (**7**) could have the same truth conditions (cf. Edgington [1997], p. 106).

Ex. 3.1 What degree of belief should you have in the following conditional:

If I draw a court card, it will be an ace?

This chapter explores degrees of confidence appropriate to conditionals, and the upshot for logic and for the truth conditions of conditionals. The initial seed of this approach was sown by Frank Ramsey in the 1920s, and it was developed in detail by Ernest Adams in the 1960s and 1970s and more recently by Dorothy Edgington.

2 Conditional probability

The crucial notion is that of *conditional probability*. We will work towards it from the notion of a *possibility space*. We can represent the fact that someone considers the chance of a die landing an even number to be the same as that of it landing an odd number, and that both these chances are 1/2, by the picture shown in figure 3.1. The main rectangle represents the whole of the "possibility space" relative to the proposition "the die lands even": it encloses all the possibilities. Its parts, marked "even" and "not even", occupy the same vertical distance (i.e. height). If we associate the whole height with 1, then the parts will each be associated with 0.5 (or 50 per cent), representing the fact that the two possibilities which have been distinguished are considered equally likely, and between them exhaust the space: the chance of the die landing an even number = the chance of the die landing an odd number = 1/2. Each of the smaller rectangles could be subdivided into three equal parts, to reflect the probability, for each of the six numbers, of the die landing that number.

Is it really a matter of chance which side the die lands? Isn't it really determined (though in ways we don't know) by how it is thrown? This may be so; but the notion of "chance" or "probability" to be discussed in this chapter is quite consistent with this view. It makes no claim to apply directly to how things are, but only to degrees of confidence or

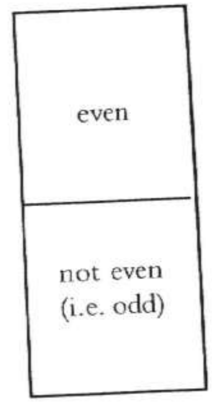

Figure 3.1

belief in how things are. If in the present discussion I say that the chance or probability of the die landing an even number is 0.5, I am to be understood as saying that some unspecified but rational person considers the likelihood of this being so to be 0.5 or 50:50. We could express this by saying that, for this person in this state of information, "Pr(the die lands an even number) = 0.5". One sign that a person has this degree of confidence is his betting behaviour: for example, he can be expected to regard a bet on heads in which he pays a dollar if he loses and otherwise wins two dollars as attractive, and one in which these figures are reversed as unattractive.

Subjective degrees of confidence, represented by expressions of the form $\Pr(A)$, can be evaluated as correct or incorrect, rational or irrational. For example, a person who thinks the die is not biased is irrational if, in addition, his beliefs can be described by "Pr(the die lands an even number) = 0.75".

The method of representation incorporates some of the minimum principles of rationality, for example: "$\Pr(A) = 1 - \Pr(\text{not-}A)$". If you believe it almost certain that John will come (call this A) then you should regard it as extremely unlikely that he will not come. Figure 3.2 makes plain that the values of A and not-A must sum to 1 (the total height of the possibility space). Certainty in A would be represented by making it fill the whole space, leaving no room for $\neg A$.

We can juxtapose possibility spaces for distinct propositions in a way which reveals the relations between them. For example, suppose A and B

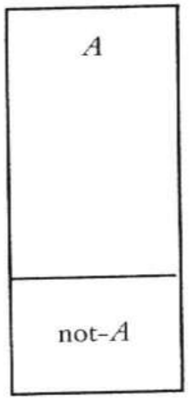

Figure 3.2

are contradictory, and that this is reflected in a subject's degrees of belief. The situation can be represented by figure 3.3. Reading horizontally, the idea is that the possibilities in which A holds are just those in which not-B holds, and those in which not-A holds are just those in which B holds.

We are not only more or less confident about simple propositions, like whether John will come back today, but also about complex ones, like whether John will come back and it will rain, and whether John will come back if it rains. We can use the diagrams to show how the degrees of belief in the more complex are related to degrees of belief in the less complex. Suppose I believe that A is quite likely and B significantly less so. One way for this to be so is represented by the diagram in figure 3.4. In order to represent some relationships between A and B, the probability associated with not-B has had to be displayed non-contiguously: it is the sum of the heights labelled not-B $(x + z)$. The significance of the horizontal juxtaposition in the representation of degrees of belief can be explained by example:

> The height y represents the probability associated with "A and B".
> The height $x + y + w$ represents the probability associated with A and also with "A or B".
> The probability associated with "A or not-B" is 1.

The probability of P given Q is the proportion of the Q height with horizontally matched P height. In figure 3.4:

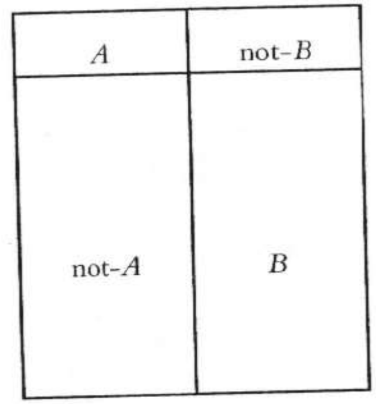

Figure 3.3

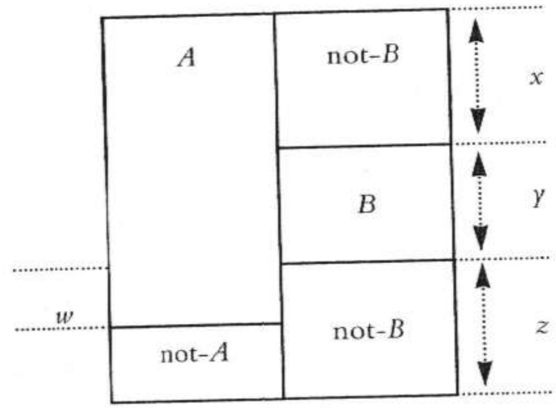

Figure 3.4

The totality of the B height has horizontally matched A height; so the probability of A *given* B is 1; equivalently, the probability of A *conditional upon* B is 1. In condensed form:

$\Pr(A \mid B) = 1$.

Alternative formulation: in all the possibilities in which B is true, so is A. So on the supposition that B, A is certain (though, as the height of not-A shows, in the absence of that supposition it is not certain).

Consider the not-B height $(x + z)$. The majority of it has horizontally matched A height $(x + w)$. (The part of its height not having horizontally matched A height is only $(z - w)$.) So the probability of A conditional upon not-B is greater than 1/2:

$Pr(A \mid \text{not-}B) > 1/2$.

Alternative formulation: On the supposition that not-B, our confidence in A would be measured as the proportion of the cases in question in which A is true to all the relevant cases. This ratio is $(w + x)/(x + z)$, and is more than a half.

In calculating the size of the conditional probability of P given Q, one pays no heed at all to not-Q: a conditional probability is the probability of a proposition on a supposition, so only the supposition, and not its negation, is relevant.

The explanation takes for granted that one cannot believe to be true a proposition which one is certain is not true. There is no such thing as the conditional probability of a proposition given some proposition assigned probability zero. This is encoded in the diagrams: we are to calculate $Pr(A \mid B)$ by taking the possibilities in which B, and comparing them with those in which A also holds. The procedure is inapplicable if there are no possibilities in which B; that is, if B is represented as having zero height.

Conditional probability is closely related to the probability of conjunctions, in a way that has sometimes been used to offer a definition of it:

1) if $Pr(A) > 0$, $Pr(B \mid A) = Pr(B \,\&\, A) \div Pr(A)$.

One cannot challenge the equation; but one can question whether it provides a useful definition of conditional probability. For how are we to understand the probability of a conjunction? A standard answer is given by the equation:

2) $Pr(B \,\&\, A) = Pr(B \mid A) \times Pr(A)$, if $Pr(A) > 0$.

We could either take conditional probability as a primitive notion and use it to define the probability of a conjunction, or take the probability of a conjunction as primitive and use it to define conditional probability. The former seems the preferable option. To use **(1)** as a definition of $Pr(B \mid A)$ would imply that there can be no such thing as $Pr(B \mid A)$ if there is no such thing as $Pr(A)$. However, we may assign a conditional probability even when we do not assign any probability to the components. We may think that

3) Pr(The streets in Moscow are wet)|(it is raining in Moscow)

is close to 1, while being unwilling to assign any definite probabilities to "The streets in Moscow are wet" or to "it is raining in Moscow" or to the conjunction of these. What one is confident of is that, supposing that it is raining in Moscow, the chance of the streets being wet is high, that is, high relative to the supposition, whatever the probability of the supposition may be. This means that there is such a thing as $Pr(B \mid A)$ even when there is no such thing as $Pr(A)$; so **(1)** would not be a useful definition in the sense of one which we would always or typically use in estimating conditional probabilities.

Conditional probability plays a more fundamental role in our reasoning than the probability of a conjunction, to which **(1)** accords a defining role. It is very common to suppose that such-and-such is so, in order to assess, for example, how to act: supposing I do such-and-such, how will things turn out? Or supposing such-and-such happened, what should I do? The notion of how likely one thing is, on the supposition of another, belongs with very familiar notions. By contrast, it is not clear that we have so natural an everyday use for reaching an assignment of a probability to a conjunction.

3 Probabilistic logic

The conditional probability of B given A seems to express the right probability to assign to the conditional "if A then B". In considering the degree of belief I should assign to "if A then B", it is natural to reason: suppose that A; how likely it is that B? We can state this as the following hypothesis:

1) $Pr(\text{if } A \text{ then } B) = Pr(B \mid A)$.

Ex. 3.2 Show that the conditional probability of B given A may differ from the degree of belief in "if A then B" based on the rule:

Suppose *you believe that* A; how certain are you that B?

Consider (**1.3**):

If this is urn B, I am not holding a black ball.

Given that you know that just one of the million balls in urn B is black, it seems right to believe (**1.3**) with great confidence. The related conditional probability is likewise very high: given that this is urn B, the chances of holding a white ball are nearly a million to one. One can arrive at this conditional probability without having assigned any absolute probability to "this is urn B".

The assignments of probability claimed in §**1** to pose a difficulty for the view that English conditionals are material conditionals coincide with what one would expect if (**1**) were true. It is consistent to assign a low absolute probability to "I will win the lottery" and also a low probability to "I will not pay off the mortgage" on the supposition that I do win the lottery. I suppose, unlikely as it is, that I win the lottery, and I consider what I do; paying off the mortgage presents itself as a highly likely first step, not paying it off seems most unlikely; so the conditional probability of not paying off the mortgage, given that I win the lottery, is low. If (**1**) is right, this means that I assign low probability to (**1.4**):

If I win the lottery I will pay not off the mortgage.

What matters to the conditional probability of *B* upon *A* is not the size of the probability one awards to *A*, but how much of the probability space awarded to *A* is also occupied by *B*. This emerges clearly with examples like (**1.6**):

If the die lands with an even number, it will land 6.

The reasoning given in §**1** for the conclusion that one should believe this to degree 1/3 in effect invited one to suppose the die lands with an even number, and then to count the ratio of those cases in which that even number is 6 to the totality of cases (the totality, that is, consistent with the supposition, that is, all the even numbers). This is diagrammatically presented in figure 3.5. Because a third of the possibilities, given that the die lands even, are that it lands 6, one should, according to (**1**) and common sense, assign 1/3 as the probability of (**1.6**).

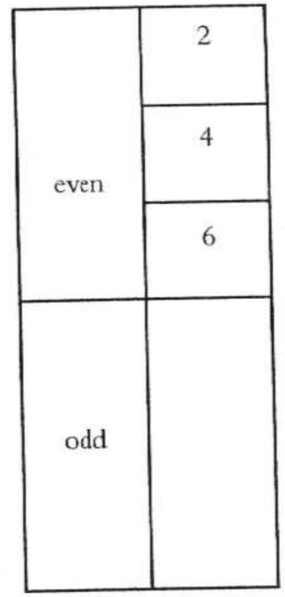

Figure 3.5

Typically, our premises when we reason are less than certain, but we want to reason in such a way as not to increase our uncertainty. If we believe a proposition to degree 0.75, that is, if we assign it this probability, then our uncertainty in it is 0.25. In other words, the uncertainty of a proposition is equal to the probability of its negation. Not increasing uncertainty is a constraint upon reasoning, and indicates how we can define a notion of *probabilistic validity*:

2) An argument $A_1, \ldots A_n; C$ is probabilistically valid (for short, $A_1, \ldots A_n \mid\!\sim C$) iff necessarily, the uncertainty of the conclusion does not exceed the sum of the uncertainties of the premises.

The "necessarily" requires that however a rational person assigns probabilities to the components of the arguments, there can be no increase in uncertainty. Valid reasoning, according to (**2**), is all and only reasoning which *cannot* introduce new uncertainty. It can be shown (cf. Adams [1975]) that if we consider a propositional language with no conditionals, the classically valid arguments (those for which $\models_P$ holds) are just those which are probabilistically valid (those for which $\mid\sim$ holds). If we now extend the language to include English conditionals, and we treat their probability as the corresponding conditional probability as (**1**) commends, and treat (**2**) as the right account of validity, we get a logic which departs from the classical: let us call it *probabilistic logic*.

Ex. 3.3 Use (**3.2**) to show the correctness of the following claim: $A, \neg A \mid\sim B$ (whatever B may be).

In the remainder of this section, we will consider some of these departures. In particular, we will see whether some of the argument patterns which presented problems for the view that English conditionals are material conditionals are validated in probabilistic logic.

(**2.5.7**) was:

$X \models_P [\neg X \rightarrow Y]$, whatever Y may be.

Correspondingly, we enquire whether:

3) $A \mid\sim$ [if not-A then B], whatever B may be.

(**2.5.10**) was:

$Y \models_P [X \rightarrow Y]$, whatever X may be.

Correspondingly, we enquire whether:

4) $B \mid\sim$ [if A then B], whatever A may be.

To both enquiries the answer is negative: both (**3**) and (**4**) are false.

Figure 3.6 shows that (**3**) is false. The uncertainty of A, that is, the probability of not-A, is quite small, but the probability of B given not-A is zero, so its uncertainty is 1, so the condition (**2**) for probabilistic validity has not been met. If (**1**) is correct, were we to reason from A to "if not-A then B", we could end up with a conclusion that was less certain than the premise. For example, I am pretty certain that I will catch the train (A) and so will be at work on time (B), but I have zero confidence that I will be at work on time if I miss it.

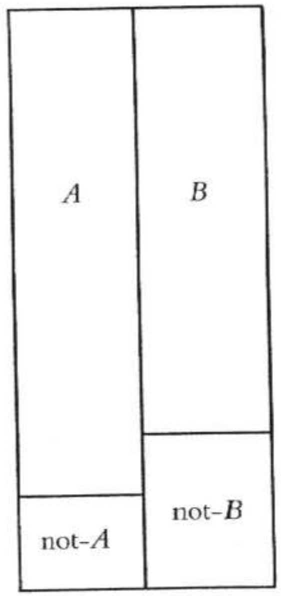

Figure 3.6

Figure 3.7 shows that (**4**) is false. The uncertainty of B is quite small, but the uncertainty of if A then B is 1.

Ex. 3.4 Give an example of an argument to which it is reasonable to assign degrees of belief in the way described by figure 3.7.

We have already seen (in §**1** above) that the view that English conditionals are material conditionals is hard put to explain our apparently rational assignments of probabilities to conditionals. We can also show that not all the principles used to argue for the material conditional treatment are probabilistically valid. This is what one would expect if the probabilistic view of conditionals provides a genuine and coherent alternative to the truth functional one.

The first argument (chapter **2.8**) for the truth functional interpretation essentially depended upon the following principle:

5) A or $B \models$ if not-A then B.

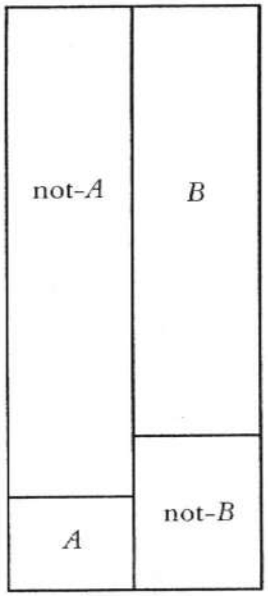

Figure 3.7

The probability of a disjunct is represented by the net sum of the heights of the disjuncts (that is, overlapping heights are not counted twice). In Figure 3.6, the uncertainty of *A* or *B* is small, but the uncertainty of if not-*A* then *B* is 1; so the argument in **(5)** is not probabilistically valid. However, arguments on the pattern of **(5)** appear valid, so perhaps we have here a reason to suspect that principles **(1)** and **(2)**, which respectively offer a conditional probability account of the probability of conditionals and a probabilistic account of validity, do not provide a correct picture of our reasoning with conditionals.

The earlier argument of the form of **(5)** was given in **(2.8.1)**, viz.:

> Either the butler or the gardener did it. Therefore if the gardener didn't do it, the butler did.

This seems to be valid, and seems to be so in virtue of its form; so it seems that such reasoning could not increase uncertainty.

The probability theorist disputes this. Suppose you believe that the gardener is overwhelmingly the most likely suspect, and you are sure

that if the butler was involved it was simply as the gardener's accomplice (he would not have acted alone). There is nothing irrational about these assignments of probability. Yet they appear to involve thinking that the premise of **(2.8.1)** is highly likely, and the conclusion certainly false. If this is correct, there is, after all, a reason for thinking that **(2.8.1)** is not valid, for if an argument is valid it is not rational to assign high probability to its premise and low probability to its conclusion.

Ex. 3.5 If you think the premise of **(2.8.1)** can be highly likely, yet its conclusion certainly false spell out why. Would your reasons be accepted by one who analysed conditionals truth functionally?

Other examples suggest that there are invalid instances of the form in **(5)**, for example:

6) Either he will throw an even number or he will throw a 6. Therefore, if he does not throw an even number, he will throw a 6.

This strikes many as invalid, which would be enough to refute **(5)**. It is certainly probabilistically invalid. By contrast, a truth functional theorist will insist that if the premise is true because an even number is thrown, the conclusion is automatically true; otherwise the premise is false, so there can be no counterexample. The appearance of invalidity of **(6)**, according to the truth functional theorist, adds nothing to the already admitted oddness of conditionals whose antecedents are known to be false.

Ex. 3.6 Demonstrate the probabilistic invalidity of **(3.6)** by specifying the uncertainty of premise and conclusion.

The other argument given in chapter **2.8** for the truth functional interpretation of conditionals was based on the principle of conditional proof, **(2.8.8)**, viz.:

If $[A_1, \ldots A_n, B \models C]$ then $[A_1, \ldots A_n \models \text{if } B \text{ then } C]$.

We can show that we can have

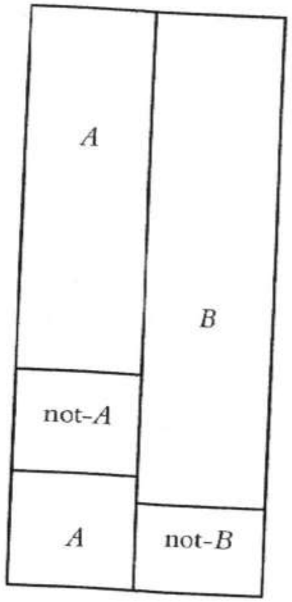

Figure 3.8

7) $A, B \mid\!\sim C$

without having

8) $A \mid\!\sim$ if B then C.

One way to do so is to consider the instance in which "$A \vee C$" replaces "A" and "not-A" replaces "B". Since this involves no conditionals, and since $\mid\!\sim$ and $\models_P$ coincide for such cases, we can be confident that we have a truth of the form of (7). But we have already seen that the corresponding instance of (**8**) is false (in the discussion of (**2.8.1**)).

Other classical principles of validity which hold for "→" fail on the probabilistic approach, for example, *contraposition* (cf. **2.5.19**) and transitivity (cf. **2.5.22**). Contraposition can be represented as:

9) If A then B; if not-B then not-A.

Figure 3.8 demonstrates the probabilistic invalidity of the argument in (**9**).

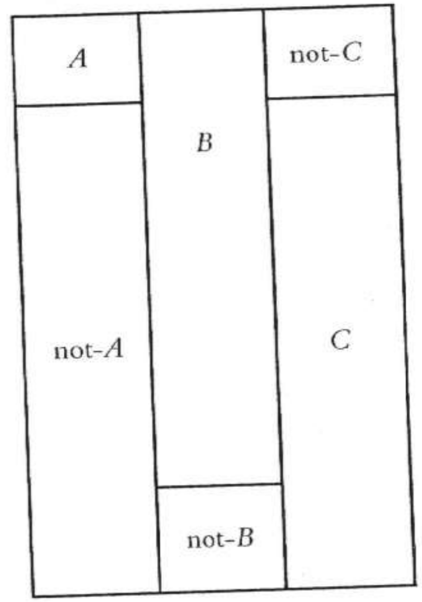

Figure 3.9

Ex. 3.7 Stalnaker (1984, pp. 124–5) gives a general argument for thinking that contraposition is unacceptable. The crucial assumption is:

If ($B \models C$) then (if A then $B \models$ if A then C).

Can you complete the argument? What would be the best response that could be made by one who thought conditionals were truth functional?

Transitivity can be represented as:

10) If A then B, if B then C; if A then C.

Figure 3.9 demonstrates the probabilistic invalidity of (**10**). In chapter **2**, the following example (**2.5.23**) was offered to suggest that our ordinary reasoning with conditionals does not always conform to transitivity:

If Smith dies before the election, Jones will win.
If Jones wins, Smith will retire from public life after the election.
So, if Smith dies before the election, he will retire from public life after the election.

Some arguments which instantiate transitivity seem valid, for example

11) If the battery's dead, then the car won't start.
If the car won't start, then I won't be able to get to work.
So, if the battery's dead, then I won't be able to get to work.

The fact that this argument is not probabilistically valid needs to be explained by the probability interpretation. If one could show how uncertainty could be increased by reasoning in this way, one would show that it cannot really be valid, however it may appear. One might also be able to explain the semblance of validity, perhaps along the following lines: this argument form *is* probabilistically valid:

12) If *A* then *B*, if *A* and *B* then *C*; if *A* then *C*.

Ex. 3.8 Can you suggest a scenario in which it is reasonable to accord high probability to the premises of **(3.11)** but low probability to its conclusion?

Perhaps arguments like **(11)** strike us as valid because we are quite happy to add the antecedent of the first conditional to the antecedent of the second; in other words, we assimilate the form of **(11)** to **(12)**. Were we to do this in the case of **(2.5.23)**, we would have the following:

13) If Smith dies before the election, Jones will win.
If Jones wins and Smith dies before the election, Smith will retire from public life after the election.
So, if Smith dies before the election, he will retire from public life after the election.

Many people find the second premise unacceptable, and so would have no use for the argument.

A seemingly convincing argument concludes that we should not expect transitivity. The relevant assumptions are:

14) For some *A*, *B*, "if *A* then *B*" is true even though *A* does not entail *B* (so it is logically possible that *A* is true yet *B* false).

15) Any conditional with a possibly true antecedent and a consequent incompatible with its antecedent is false.

Ex. 3.9 Show with an example why **(3.15)** would be unacceptable to the theorist who views English conditionals as material conditionals. Compare with **(2.3.28)**, which might be abbreviated to: "If common sense is correct, then common sense is incorrect. Therefore common sense is incorrect."

Consider the following instance of **(10)**:

16) If *A* and *C* then *A*, if *A* then *B*; if *A* and *C* then *B*

(putting "*A* and *C*" for "*A*", "*A*" for "*B*" and "*B*" for "*C*"). If dropping a necessarily true premise cannot make a valid argument invalid, and assuming that "if *A* and *C* then *A*" is necessarily true, then if **(16)** is valid so is

17) If *A* then *B*; if *A* and *C* then *B*.

Take an instance in which we put "not-*B*" for "*C*". "If *A* then *B*" is to be a true conditional which does not entail its consequent (**(14)** assures us that there is one), so "if *A* then not-*B*" is possibly true. So by **(15)** the conclusion of **(17)** is false, though the premise is true. So we must reject the supposition that **(16)** is valid, and so also the supposition that **(10)** is valid.

We have already noticed an apparent counterexample to the validity of **(17)** in **(2.5.28)**, viz.:

If I put sugar in this cup of tea it will taste fine.
So, if I put sugar and also diesel oil in this cup of tea, it will taste fine.

The probability theory, as characterized in this section, gives an account of the degree of belief one should have in a conditional (equal to the conditional probability of the consequent upon the antecedent) and tells us how to reason with conditionals (so as not to reduce uncertainty, characterized, for conditionals, in terms of conditional probabilities). It does not say anything about the truth conditions of conditionals, which is a topic we turn to later (in §4). Moreover, I have not presented this theory as saying anything about the conditions under which conditionals should be asserted. Assertibility will be discussed briefly now, under the heading of assorted difficulties with the probability theory.

Ex. 3.10 If any of the following are incorrect, draw a diagram to establish this, in each case giving an example of an argument which is intuitively invalid:
i) It is not the case that if *A* then *B* |~ *A*.
ii) It is not the case that if *A* then *B* |~ it is not the case that *B*.
iii) If *A* then not-*B*, *B* |~ not-*A*.

(a) Even if the theory gives the right account of degrees of belief, it is wrong to link this with degrees of assertibility. For example, it may be urged that even if I allow only a slight chance of (**1.3**), viz.:

If this is urn B, I am not holding a black ball

being false, I still ought not to assert it.

Although probability theories are sometimes cast in terms of assertibility, this is not essential (and that casting has been avoided here: for a discussion, see Edgington [1997] p. 101–5). The theorist does best to admit that the relation between the degree of one's belief and the appropriateness of assertion is a complex issue. There is certainly no immediate inference from high degree of credence to appropriateness of assertion, as one can see at once by considering the demands of etiquette.

Ex. 3.11 Give an example of a conditional belief with a high degree of probability and a situation in which asserting it would be highly misleading or otherwise inappropriate.

(b) The conditional probability of *B* given *A* is undefined when Pr(*A*) = 0. Yet surely we believe some conditionals which have antecedents we regard as certainly false. I will consider some putative examples.

18) If the longhorns lose, then I'm a monkey's uncle.

One who says this represents himself to be as sure that the longhorns will win as he is that he is not uncle to a monkey, that is, as assigning Pr(the longhorns lose) = 0. Yet the conditional appears entirely in order.

Options for the probability theorist include (i) saying that this is a special idiom from which no general lesson can be inferred; (ii) saying that one who asserts or believes (**18**) must allow the possibility of the longhorns' defeat: (**18**) expresses, not complete certainty in victory, but merely very high probability; (iii) having a special rule for the case in which the antecedent has zero probability, for example, that every such conditional is to count as certain.

The special rule proposal is antithetical to the spirit of the probability theory, and would bring its account closer to the material conditional account. More than one rule is possible, but if we take (**18**) at face value, an appropriate one is: a conditional to whose antecedent I assign zero probability is one I should regard as certain. Suppose my confidence in *A* decreases over time, finally reducing to zero. As far as "$A \rightarrow B$" goes, this corresponds to gradually increasing confidence in its truth. But if no part of the *A* height at the beginning of this process is horizontally opposite to the *B* height, then on the probability theory, if it adopts the envisaged special rule, I start with certainty of the falsity of "if *A* then *B*" and suddenly, as I move from being almost to completely certain of the falsehood of *A*, the conditional becomes certainly true. This is not plausible. An analogous difficulty would affect the special rule that if Pr(*A*) = 0 then Pr(if *A* then *B*) = 0; and either rule will have difficulty explaining away the intuition that some conditionals with certainly false antecedents are true and others false.

Ex. 3.12 Can you give an example of a true conditional with a certainly false antecedent, and an example of a false conditional with a certainly false antecedent?

A more everyday example than (**18**) makes the "special idiom" option less attractive:

19) If someone broke in, they must have repaired the damage before they left.

This might be said as a way of reassuring a worried homeowner: since we all know that burglars do not repair the damage they cause, the audience is intended to achieve complete confidence that no one broke in. The audience is supposed to believe the conditional and so, like the speaker, assign Pr(someone broke in) = 0.

There are various similar kinds of case in which we seem to have a use for a conditional with an antecedent we are absolutely certain is false. I know I am innocent of the murder, which occurred in London. In my interrogation I affirm:

20) If I committed the murder, then I was in London on Tuesday evening.

If I can get the police to agree, then their acceptance of my alibi showing I was in Oxford on Tuesday evening commits them to accepting my innocence. In cases like this, I don't expect my audience already to be certain of the falsehood of the antecedent. The probability theorist might see both (**19**) and (**20**) as conniving or deferential uses: I adopt for the sake of argument the degrees of belief of my of my audience, which, at that point, is non-zero for the antecedent. But it would seem that I can steadily believe (**20**), and perhaps also (**19**), regardless of the rhetorical or conversational context which makes their utterance natural.

In mathematical reasoning, it is common to introduce assumptions for *reductio ad absurdum*, and these may figure as antecedents of conditionals. For example, a famous proof that there is no greatest prime may be represented as starting from

21) If there is a greatest prime, then it is odd.

There seems no question but that many users of (**21**) assign zero credence to the antecedent. While we are no doubt less than certain of many necessary truths (of logic or mathematics, for example) there seem to be many of which we are completely certain, and it seems that this confidence does not prevent us using their negations as antecedents of intelligible conditionals.

The best response for the probability theorist, in my view, is to pursue the notion of conniving or deferential use, an idea mentioned (rather briefly) by Edgington: a conditional is only used when the speaker or thinker takes the antecedent to be not certainly false "at least for the sake of argument, at least temporarily, at least to co-operate with her audience" ([1995], p. 265). All that is required for there *to be* a conditional probability is that there *be* a rational non-zero assignment of a positive degree to the antecedent. We are good at putting ourselves in the epistemic states of others or in imaginary epistemic states. We are willing suspenders of disbelief; we can suppose to be so what we know not to be so. There would be nothing strange in the idea that a full description of these activities saw in them pretended or imagined shifts in our degrees of belief. It is not that when I suppose that A I suppose that I believe that A, but that seriously entering into the supposition may also involve rearranging actual degrees of credence.

(c) The final issue for probabilistic logic to be mentioned here is that it ought to provide an explanation of why the probability of a conditional should be equal to the probability of its consequent, given its antecedent. This is the main topic of §**4**.

4 Lewis's proof

One possible explanation for why it is that (**3.1**), viz.

$$\Pr(\text{if } A \text{ then } B) = \Pr(B \mid A),$$

is that a conditional "if A then B" expresses some proposition, say X, meeting the condition that in every reasonable distribution of degrees of belief,

1) $\Pr(X) = \Pr(B \mid A)$.

Having initially identified X in terms of appropriate degrees of belief, we might be able go on to determine its truth conditions. David Lewis [1976, 1986] has proved the remarkable fact that, on certain assumptions, there is no such proposition X. The proof requires more precision and some additional material. Many readers will prefer to skip it, and take up the thread again at §**4(2)** on p. 47.

Ex. 3.13 On present assumptions, one way to show that conditionals are not material conditionals is to show that X cannot be $A \rightarrow B$. Show that:

$\Pr(A \rightarrow B) \neq \Pr(B|A)$.

(1) Details of the proof

A *probability function*, Pr, is an assignment of numbers to all the sentences of the language in accordance with the following rules:

2) $1 \geq \Pr(A) \geq 0$.
Pr assigns every sentence a number between 1 and 0.

3) If $\models (A \leftrightarrow B)$ then $\Pr(A) = \Pr(B)$.
Pr assigns logical equivalents the same value.

4) If $\models (A \rightarrow \neg B)$ then $\Pr(A \vee B) = \Pr(A) + \Pr(B)$.
Pr assigns to a disjunction of incompatible sentences the sum of the numbers it assigns to each disjunct.

5) $\Pr(A \;\&\; B) = \Pr(A \mid B) \times \Pr(B)$, if $\Pr(B) > 0$.
If $\Pr(B) > 0$, Pr assigns a conjunction of A and B the product of the probability of A given B with $\Pr(B)$.

6) $\Pr(A) = 1 - (\Pr(\neg A))$.
The probability of a sentence and its negation sum to 1.

7) If $\models A$ then $\Pr(A) = 1$.
Pr assigns 1 to logically necessary truths.

There are no constraints on how a probability function assigns numbers to atomic sentences: there is a probability function $\Pr_1$ such that $\Pr_1$(the earth is flat) = 0 and also a probability function $\Pr_2$ such that $\Pr_2$(the earth is flat) = 1, and indefinitely many other functions assigning intermediate values.

Ex. 3.14 Why does **(4.7)** not constrain how a probability function assigns numbers to atomic sentences?

The claim **(3.1)** is intended to generalize over probability functions: for any such function, or at least any reasonable one (that is, any function which could represent the way a rational thinker could assign probabilities to his beliefs), Pr(if A then B) = $\Pr(B \mid A)$. In his original proof, Lewis assumed that reasonable probability functions would be *closed under conditionalization* (subsequent work, by Lewis and others, has shown how to weaken this assumption). The assumption is that if $\Pr_i$

is reasonable and $\Pr_i(C) > 0$, then there is a reasonable function $\Pr_j$ such that, for all A, $\Pr_j(A) = \Pr_i(A \mid C)$.

For *reductio*, suppose that "if" is a non-truth functional sentence connective whose meaning ensures that **(3.1)** holds for any reasonable probability function. Reasonable probability functions satisfy the following condition:

8) Pr((if A then B) | C) = $\Pr(B \mid (A \text{ and } C))$, if $\Pr(A \text{ and } C) > 0$.

Here we find "if A then B" used as a component in a conditional probability. This is justified by the assumption for *reductio*: "if A then B" is a proposition like another, and so can be an instance of a sentence letter.

To establish **(8)**, take an arbitrary probability function, say $\Pr_i$. Since the functions are closed under conditionalization, there is also a function, $\Pr_k$, such that, for some C for which $\Pr_i(C) > 0$, for all A, $\Pr_k(A) = \Pr_i(A \mid C)$. The remainder of the argument for **(8)** can be set out thus:

9) $\Pr_k(B \text{ and } A) = \Pr_i((B \text{ and } A) \mid C)$ (definition of $\Pr_k$)
$= \Pr_i(A \mid C) \times \Pr_i(B \mid (A \text{ and } C))$ (by **(5)**)
$= \Pr_k(A) \times \Pr_i(B \mid (A \text{ and } C))$ (definition of $\Pr_k$)
$\Pr_k(B \text{ and } A) = \Pr_k(A) \times \Pr_k(B \mid A)$ (by **(5)**)
$= \Pr_k(A) \times \Pr_k(\text{if } A \text{ then } B)$ (by **(3.1)**)

Ex. 3.15 Show how to use **(4.5)** to establish:
$\Pr_i((B \text{ and } A) \mid C) = \Pr_i(B \mid (A \text{ and } C))$.
Hint: you have to use **(4.5)** more than once; and a strict setting out of the proof would also invoke **(4.3)**.

Hence

$$\Pr_k(A) \times \Pr_k(\text{if } A \text{ then } B) = \Pr_k(A) \times \Pr_i(B \mid (A \text{ and } C))$$

and so

$$\Pr_k(\text{if } A \text{ then } B) = \Pr_i(B \mid (A \text{ and } C))$$

which, given the definition of $\Pr_k$, and the fact that $\Pr_i$ was arbitrary, amounts to **(8)**.

Now take arbitrary sentences *A*, *B* and some probability function such that Pr(*A* and *B*) > 0 and Pr(*A* and not-*B*) > 0. From **(2)**–**(7)** it follows that

10) Pr(if *A* then *B*) = Pr((if *A* then *B*) and *B*) + Pr((if *A* then *B*) and not-*B*).

Hence by **(5)**

11) Pr(if *A* then *B*) = (Pr((if *A* then *B*) | *B*) × Pr(*B*)) + (Pr((if *A* then *B*) | not-*B*) × Pr(not-*B*)).

Applying **(8)** to this gives:

12) Pr(if *A* then *B*) = (Pr(*B*|(*A* and *B*)) × Pr(*B*)) + (Pr(*B*|(*A* and not-*B*)) × Pr(not-*B*)).

This simplifies to

13) Pr(if *A* then *B*) = (1 × Pr(*B*)) + (0 × Pr(not-*B*)) = Pr(*B*).

(13) is unacceptable: it says that for an arbitrary probability function, and an arbitrary conditional, the probability of the conditional is the same as the probability of its consequent. There are plenty of counterexamples. A premise which leads to this absurdity must be rejected, so Lewis claims that we must reject the assumption that "if" is a non-truth functional sentence connective whose meaning ensures that **(3.1)** holds for any reasonable probability function.

Ex. 3.16 Give a counterexample to **(4.13)**.

If we are willing to hold on to **(3.1)** (as Lewis is not) it would have been enough to assume something more general: that conditionals are propositions, as **(1)** affirms. By **(3.1)**, if they are propositions they are ones whose probability is always the conditional probability of the consequent, given the antecedent. The assumption that they are propositions is used in taking **(8)** to be intelligible: if "if *A* then *B*" is not a proposition, we cannot intelligibly ask after its probability given *C*. Lewis's result has therefore been taken by some (though not by Lewis himself) to show that there are no conditional propositions. This view will be considered in **§5**. In order to get a clearer fix on what it is, let us consider an alternative argument for this conclusion, given by Edgington.

(2) Alternative arguments

If there is a proposition *X* which satisfies **(1)**, then either it is a material conditional or else it affirms a stronger-than-material link between its components. Edgington suggests we can eliminate both possibilities. For:

(i) Minimal certainty that *A* or *B* is enough for certainty that if not-*A* then *B*;
(ii) It may be rational to disbelieve both *A* and also if *A* then *B*. (cf. Edgington [1995], p. 279)

The idea is that the truth functional account meets (i) but not (ii), and any stronger-than-material link meets (ii) but not (i).

To be minimally certain that *A* or *B* is to assign 1 to "*A* or *B*" but not to *A* and not to *B*. (We saw earlier that one who assigns high credence to "*A* or *B*" just because, or largely because, they assign high credence to *A*, may also assign low credence to "if not-*A*, then *B*". Such persons are certain but not "minimally" certain that *A* or *B*.) We have already seen both that the truth functional account meets (i) and that it fails to meet (ii).

A stronger-than-material account of conditionals might require that the conditional should represent a causal or inferential connection between antecedent and consequent. The details do not matter. The point is that though such an account may meet (ii), it cannot meet (i), since minimal certainty that *A* or *B* is not enough for certainty of a special link between *A* and *B*.

The divergence which upsets **(1)** arises because a degree of belief in *B* given *A* is insulated from not-*A* facts: the only thing relevant is the proportion of ways in which *A* is true which are also ways for *B* to be true. By contrast, a proposition *X*, with *A* and *B* as components, and purportedly satisfying **(1)**, will not be thus insulated: some such propositions will be true when not-*A* and some will be false when not-*A*. (It cannot be that not-*A* is true whenever *X* is true, nor that not-*A* is false whenever

X is true: not only would either link be inconsistent with our use of conditionals, either link would prevent pr(X) always aligning with pr($B|A$).) Distinct reasonable probability functions may diverge in what they assign to not-A while coinciding on $B|A$ (to which assignments to not-A are irrelevant), and hence diverge in what they assign to X. This would be enough to show that X does not satisfy (**1**).

Ex. 3.17 Show that if X (a supposedly conditional proposition with A and B as components) was either true whenever not-A, or false whenever not-A, it would not correspond to "if A then B".

If there is no proposition satisfying (**1**) then conditionals are not propositions. How, then, can we use them in reasoning? The positive account might start by pointing out that there is a range of conditional acts: conditional bets, conditional questions, conditional commands:

14) I bet you that if Oswald doesn't kill Kennedy, someone else will.

15) If Oswald doesn't kill Kennedy, was it you?

16) If Oswald doesn't kill Kennedy, kill him yourself!

These acts do not have a content which is exhausted by some proposition. This would encourage one to add to the list conditional assertion and conditional belief. Conditional bets, questions and commands are in some sense "cancelled" if their antecedent does not obtain: the bet is off (no one wins and no one loses), it is as if no question had been asked, and nothing you can do counts as obeying or disobeying the order. In the case of belief and assertion, the analogue might be that if the antecedent does not obtain, you have no commitment to the consequent. This does not mean that to make a "conditional assertion" of B, given A, is to do something which, given B, amounts to an assertion of A. If I utter "if A then B" assertively, and A is true, it does not follow that I have *asserted* that B (though normally I am in that case somehow committed to B). Similarly, if my belief state can be described by "if A then B" and A is true it does not follow that I *believe* that B. The notion of "cancellation" associated with falsehood of the antecedent of a conditional bet, question or order could be applied in more than one way to conditional assertion or belief.

On this theory there can be no account of the truth conditions of conditionals: they have none. In the following section, we will consider a difficulty for this theory.

5 Conditionals without truth conditions?

One main difficulty for the view that conditionals lack truth conditions is that conditionals appear to embed within propositions, and this seems hard to explain unless they are themselves propositions, where a proposition is, by definition, something which has truth conditions. Let us start with a case which enables us to bring the "no-proposition" view more clearly into focus:

1) John believes that if Mary is not in Rome, she is not in Italy.

In the first instance, the no-proposition view is that this is true just in case John believes that Mary is not in Italy on the supposition that she is not in Rome. The theorist might decline to say what it is to believe on a supposition, other than in terms of assigning a fairly high conditional probability. The mental state is not purely hypothetical: it is not that there is nothing special to John's belief state unless or until the supposition is realized. The negative part of the view is the denial of the orthodoxy that (**1**) affirms a relation between a person (John) and a proposition (that if Mary is not in Rome, she is not in Italy).

One whose probability function is like John's will often be ready to use the sentence

2) If Mary is not in Rome, she is not in Italy

assertively. On conventional views, John is asserting the proposition that if Mary is not in Rome, she is not in Italy. On the no-proposition view, John is engaged in an act of conditional assertion. If it turns out that Mary is not in Rome, John is committed to her not being in Italy. But what if it turns out that Mary is in Rome? Clearly John is not committed to her not being in Italy; but his assertive use of (**2**) commits him to something. If I believe what John believes, I know that there is no point searching for Mary anywhere in Italy except Rome. This knowledge is not purely hypothetical, but can immediately guide action.

Conditionals are said to embed within a variety of contexts other than those of which (**1**) is an example. Here are four troublesome kinds of case:

3) It's not the case that if you *will* to succeed then you'll succeed.

4) If you apply to us then, if you get other offers, you'll be in a position to choose.

5) If John is someone who tends to back down if challenged, then your best bet is to challenge him.

6) There is a student who, if I criticize him, will get angry (cf. Kölbel [2000]).

The cases are troublesome because they suggest that a conditional can be an element of a compound proposition in a way that would seem to require that conditionals themselves are propositions.

In many circumstances, the natural way to deny a conditional "if *A* then *B*" is to make an assertive utterance of "if *A* then not-*B*". You say that if the match is played at home then Arsenal will win, but I disagree and could naturally express my disagreement by

7) If the match is played at home, Arsenal won't win.

This poses no problem for the no-proposition view, for an assertive utterance of (**7**) can be regarded as a conditional assertion of "Arsenal won't win". However, there are also cases in which denial doesn't happily take this form. My son says that he doesn't need to go on a course: all he needs to do, in order to succeed, is to exercise his will to succeed. I am not convinced. It's not that I think that if he exercises his will he won't succeed; I merely think (**3**). This appears to be the negation of a proposition my son has affirmed; but this cannot be so on the no-proposition view. It's true that I don't have to use what appears to be a negated conditional to express my view. I could instead say something like: willing to succeed isn't a sufficient condition for success. But the fact remains that (**3**) is an entirely natural mode of expression. (Contrast Edgington [1995], pp. 283, 284.)

(**4**) appears to be a conditional with a conditional consequent. Appearances must deceive, if the no-proposition view is right: if the consequent is a conditional, it cannot as such have truth conditions,

and so cannot be asserted, even conditionally. One response is that we can understand such conditionals as having conjunctive antecedents, and so not as embedding a conditional at all. The proposal is that (**4**) really has the logical form

8) If you apply to us and you get other offers, you'll be in a position to choose.

Within the no-proposition framework there is a specific justification: there is no difference between on the one hand supposing *A* and going on to suppose *B*, and, on the other, supposing *A* and *B*.

To say that John tends to back down if challenged is pretty much to say that John is lacking in assertiveness. It would therefore be surprising if there was anything significantly more difficult about explaining (**5**) than about explaining

9) If John is lacking in assertiveness, then your best bet is to challenge him.

Yet according to the no-proposition view, while one can suppose that John is lacking in assertiveness, one cannot suppose that he will tend to back down if challenged, for what can be supposed can be true or false. According to this view, whereas one should regard an assertive utterance of (**9**) as a conditional assertion that your best bet is to challenge him, one cannot take the same view of (**5**).

One way of dealing with these cases of conditionals with apparently conditional antecedents is to claim that we "decode" sentences like (**5**) by finding a "basis" for the antecedent conditional, say John's lack of assertiveness, and then treat the sentences as if they were simple conditionals, like (**9**). This strategy will not work in all cases, and it can also backfire: as I have just suggested, some kind of equivalence between (**5**) and (**9**) is no less apt to suggest that (**5**) can be taken as genuinely a case in which a conditional is antecedent to a conditional.

Ex. 3.18 Give an example of what appears to be a conditional with a conditional antecedent for which it is less plausible to say: the antecedent introduces a basis, so the whole conditional can be understood as having a non-conditional antecedent.

This widespread phenomenon not only offers many cases in which a conditional apparently has a conditional antecedent; it also brings out a more direct conflict between the no-proposition view and naive assumptions. For example, "fragile" means something like "tends to break if dropped". It is easy to think of apparent conditionals whose antecedent ascribes fragility ("if this vase is fragile, it needs careful handling"). The more direct conflict is that we would be extremely reluctant to think that "fragile" cannot be true of some things and false of others, yet if it has the supposed conditional meaning, this ought not to be so, given the no-proposition account of conditionals.

The final case, (**6**), involves a conditional apparently embedded in an existential quantification. In uttering (**6**) assertively, I seem to make an unconditional assertion to the effect that there is a student who . . . . Intuitively, what I say is false if "if I criticize him, he will get angry" is false of each student. But on the no-proposition view, a conditional can no more be true *of* someone than it can be true.

Many of these examples pose serious difficulties for any theory, especially examples like (**5**) (cf. Edgington [1995], pp. 280 ff.). (A particular interest of (**6**) is that it does not pose special problems for propositional accounts of conditionals.) No-proposition theorists have offered various defences, though they, in common with propositional theorists, have not found a single strategy which looks plausible for all cases.

The considerations of this chapter relate to "indicative" conditionals. A somewhat more detailed exploration of the taxonomy, and a discussion of non-indicative conditionals, is in chapter **5.2**.

Bibliographical notes

Serious work in this area could well begin with Edgington [1995] which reviews the main positions, and argues for the author's "no-proposition" view. Two good collections are Harper, Stalnaker and Pearce [1981] and Jackson [1991]. Adams [1975] is the classic text for probabilities of conditionals (though the view traces back to Ramsey [1929]). A similar logic, though on a very different basis, was devised by Stalnaker [1968, 1975]; the account is briefly discussed in chapter **5.2** below. Lewis [1976b] is the classic "bombshell" paper, developed in his [1986b]. The embedding problem for the no-proposition theory is interestingly discussed in an exchange between Kölbel [2000] and Edgington [2000]. Eels and Skyrms [1994] contains some advanced material.

4

Quantification

This chapter introduces a richer artificial language, **Q**, capable of representing *quantifiers* like "all" and "some". Readers already familiar with this language should merely skim the first two sections, to check on the terminology and symbolism used here. (Of particular importance is a grasp of the precise notions of interpretation and **Q**-validity.) Later sections consider problems of formalizing English in **Q**.

1 The classical quantificational language

In **P**, the smallest unit was the sentence-letter. English sentences are composed of parts that are not themselves sentences, and this finer structure is sometimes relevant to validity. A sentence like "John runs" is composed of a *name*, "John", and a *verb*, "runs", yet the deepest **P**-formalization of this sentence is simply a **P**-letter, in which this structure is obscured. The valid argument

1) John runs; so someone runs

has no more revealing **P**-logical form than

2) p; q

which is plainly not **P**-valid.

The language **Q** is to include **P**, so that every **P**-sentence automatically counts as a **Q**-sentence, but is also to include further devices to reach structures **P** cannot reach.

Sentences of **Q** are composed of the following kinds of symbol:

Sentence-letters: p, q, r, p', . . . etc.: These are exactly as in **P**.

Name-letters: α, β, γ, α', . . . etc.: These will be used to correspond to ordinary English names like "Ronald Reagan".

Predicate-letters: F, G, H, F', . . . etc.: These will be used to correspond to English verbs, like "runs", some expressions involving adjectives, like "is hungry", and some involving nouns, like "is a man". "John runs" will be **Q**-formalized as "$F\alpha$" (to be read "α is F" or "F of α"), and so will "John is hungry" and "John is a man".

Predicates: = : The only **Q**-predicate (as opposed to predicate-letter) is "=", the sign for identity (being the same as); we can formalize "Hesperus is Phosphorus" as "$\alpha = \beta$".

Variables: x, y, z, x', . . . etc.: Their role will be explained shortly.

Operators: the sentence-connectives of **P**, together with "$\forall$" (the universal quantifier, corresponding to "all" or "every") and "$\exists$" (the existential quantifier, corresponding to "some" or "a").

Giving the meaning of the operators new to **Q** (as opposed to the sentence connectives carried forward from **P**) will be deferred until the basic ideas of how the language works have been introduced informally.

A *predicate* is an expression which takes one or more names to form a sentence. "Is a man" is a predicate which takes one name (e.g. "John") to form a sentence ("John is a man"). "Loves" is a predicate which takes two names (e.g. "John" and "Mary") to form a sentence ("John loves Mary"). "Is between . . . and . . ." is a predicate which takes three names (e.g. "Austin", "San Antonio" and "Waco") to form a sentence ("Austin is between San Antonio and Waco"). The number of names a predicate takes to form a sentence is called its *degree*. "Runs" is of degree 1, "loves" of degree 2, "is between . . . and . . ." of degree 3. Every predicate and predicate-letter of **Q** is stipulated to have some one fixed degree. (In a fully explicit notation, we could superscript every predicate and predicate letter with its degree, e.g. writing "F^3" for a predicate-letter fit to formalize a predicate of degree 3. We adopt the more flexible policy of letting predicate-letters take on whatever degree is appropriate to the task in hand.) The notion of a predicate makes no distinction among categories which traditional grammar distinguishes (e.g. verbs, nouns, adjectives) and includes expressions which traditional grammar does not recognize (e.g. "is between . . . and . . .").

Ex. 4.1 What is the degree of "="?

The *atoms* of **Q** are predicates or predicate-letters, combined with an appropriate number of names (the number equals the degree of the predicate or predicate-letter). Examples of atoms which formalize the three sentences mentioned in the last paragraph are "$F\alpha$", "$F\alpha\beta$", "$F\alpha\beta\gamma$". In the first example, "α" corresponds to "John", and "F" to the degree-1 predicate "is a man"; in the second, "α" corresponds to "John", "β" to "Mary", and "F" to the degree-2 predicate "loves"; in the third, "α" corresponds to "Austin", "β" to "San Antonio", "γ" to "Waco" and "F" to the degree-3 predicate "is between . . . and . . .".

The atoms can be compounded using the sentence connectives just as if they were **P**-letters. Thus all of the following are **Q**-sentences:

$F\alpha \rightarrow G\beta$
$(H\alpha\beta \;\&\; F\alpha) \vee p$
$F\gamma\alpha\beta \rightarrow \alpha = \beta.$

It remains to introduce the role of quantifiers and variables. Physicalists wish to affirm that everything is physical. They could use any of the following English, or more or less English, sentences:

Everything is physical.
Take anything you like, it is physical.
For all things, x, x is physical.

The standard **Q**-formalization of the English is more like the last of these variants:

3) $\forall xFx$

read "for all x, F of x". Here the *quantifier* "$\forall$" corresponds roughly to "for all things", and "F" corresponds to "physical". The *variable*, "x", plays something like the role of the pronoun "it" in "Take anything you like, *it* is physical". "Something is physical" is formalized

4) $\exists xFx$.

(3) and (4) are called, respectively, a *universal quantification* and an *existential quantification*. By contrast,

5) $\forall xFx \rightarrow \exists xFx$

is a conditional and neither a universal nor an existential quantification. This is because "$\rightarrow$" has wider scope than either quantifier, and so is the *dominant operator.* (Compare the definition of scope in (**2.1.6**).)

We can now give a fuller account of what it is to be **Q**-sentence:

6) X is a **Q**-sentence iff
 (i) X is a sentence-letter or
 (ii) X is an atom (a predicate or predicate-letter of degree n combined with n names) or
 (iii) there are **Q**-sentences Y and Z and a variable v not occurring in Y such that X is one of
  a) $\neg Y$
  b) $(Y \& Z)$
  c) $(Y \vee Z)$
  d) $(Y \rightarrow Z)$
  e) $\forall vY^{\star}$
  f) $\exists vY^{\star}$ (where $Y^{\star}$ results from Y by replacing at least one occurrence of a name by v).

We can show that (**3**) and (**4**) are **Q**-sentences as follows: "$F\alpha$" is a **Q**-sentence, being an atom; it does not contain "x"; "Fx" results from "$F\alpha$" by replacing at least one occurrence of a name by "x"; then (**6iiie**) and (**6iiif**) respectively entail that placing "$\forall x$" or "$\exists x$" before "Fx", forming (**3**) and (**4**), yields a **Q**-sentence. Given this, (**6iiid**) tells you that (**5**) is a **Q**-sentence.

7) $\exists x\forall yFxy$

is also a **Q**-sentence; we must see it as built up in stages. Starting with, say, "$F\alpha\beta$", we see that "$\forall yF\alpha y$" is a **Q**-sentence, and from this it follows that (**7**) is also.

8) $\exists y\forall yFyy$

is not a **Q**-sentence, as we can see by trying the same pattern of construction. "$F\alpha\beta$" is a **Q**-sentence, so "$\forall yF\alpha y$" is also a **Q**-sentence, as before. But we cannot go on to infer that (**8**) is, since "$\forall yF\alpha y$" contains "y", so the condition at the beginning of (**6iii**) is not met.

2 Interpretations and validity

An interpretation of **P** is an assignment of truth values to all **P**-letters. The definition of the sentence connectives ensures that every interpretation of **P** determines a truth value for *every* sentence of **P**, not just the sentence-letters. We can see the process of interpretation as containing two elements: the assignment of entities (here, truth values) to the simple expressions (the letters); and rules or definitions which bring it about that assignments to the simple expressions have determinate consequences for all the sentences.

P-validity, we saw in (**2.1.5**), is defined in terms of interpretations as follows:

> An argument in **P** is **P**-valid iff every interpretation upon which all the premises are true is one upon which the conclusion is true.

The notion of *interpretation of* **Q** is designed to play an essentially similar role in the definition of **Q**-*validity*. Each **Q**-interpretation is associated with a specific *domain* of entities. These are, intuitively, the objects the language talks about. In interpreting natural language, we may need to consider different domains in different contexts. If, as a physicalist, I say "Everything is physical", then I mean to speak of absolutely everything, or at least of all concrete things. But if, in a lecture, I say "There's no more chalk", the domain is presumably restricted to the objects in the room, or to hand: I say of all of these that none is chalk, and I do not say that no object in the whole wide world is chalk. In defining domains of interpretation for **Q**, the only restriction we place is that every domain contains at least one object. The stipulations with regard to the simple expressions are as follows:

1) In any interpretation of **Q**, with respect to any domain, D:
each sentence-letter is assigned a truth value (just as in **P**);
each name-letter is assigned one object in D (e.g. "α" might be assigned Ronald Reagan);
each predicate-letter of degree n is assigned a set of *n-tuples* (n-membered *sequences*) of objects in D (e.g. if "F" is of degree 2, some interpretation will assign to it a set of *ordered pairs* (2-membered sequences) such that the first member of each pair loves the second);

"=" is assigned the set of ordered pairs of members of D such that in each pair the first object is the same thing as the second; *variables* are not assigned anything.

Expressions of the form "$i(\ldots)$" can be used to denote what some **Q**-interpretation, i, assigns to some expression $(\ldots)$. Instead of saying that i assigns Ronald Reagan to "α", we can say: $i(\alpha)$ = Ronald Reagan.

In virtue of the rules for the operators, an interpretation of the simple expressions determines a truth value (relative to that interpretation) for every **Q**-sentence. (For the sentence connectives, these are just the rules from **P**, but they are restated for completeness.) A preliminary notion:

> The expression "X^n_ν" is to be read "the result of replacing every occurrence of the variable ν in X by the first name-letter n not occurring in X".

Talk of the *first* name-letter is to ensure that there is a unique result of the transformation. It presupposes some conventional (e.g. alphabetical) ordering of the name-letters. For example, if X is $G\alpha x$, then X^n_ν is $G\alpha\beta$.

2) (i) $\neg X$ is true upon an interpretation iff X is false upon that interpretation.
   (ii) $(X \,\&\, Y)$ is true upon an interpretation iff X is true upon that interpretation and so is Y.
   (iii) $(X \vee Y)$ is true upon an interpretation iff X is true upon that interpretation or Y is true upon that interpretation.
   (iv) $(X \rightarrow Y)$ is true upon an interpretation iff X is false upon that interpretation or Y is true upon that interpretation.
   (v) An atom is true upon an interpretation whose domain is D iff the sequence of the object(s) from D assigned to the name-letter(s) by the interpretation belongs to the set it assigns to the predicate or predicate-letter; it is false iff the sequence does not belong to the set.
   (vi) $\forall\nu X$ is true upon an interpretation whose domain is D iff X^n_ν is true upon that interpretation, and also upon every other interpretation whose domain is D and which agrees with that one except in point of what is assigned to n; otherwise it is false. If we say that an n-variant of i is an interpretation which agrees with i on its domain and on all its assignments except, perhaps, that to n, we can abbreviate this condition as follows: $\forall\nu X$ is true upon an interpretation, i, whose domain is D iff X^n_ν is true upon every n-variant of i; otherwise it is false.
   (vii) $\exists\nu X$ is true upon an interpretation, i, whose domain is D iff X^n_ν is true upon some n-variant of i; otherwise it is false.

I will offer some explanation of the **Q**-specific clauses (v), (vi) and (vii).

For (v), consider an atom, say "$G\beta$". Suppose that an interpretation assigns the set of 1-membered sequences whose members are presidents of the USA to "G" and Ronald Reagan to "β". Then "$G\beta$" is true upon this interpretation, since the 1-membered sequence <Reagan> does belong to the set the interpretation assigns to "G". A given **Q**-sentence may be true upon one interpretation, false upon another. Consider an interpretation which assigns Reagan to "β" but the set of 1-membered sequences each of whose members is a Chinese emperor to "G". "$G\beta$" is false upon this interpretation.

For the point of the talk of sequences in (v), consider the **Q**-sentence "$F\alpha\beta$" and an interpretation i which assigns to "F" the set of ordered pairs (2-membered sequences) such that the first member of each pair loves the second. Suppose that the interpretation assigns John to α and Mary to β: in shorthand, suppose that $i(\alpha)$ = John and $i(\beta)$ = Mary. Then (v) rules that $i(F\alpha\beta)$ is true iff the ordered pair <John, Mary> belongs to $i(F)$, that is, iff John loves Mary. *Order* is crucial here: it may be that <John, Mary> belongs to $i(F)$, yet <Mary, John> does not.

Turning to (vi), intuitively we want to say that a universal quantification in English, say "Everything is physical", is true just on condition that "physical" is true of everything. Applying this to **Q**, we want to say that "$\forall xFx$" is true upon an interpretation, i, just on condition that i assigns to "F" the set of all the 1-membered sequences that can be formed out of members of its domain. (vi) instructs us to consider the sentence "$F\alpha$" which results from "$\forall xFx$" by deleting the quantifier with its adjacent "x", and replacing the other occurrence of "x" by the first name-letter not in "$\forall xFx$"; it holds that our target sentence is true upon i iff "$F\alpha$" is true upon all α-variants of i: all interpretations which agree with i on their domain and on what set they assign to "F", though perhaps differing from i in what they assign to "α". For each object in the domain of i, D, at least one of these interpretations will assign "α" to it. Just on condition that "$F\alpha$" is true whatever object "α" is interpreted as standing for, that is, just on condition that "$F\alpha$" comes out true upon all these interpretations, "$\forall xFx$" is true upon i. "$\forall xFx$" is true upon any interpretation that assigns to "F" the set containing all the singleton sequences which can be formed from the D-members; that is, it is true upon any interpretation, i, such that for every object, x in D, $\langle x \rangle$ belongs to $i(F)$. It is, for example, false upon an interpretation with respect to the domain of all persons that assigns to "F" the set of all 1-membered sequences whose member is happy.

Talk of sets of sequences in connection with assignments to monadic predicate-letters is tiresome, and is needed only in order to achieve a smooth general statement of truth-upon-an-interpretation conditions, applicable to predicates of arbitrary degrees. Otherwise one could think of an interpretation as associating monadic predicate-letters with subsets of its domain, and I will routinely adopt this convention when convenient. Then one could say, for example, that "$\forall xFx$" is true upon any interpretation which assigns its domain to "F".

(vii) works just like (vi), except that instead of *every* n-variant we now talk of *some* n-variant. Simplifying as in the previous paragraph, "$\exists xFx$" is true upon any interpretation which assigns to F a non-empty set.

The definition of **Q**-validity is as follows:

3) An argument is **Q**-valid iff every interpretation upon which all the premises are true is one upon which the conclusion is true.

Since each interpretation has a domain, and since every domain is the domain of many interpretations, there is an implicit quantification over domains in this definition.

There are a number of different ways in which semantics for **Q** can be given. One motivation for my choice is that I wanted to bring out the unity of functioning of the sentence connectives as they occur in **P** and as they occur in **Q**. Consider a **Q**-sentence like

4) $\exists x(Fx \,\&\, Gx)$.

By our present criteria, "&" is not a sentence connective in (**4**) since "Fx" and "Gx" are not sentences. Some people like just to stipulate that they *are* sentences, but this loses the connection between something's being a sentence and its being usable in a complete act of communication, for example, to make an assertion: you cannot use the expression "x is a man" to make an assertion; for who is x? On the semantics given here, "&", even as it occurs in (**4**), makes its contribution to truth conditions through functioning as a genuine sentence connective connecting the **Q**-sentences "$F\alpha$" and "$G\alpha$". The fact that it can be treated by exactly the same interpretation rule for both languages shows that its semantic functioning is the same in both.

The present account is closely related to Tarski's [1937], but is simplified: truth is defined directly rather than via satisfaction, and infinite sequences have been avoided. The difference is that his language makes no distinction between name-letters and variables. He therefore counts "open sentences", sentences containing variables not bound by a quantifier, as sentences which receive truth-upon-an-interpretation conditions. This means that there is not an immediate correspondence between the kind of expression which has truth-upon-an-interpretation conditions and the kind which has truth conditions. On both Tarski's account and the present variant, what matters to the truth of a universal quantification is that all the objects in the domain be a certain way.

This initially contrasts with an approach, sometimes associated with Frege, which begins by identifying the truth of a universal quantification with the truth of its instances: "$\forall xFx$" is true iff every instance ($F\alpha$, $F\beta$, etc.) is true. This makes the truth of quantifications depend upon what instances are available in a language. As the account stands, to get the right result one would need to ensure that there is an instance in the language for every object in the world; since there are very many objects, there would be very many names in such a language, which would make it unlike any natural language. This is just a technical difficulty, which can be overcome by talking of extensions of the language (for any object there will be some small extension of the actual language which will contain a name for it) but the net result has no advantages over the present approach.

3 Universal quantification

What makes a **Q**-formalization of an English sentence adequate? The "truth conditions" test provides only a necessary condition: the truth-upon-an-intended-interpretation conditions of the **Q**-sentence should match the truth conditions of the English sentence. The other test, the "same saying" test, comes close to sufficiency: the recovered sentence or argument – the result of replacing the name- and predicate-letters by the English names and predicates specified in the correspondence scheme, and replacing the **Q**-operators by their stipulated English correlates – should say the same as the original. An intended interpretation is one which assigns to the name-letters the same entities as those to which the corresponding names (in English) refer, and to the predicate-letters the same sets as the sets of things in the domain of which the corresponding predicates (in English) are true. The relevant correspondence is that given by the correspondence schema associated with each formalization. Truth-upon-an-interpretation conditions typically cannot be the same as the truth conditions of a quantified English sentence, for the former is specific about the domain of quantification in a way that the latter usually are not. I say that "match" obtains when the interpretation is over a domain that is appropriate to the English sentence.

The formalization of "Everything is physical" by "$\forall xFx$" meets both tests relative to an interpretation that assigns the set of all physical things in its domain to "F". The **Q**-sentence is true on such an interpretation, i, iff $i(F)$ is its domain (that is, is the same set as the interpretation's domain). This matches the truth conditions of the English, setting aside the potentially greater specificity of domain, so the formalization passes the truth conditions test for adqequacy. (The English sentence does not make explicit the domain of the English quantifier (should it, perhaps, include only concrete things?), so various interpretations, differing in their domains, will count as intended.) Moreover, the recovered sentence ("For all x, x is physical") says the same as the original English, so the formalization also passes the same saying test.

How should sentences like

1) All men are happy

be formalized? The standard answer is:

2) $\forall x(Fx \rightarrow Gx)$,

with "F" corresponding to "is a man" and "G" to "is happy". Despite the fact that **(2)** contains no visible occurrence of "if", it is treated as if it were the same as

3) For any object, x, if x is a man then x is happy.

Is this adequate?

It is not clear how one should proceed in order to test its adequacy relative to the same saying test. However, we can use the truth conditions test. Take an intended interpretation, i, with some domain, D. It will assign to "F" the set of men in D, and to "G" the set of happy things in D. From **(2.2**vi) we know that "$\forall x(Fx \rightarrow Gx)$" is true upon i iff "$F\alpha \rightarrow G\alpha$" is true upon every α-variant of i. These interpretations fall into three classes:

(i) Those which assign something in D other than a man to "α". Since "$F\alpha$" is false on all these interpretations, "$F\alpha \rightarrow G\alpha$" is true upon all of them, by **(2.2**iv).
(ii) Those which assign a happy man in D to "α". "$F\alpha \rightarrow G\alpha$" is true upon all these interpretations, by (**2.2**iv)
(iii) Those which assign an unhappy man in D to "α". "$F\alpha \rightarrow G\alpha$" is false upon all these interpretations, by **(2.2**iv).

The upshot is that "$\forall x(Fx \rightarrow Gx)$" is true upon i iff no α-variants belong to case (iii), iff there are no unhappy men in the domain of i. So all we have to decide, to determine whether the formalization passes the truth conditions test of adequacy, is whether **(1)** is true iff there are no unhappy men.

There is no doubt that **(1)** is false if there are unhappy men. The doubt concerns whether there being no unhappy man is sufficient for its truth. Suppose there were no men at all. Then there would be no unhappy ones. But would this make **(1)** true?

Peter Strawson ([1952], pp. 173–6) has argued that if John has no children then

4) All John's children are asleep

is not true, whereas a **Q**-formalization of **(4)** as **(2)** would be true upon an interpretation assigning to "F" the set of children of John and to "G" the set of all sleepers. The first set would be empty, and so all the relevant interpretations will make "$F\alpha$" false, and thus verify "$F\alpha \rightarrow G\alpha$". On Strawson's view, truth conditions would not match truth-upon-an-intended-interpretation conditions.

A problem with the suggestion that (4) is not true if John is childless is that it is unclear how it should be generalized. We think the following is true, despite the fact that there are no bodies acted on by no forces:

5) All bodies acted on by no forces continue in a uniform state of rest or motion.

So the doctrine cannot be that *no* sentences of the form of (4), and whose first predicate is true of nothing, are true.

The situation is markedly similar to the case of conditionals with false antecedents. Indeed, on the standard kind of **Q**-formalization, the relevant problem is precisely the same. So the standard dialectic can be replayed. For example, it may be suggested that (4) conversationally implicates that John has children, but does not entail it. To support this, it may be urged that even a community whose natural language had **Q**-quantifiers would tend to utter sentences of the form "$\forall x(Fx \rightarrow Gx)$" only if they believed that something is F, because they would otherwise be more informative by saying that nothing is F.

This response does not do justice to all the issues. As well as being affirmed, the relevant sentences express things which we may believe or disbelieve; there need be no conversation, and so no operation of a conversational mechanism. One who thinks that (4) is not true, given that John has no children, will either say that there is no corresponding belief or that the corresponding belief is not true.

Moreover, the response does not do justice to the fact that we have different attitudes to different universal quantifications, alike in having an empty first predicate. For example,

6) All bodies acted on by no forces undergo random changes of velocity

is intuitively false, yet its standard **Q**-formalization as (2) would be true upon an intended interpretation (one assigning to "F" the set of bodies acted on by no forces).

A defence of (2) as **Q**-formalization of both (5) and (6) might draw on the idea that in some contexts the assertion of a quantified conditional (as (5) and (6) are taken to be) implicates that what is asserted is a law of nature. Since this is a false implicature in the case of (6), we think of (6) as unassertible and then, confusedly, come to think of (6) as false. A hearer, in order to understand a speaker's full intentions in uttering (5) and (6), would have to recognize him as intending to state a natural law. As one sentence expresses a natural law and the other does not, there is room for a contrast between them, a contrast which need not be reflected in truth values. The full development of the defence needs to do justice to the fact that (4) does not purport to express a natural law. Perhaps (5) and (6), but not (4), are reformulable in counterfactual terms:

If any bodies were acted on by no forces, they would continue in a uniform state of rest or motion

and

If any bodies were acted on by no forces, they would undergo random changes of velocity.

The "implicates natural law" idea would be applied just to the counterfactually reformulable quantifications, leaving some other story to be told for quantifications like (4). It is not easy to see how these various ideas could be applied in a systematic and illuminating way.

There are more straightforward problems with formalizing universal quantifications in English.

7) Not all men are happy

is standardly formalized as

8) $\neg\forall x(Fx \rightarrow Gx)$.

On the other hand

9) All men are not happy

is probably ambiguous. On one reading, encouraged by heavy stress on "all", it is equivalent to (7), and so to be formalized as (8). The other reading, in my view more correct, treats it as appropriately formalizable as

10) $\forall x(Fx \rightarrow \neg Gx)$

and thus as equivalent to

11) No men are happy.

"Every", "any" and "whatever" are often to be formalized by "$\forall$". Thus (2) ($\forall x(Fx \rightarrow Gx)$) is standardly held to formalize all of:

12) Every person in the room was happy.

13) Any person who interferes will be shot.

14) Whatever you buy you will charge to me.

This formalization is not always the "deepest". For example, it does not discern a possible conjunctive structure in "person in the room". A candidate for a deeper formalization of (**12**) is

15) $\forall x(Fx \,\&\, Gx \rightarrow Hx)$

where "F" corresponds to "is a person", "G" to "is in this room" and "H" to "was happy".

Ex. 4.2 (a) The following arguments present problems for the standard **Q**-formalizations of English universal quantifications. Formalize the arguments, showing your correspondence scheme, state the problems, and briefly indicate how, if at all, you think they might be resolved:

(i) It is not the case that all bodies acted on by no forces undergo random changes of velocity. Therefore there are bodies acted on by no forces.

(ii) If anyone plays cricket, he does not also play squash. So anyone who plays both cricket and squash does not play cricket.

(iii) Every parent who loves all of his children is saintly. So every person who loves all of his children is saintly. (Cf. McCawley [1981], p. 163.)

(b) Formalize the following in **Q**, noting any problems or deficiencies. The only quantifier to be used is "$\forall$".

(i) If Pedro owns a donkey, he beats it.

(ii) If Pedro owns a donkey, he is lucky.

(iii) Old soldiers never die.

(iv) It is not invariably the case that love is requited. [α's love for β is requited iff, in addition, β loves α.]

(v) Anyone who needs something should have it.

(vi) Jones never leaves his desk.

(vii) If *a* is greater than 0, 0 is less than *a*.

(viii) None but the brave deserve the fair.

(ix) Someone who ever lies is someone you should never trust.

(x) If no one telephones her Jane will be miserable.

(xi) No one runs faster than John.

(xii) No one runs faster than himself.

(xiii) Not everyone runs faster than himself.

(xiv) If all the people in the world were stretched end to end, they would circle the globe.

4 Existential quantification

The English sentence

1) Someone is happy

is standardly formalized

2) $\exists xFx$.

Given the correspondence of "F" to "is happy", an intended interpretation, *i*, assigns the set of all happy things in its domain to "F". (2) is true upon *i* iff some interpretation with the same domain making the same assignment to "F" assigns someone happy to "α", and this is intuitively the correct truth condition.

How should sentences like

3) Some elephants are greedy

be formalized? A standard answer is:

4) $\exists x(Fx \,\&\, Gx)$,

pronounced: "there is an x such that F of x and G of x". Despite the fact that (3) contains no visible occurrence of "and", it is treated as if it were the same as

5) Something is both an elephant and greedy.

On an interpretation i which assigns to "F" the set of elephants, and to "G" the set of greedy things, (4) is true iff at least one α-variant of i assigns a greedy elephant to "α". The formalization represents (3) as true iff at least one thing is greedy and an elephant. (3) appears to imply that there is more than one greedy elephant, and this aspect is not preserved in the formalization. Moreover, some have said that "some" implies "some but not all", so that (3) would imply that some elephants are not greedy. These alleged implications appear stronger in some cases than others. For example,

6) She ate some of the cakes

seems to imply quite strongly that some cakes were left and that more than one was eaten.

The "some but not all" implication could be approached in at least two ways. One could explicitly add to the formalization of, say, (3) that not all elephants are greedy. This is simply the denial of "All elephants are greedy", and so the formalization proposed is

7) $\exists x(Fx \; \& \; Gx) \; \& \; \neg\forall x(Fx \rightarrow Gx)$.

Alternatively, one could say that the implication is a matter of implicature only, and so need not be registered in the formalization. To support this view, one could say it would be misleading to use "some" if one knew something expressible by "all": misleading, but not false. The cancellability of the implicature would support this view:

8) She (certainly) ate *some* of the cakes – *all* of them in fact

is consistent.

The implicature story does not work well for the implication from "some"-sentences to "more than one"-sentences. Since, for example,

9) She ate a cake

does not entail

10) She ate some cakes,

they cannot both correctly be given the same formalization, viz. (4). Similarly, we do not have the cancellability one would expect from implicature, since the following is inconsistent:

11) She ate some of the cakes – in fact, just one.

Compare also

12) A hungry man is at the door

and

13) Some hungry men are at the door.

While the first could be adequately formalized along the lines of (3), the second could not be.

We should not infer that the meaning of "some" differs from that of "∃". The entailment in question need not be seen as due to "some", because where it obtains it is adequately explained by the presence of the plural nouns ("cakes" rather than "cake", "men" rather than "man"). In §9 we will discuss how the effect of plurals can be captured in **Q**. Probably the safest way to read "$\exists x$" is "There is at least one thing, x, such that"; and we can be confident that, in using "∃" in the manner of (4), we are adequately formalizing English sentences containing quantifiers like "some", "something", "a" – as in "A man is here to see you" – just on condition that we can rephrase without distortion in terms of "There is at least one thing such that . . .".

Ex. 4.3 Consider the following pair:
(i) Some Buddhists are vegetarians.
(ii) Some vegetarians are Buddhists.
Is the apparent difference between them one which could affect the validity of some argument? Can the apparent difference be reflected in different Q-formalizations? If so how? If not, does this give a new reason for thinking that the meaning of "some" differs from that of "∃"? (Cf. McCawley [1981], p. 123.)

There is no need for **Q** to contain both "∀" and "∃", since they are interdefinable. This corresponds to the facts about English that "everything is . . ." means "it is not the case that something fails to be . . ." and that "something is . . ." means "it is not the case that everything fails to be . . .".

Ex. 4.4 (a) Assuming "∀" is primitive in **Q**, show how you could introduce "∃" by a definition.

(b) Assuming "∃" is primitive in **Q**, show how you could introduce "∀" by a definition.

(c) Formalize the following in **Q**, noting any problems or deficiencies:
- (i) If Pedro owns a donkey he is lucky. [Do not use "∀".]
- (ii) There is a town between Oxford and London.
- (iii) I met a man.
- (iv) A puppy is a young dog.
- (v) Some men are touchy and vain.
- (vi) Some people have everything.
- (vii) Every cloud has a silver lining.
- (viii) When beggars die there are no comets seen.
- (ix) There is a skeleton in every cupboard.
- (x) Nothing lasts for ever.
- (xi) Jones never feeds any of his dogs.
- (xii) Who laughs last laughs longest.

5 Adjectives

A sentence like

1) Tom is a greedy man

is usually formalized as

2) $F\alpha \ \& \ G\alpha$,

with "F" corresponding to "is greedy", "G" to "is a man" and "α" to "Tom". The adjectival construction of the English, in which the adjective "greedy" qualifies the noun "man", is rendered as a conjunction in **Q**. The truth conditions appear to come out correctly. (**1**) does indeed seem equivalent to

3) Tom is a man and is greedy.

Moreover, upon an interpretation which assigns to "F" the set of men, to "G" the set of greedy things, and to "α", Tom, (**2**) is true iff Tom is a man and is greedy, which seems correct with respect to (**1**).

In **Q**, adjective phrases like "is greedy", noun constructions like "is a man", and verbs like "runs" are treated as all of essentially the same kind – predicates – and hence are matched with predicate-letters in **Q**-logical forms. This policy ensures that some familiar English adjectival constructions cannot be adequately formalized in **Q** to a reasonable depth. Consider

4) Tom is a large man

and suppose we formalize this as (**2**): "$F\alpha \ \& \ G\alpha$", with "F" corresponding to "is a man" and "G" to large. Continuing this policy, how should we formalize

5) Tom is a businessman but not a large businessman?

The policy dictates:

6) $H\alpha \ \& \ \neg \ (H\alpha \ \& \ G\alpha)$

with "H" corresponding to "is a businessman" and "G" to "large". However, (**4**) and (**5**) are consistent, whereas (**2**) and (**6**) are not. Hence the policy of formalizing (**4**) and (**5**) by the conjunctive treatment of adjectives has failed. It has wrongly represented consistent sentences as inconsistent.

Ex. 4.5 Explain why (**5.2**) and (**5.6**) are inconsistent.

There is a class of adjectives like "large" (and "heavy", "expensive", etc.) which resist **Q**-formalization. And there are other resistant adjectives. For example, it is obvious that

7) Tom is an alleged murderer

is not to be formalized as (**2**) (with "F" corresponding to "is a murderer" and "G" to "alleged"). Although it would be reasonable to formalize

8) Tom loves only happy women

as

9) $\forall x(H\alpha x \rightarrow (Fx \; \& \; Gx))$

(with "F" corresponding to "is a woman", "G" to "is happy" and "Hxy" to "x loves y"), it would not be reasonable, but nonsense, to adopt the same formalization of

10) Tom loves only three women.

An intended interpretation would assign to "G" the set of all things which are three. But it is nonsense to speak of an object (one single object) *being three*. There is no such property. I discuss in §9 how adjectives like "three" are **Q**-formalized.

In general, we can reflect adjectival modification by conjunction only if, where "n" is a name, "A" an adjective, and "C" the common noun it qualifies, the following is true (and has an intelligible conclusion):

11) n is a (or an) $AC \models n$ is A and n is a (or an) C.

Adjectives (or adjectival phrases) satisfying (**11**) (like "happy", "greedy", "red", "weighs 12 pounds") are called "*predicative*". Among non-predicative adjectives, there are various distinctions. There is a category of adjectives which qualify other adjectives, for example "dark" (qualifying colour adjectives), and these, though ruled non-predicative by (**11**), behave rather differently from "large". "Three" and "alleged", though both classified as non-predicative, differ in that while "three" can be adequately represented in **Q**, "alleged" cannot.

Ex. 4.6 Show by example that the following suggestion is mistaken: If "weighs 20 stone" is predicative "heavy" must also be predicative.

Sentences containing non-predicative adjectives, like (**4**), can be **Q**-formalized. We could let "F" correspond to "is a large man" and formalize (**4**) as "$F\alpha$". We can make sense of the idea of an interpretation which assigns to "F" the set of large men. However, this **Q**-formalization is not deep enough to bring out the distinctive contribution of "large". "Is a large man" is complex, composed of "large" and "man". This complexity of structure is not revealed in the envisaged **Q**-formalization.

Ex. 4.7 Explain why we cannot make sense of an interpretation which assigns to "F" the set of large things.

The formalization could not reflect the validity of arguments like:

12) Tom is a small businessman. Therefore Tom is a businessman.

The most revealing **Q**-formalization would be

13) $F\alpha$; $G\alpha$

(with "F" corresponding to "is a small businessman" and "G" to "is a businessman"), which is plainly **Q**-invalid. Does this matter? The idea of artificial languages was to capture generalizations about *formal* validity, but, by the standards set in chapter **1.10** and **1.11**, (**12**) is not formally valid. While this may give the **Q**-formalizer relief from the burden of finding a formalization of (**12**) which reflects its validity, it also carries with it an objection concerning cases in which there are validity-reflecting **Q**-formalizations. For example, the validity of

14) Tom is a happy man. Therefore Tom is a man

is reflected by the validity of the obvious **Q**-formalization. Hence the conjunctive method of **Q**-formalization has captured more validity than was intended: it has reflected as *formally* valid an argument that, though valid, is not *formally* valid, according to our original definition. There is a tension here. The original definition of formal validity could be revised, or else the project of formalization recharacterized.

Ex. 4.8 Explain why (**5.12**) does not count as formally valid, by the standard of chapter **1**. Is there any other standard of formal validity which would be more appropriate?

I have given examples of supposedly predicative adjectives – "happy", "greedy", "red", "weighs 12 pounds" – but a doubt remains. Are all these adjectives really predicative? One sign of non-predicativity comes from the pattern discerned in the case of "large":

you can be a mouse and an animal and a large mouse without being a large animal. Suppose that humans are on the whole rather less greedy than most animals. Then you could be a man and an animal and a greedy man without being a greedy animal. Like "large", there is a relativization to a comparison class, which is introduced by the noun, and which could be expressed more explicitly by saying "He's large *for* a mouse, but not large *for* an animal", "He's greedy *for* a man, but not greedy *for* an animal". So perhaps "greedy", and by similar considerations "happy", are, after all, non-predicative.

Ex. 4.9 Can you find expressions T_1 and T_2 such that there is no inconsistency in supposing that someone could be happy for a T_1 but not happy for a T_2?

The difference between "large" and "greedy" is this: "large", associating with the noun it qualifies, thereby determines, at least in part, the appropriate standards of size to be applied. This does not hold for "greedy", which in this resembles adjectives like "tall". Even when "tall" occurs modifying a noun, as in "Tom is a tall man", it may not be this noun which determines the relevant comparison class. That determination is left to context. Being tall for a Swede involves being taller than does being tall for an Eskimo.

Does the growing shadow of non-predicativity extend even to an adjective like "red"? It's one thing for a house to be red (quite in order for the windows, doors, mortar and interior not to be red), another for a colour sample to be red (should be uniform red all over), another for a tomato to be red (*painted* red doesn't count) and yet another thing for hair to be red. Correct as these observations are, they do not establish non-predicativity. There is no general term, T, not itself having colour entailments, yielding the pattern: "This is a tomato and a T and is a red tomato but not a red T". Moreover, (**11**) holds for "red" combined with any appropriate noun. So the phenomenon is not non-predicativity. Although the different nouns it qualifies impose different standards for what is to count as a red thing of that kind, it appears impossible for a thing to belong to two different kinds and count as red by the standards appropriate to one and as non-red by the standards appropriate to the other. With non-predicativity, like that of "large", by contrast, the different possible comparison classes can give different verdicts with respect to the applicability of the adjective.

Ex. 4.10 (a) Formalize:
This is a gold ingot. Therefore it is made of gold.
(b) Give three examples each (not in the text) of predicative and non-predicative adjectives.
(c) Provide **Q**-formalizations of the following, noting any inadequacies:
(i) Some greedy men are vain.
(ii) Some vain men are greedy.
(iii) John loves a beautiful actress.
(iv) John loves a former actress.

6 Adverbs

In traditional grammar, an adverb or adverbial phrase is seen as modifying a verb. No such construction is available in **Q**. Hence there is no straightforward validity-reflecting **Q**-logical form of, for example,

1) John walked quickly. Therefore John walked.

We could merge "walked" with "quickly", and formalize "walked quickly" by a single predicate letter; but we would have to choose a different letter to correspond to plain "walked", as it occurs in the conclusion, so the formalization would patently be **Q**-invalid.

We can reach a deeper formalization by a roundabout route, which has been proposed by Donald Davidson. It seems that

2) John walked

is true iff *there is* something – a walk – which John walked. So, though there is no visible existential quantifier in (**2**), we could offer the following formalization of it:

3) $\exists x F\alpha x$

with "Fxy" corresponding to "x walked y", and "α" to "John".

John's walk might be quick or slow, uphill or downhill, with a stick or with a dog. This suggests that (**1**) should be formalized:

4) $\exists x(F\alpha x \,\&\, Gx)$; $\exists x\; F\alpha x$

with "F" and "α" as before, and "G" corresponding to "is quick". It is perfectly in order for an interpretation of "G" to assign a set of quick things: quick walks and quick swims and quick runs, but not slow walks, etc., and not trees or lamp-posts, since these things cannot be quick.

(**4**) is **Q**-valid. The premise says that there is something with two properties – being F to α and being G – and it is clearly correct to infer from this that there is something with just the first of these properties.

In general, the sort of thing that can be walked or can be quick is an *event*. We can sum up Davidson's proposal as involving two theses: (i) many verbs (those like "walk", "run" and "swim") introduce events, and many sentences containing these verbs, like (**2**), introduce (implicit) existential quantification over events; (ii) adverbs are adjectives which qualify events. Thus Davidson's proposal sees the premise of (**1**) as containing an implicit existential quantifier, and sees "quickly" as an adjective whose intended interpretation associates it with a subset of events. (The notion of a sentence containing an "implicit" occurrence of a quantifier is discussed in chapter **6**, especially **6.6**.)

How widely can this account be applied? One class of adverbs can be exempted from the account at the outset: those which, like "allegedly", "necessarily", "probably" and "rarely", modify sentences rather than verbs: they are *sentence adverbs* (cf. Taylor [1985]). It is plain that in

5) John allegedly shot Jane

we are not saying that there is (really is!) a shooting with the property of being alleged. Rather we are saying something like: it is alleged that John shot Jane. We are not asserting that a shooting has a certain property, but rather qualifying a whole sentence ("John shot Jane"). Davidson ([1967b], p. 121) says that his account is not intended to apply to any sentence adverbs.

To bring out the difference between sentence adverbs and the kind with which we are concerned, consider the ambiguity of

6) John trod carelessly on a snail.

If we hear "carelessly" as a sentence adverb, we will regard this as equivalent to

7) It was careless of John to tread on a snail.

If we hear "carelessly" as genuinely modifying "trod", we will reject this equivalence: John may have set out with full deliberation and care to tread on a snail, so (**7**) is false, but in the end executed his plan carelessly, so one reading of (**6**) is true.

Ex. 4.11 What considerations, if any, should incline one to classify "necessarily", "probably" and "rarely" as sentence adverbs?

There is another resistant category of adverbs, exemplified by "intentionally". We cannot paraphrase "John intentionally trod on a snail" as "It was intentional that John trod on a snail", since the latter is ungrammatical. So "intentionally" is not a sentence adverb. Yet we cannot see the adjective "intentional" as true of events. The very same event may be an intentional switching on of the light and an unintentional scaring off of an intruder. As Davidson says ([1967b], p. 121), events as such can neither be intentional or otherwise. Hence the adverb "intentionally" cannot be treated as an adjective applying to events.

Does Davidson's account hold of all adverbs which are neither sentence adverbs nor ones like "intentionally"? No: for in some cases – including "quickly" – the adjective that would be discerned on Davidson's account is non-predicative, and hence its distinctive contribution cannot be captured in **Q**. John swam the channel, and thereby broke the record for cross-channel swims. On Davidson's account

8) John swam the channel quickly

should be true in virtue of whatever swim verifies (**8**) being a quick one. On the other hand, compared to most channel-crossings, John's self-propelled crossing was slow. So we also have as true

9) John crossed the channel slowly.

But the very same event is both the swim and the crossing, and verifies both (**8**) and (**9**). If we applied Davidson's account, we would be committed to the view that some particular event (John's swim, that is, his crossing) is both quick and slow, which is unacceptable.

Davidson's account should therefore be applied only to those adverbs not in any of the three resistant categories we have mentioned, viz.: sentence adverbs, those like "intentionally", and those corresponding to non-predicative adjectives. Whether there is a positive characterization of the remaining adverbs, and whether they genuinely form a special category of adverb, are questions we will not pursue. Failure of predicativity, or at least some closely related phenomenon, may be more prevalent than one might at first suppose (cf. Wiggins [1985]). For example, a walk can be uphill, and "uphill" appears predicative. A walk can also be a warning. But it is open to question whether it is intelligible to qualify a warning as an uphill one. We seem to have the pattern of non-predicativity: something is a walk and a warning and an uphill walk but not an uphill warning.

Ex. 4.12 (a) Formalize the following, stating your correspondence scheme, and noting any inadequacies:

(i) John ambled down the hill.
(ii) Quickening his pace, John strode up the hill.
(iii) John worked for the whole night.
(iv) John worked in Boston for four years.

(b) Davidson has argued that pairs of sentences like

(i) Shem kicked Shaum. It happened in the gymnasium.

constitute evidence in favour of the view that sentences like the first in the pair introduce events. For the "it" of the second sentence appears to be linked to the first sentence in the way that "he" is linked to "a man" in the pair

(ii) I met a man. He was bald.

(The "he" here is called an "anaphoric" pronoun.) Formalize (i), showing your correspondence scheme, and say whether your formalization supports Davidson's claim.

7 Names

Names in English, like "Tom" and "London", are supposed to be **Q**-formalized by the use of name-letters. What is the common feature of names that justifies this common treatment? When we have answered

this question, we will be in a position to say whether expressions like "the present Prime Minister of Great Britain", "water", "Pegasus" and "Vulcan", which in many ways resemble names, should also be formalized by name-letters.

In **Q**, a name-letter is simply assigned an object (in the domain of interpretation). So one would expect the category of English names to be a category of expressions which stand for objects. "Tom" and "London" fall into this category. But may it not turn out to include many other expressions as well? For example, why does not an expression like "happy", in the sentence "Tom is happy", count as standing for an object: the property of being happy, or, perhaps, the set of happy persons? Indeed, predicate-letters are interpreted in **Q** by being assigned objects (viz. sets), so should we not also construe English predicates as standing for objects, and thus, by the proposed criterion, counting as names?

A predicate takes a name to form a sentence. If predicates were a species of name, then a sentence like "Tom is happy" would be construed as two names juxtaposed. But two names juxtaposed do not make a sentence, only a list. Hence predicates cannot count as names.

This argument does not prevent one seeing common nouns as names, and analysing, say,

1) Tom is a man

as

2) $F\alpha\beta$

with "α" corresponding to "Tom", "β" to "the property of being a man", and "F" to "having" (as when we say an object *has* a property). I know of no decisive objection to the policy, though we will not follow it.[1] Pursuing it cannot construe all predicates as names: there remains the predicate "having", matched to the predicate-letter "F".

Expressions like "the present Prime Minister of Britain" are called "*definite descriptions*". Many of them stand for objects, like the one just

[1] There would be a decisive objection if there were no properties: cf. Armstrong [1978].

mentioned, so they are prima facie cases of names. Some, like "the golden mountain", do not. Hence if all definite descriptions are to be treated in the same way, none of them should be formalized by name-letters. Other reasons for not formalizing them by name-letters are given in §**10**.

Does "water" stand for an object? One is inclined to say yes: of course, "water" stands for water! But what is this object, water? Not, presumably this puddle or glassful, but the scattered totality of water throughout the universe. There is no objection to allowing this totality to count as an object. Most objects are spatially cohesive in the sense that between any two spatially separated parts of the object lies a route (not necessarily a straight line) at each point of which is another part of the object. But there is no need to insist that all objects must be like this. So far, then, "water" passes the test for being a name.

There are cases in which it is unnatural to treat it as a name, for example:

3) There is some water here.

One could try formalizing "water" as a predicate:

4) $\exists x(Fx \;\&\; Gx\alpha)$

with "Fx" corresponding to "x is water", "Gxy" corresponding to "x is present at y" and "α" corresponding to "here". There is something suspect: if we use the test of reading "$\exists$" as "there is at least one thing which", we find we are treating (**3**) as saying "There is at least one thing which is water here". Likewise, an interpretation would have to assign to "F" a non-empty set of waters, or things which are water. What things are these? Pools and lakes? If so, we misrepresent (**3**) as equivalent to something like "there is at least one pool (or lake, or quantity) of water here". If "F" is assigned the single scattered object, water, one could as well revert to the idea that "water" is a name, and adopt a formalization like:

5) $\exists x(Fx\beta \;\&\; Gx\alpha)$

with the correspondences for "G" and "α" as before, "β" corresponding to "water" (now treated as a name) and "Fxy" corresponding to "x is part of y". Again, this is not a perfect mirror of (**3**), for it treats it as equivalent to "Some*thing* is water here", and it's not clear that (**3**) involves any such thing.

More ingenuity is required to get even this close in the case of sentences like

6) Oil is lighter than water.

This sentence is true. But if we formalize as

7) $F\alpha\beta$

with "F" corresponding to "lighter than", and "oil" and "water" to "α" and "β" respectively, we say something which is true iff the weight of the total amount of oil in the universe is less than that of the total amount of water. This may still be true, but it is clearly not what (**6**) says. Rather, (**6**) says something like: any volume of oil weighs less than the same volume of water, and this would have to be the basis for an appropriate formalization, using the notion of "part of" as in (**5**).

In the case of words that are in many respects similar, for example "gold", a similar strategy would be called for in formalizing such sentences as

8) This ring is made of gold.

Made of gold it may be, but not composed of the scattered totality of all the gold in the world.

Ex. 4.13 Provide formalizations of (**7.6**) and of (**7.8**), showing your correspondence scheme. Are your formalizations true upon an intended interpretation?

Does "Pegasus" stand for an object? If it does, the object is a mythological one. Are there mythological objects, or is applying "mythological" a way of saying that there is no such object? Opinions differ. If, like me, you think that there are not *really* any such objects as Pegasus, you cannot formalize names like "Pegasus" by name-letters. One alternative is given in §**10**.

Even if you think that there is such a thing as Pegasus, you will not, I presume, think there is any such thing as Vulcan. "Vulcan" was introduced in the nineteenth century by astronomers. They posited an additional planet, lying between the planet Mercury and the Sun, whose presence they felt was required to explain the course of Mercury. They were wrong: there is no such planet between Mercury and the Sun. So you cannot think that "Vulcan" stands for an object, so you cannot formalize it by a name-letter. An alternative is given in §**10**.

8 Identity

"=" is a genuine predicate (or, in some terminologies, a predicate constant). Unlike predicate-letters, which are assigned different sets in different interpretations, "=" is assigned a set specifiable in the same way in every interpretation: the set of all ordered pairs whose members belong to the domain and whose first member is the same as the second. This stipulation about interpretations reflects the intention that "=" should mean identity.

To conform to tradition, we will abbreviate expressions like "$= \alpha\beta$" and "$\neg = \alpha\beta$" to "$\alpha = \beta$" and "$\alpha \neq \beta$".

It follows from what was said in the first paragraph that

1) $\alpha = \alpha$

is true on every interpretation, that is, $\models_Q \alpha = \alpha$. This is enough to establish that

2) $\models_Q \forall x\, x = x$.

By contrast,

3) $\models_Q \forall x Fxx$

is false. For though **(3)** is true upon some interpretations (for example, those which assign to "F" what is assigned to "="), it is false on others (for example, those which assign to "F" the set of ordered pairs such that the first is larger than the second).

The following true argument-claim illustrates the so-called principle of *substitutivity of identicals*:

4) $F\alpha, \alpha = \beta \models_Q F\beta$.

This reflects the intuitive truth that if α is the same thing as β, then any predication true of α is also true of β. The point can be put more generally as follows:

5) An interpretation upon which "$\alpha = \beta$" is true is one upon which: "... α ..." is true iff "... β ..." is true.

The last line is to be read as follows: any sentence containing "α" is true (upon the interpretation) iff the corresponding sentence, but with "α" replaced by "β", on any number of occurrences, is true (upon the interpretation). A rough summing up: identical objects have identical properties.

Ex. 4.14 (a) Formalize the following, using "=" where appropriate:

(i) Mr Hyde is the same person as Dr Jekyll.
(ii) Everything is what it is and not another thing.
(iii) John is wiser than anyone else.
(iv) Apart from John no one brought a present.
(v) Mary loves only John.
(vi) If Mary loves anyone John is that person.
(vii) Mary loves John but John loves someone else.
(viii) Mary loves John and Sally loves someone else.
(ix) John loves himself and so does Mary.
(x) Who laughs last laughs longest.

(b) Show that in English there are counterexamples to the claim that if $x = y$, then "...x..." is true iff "...y..." is true. Do these discredit **(8.5)**?

9 Numeral adjectives

Consider

1) There are two men at the door.

Though "two" is, by some standards, an adjective attached to "man", it does not modify "man" in the predicative way. We cannot formalize **(1)** as

2) $\exists x(Fx \,\&\, Gx)$

with "F" corresponding to "is a man at the door" and "G" corresponding to "two". An intended interpretation would have to assign to "G" the class of all things which are two, and this is nonsense: no (one) object is two.

Ex. 4.15 Discuss the following objection:

One book may be a hundred pages; so one object can be a hundred.

We can provide adequate formalizations of sentences like **(1)** by using quantifiers and "=". **(1)** is probably ambiguous between

3) There are exactly two men at the door

and

4) There are at least two men at the door.

We can formalize **(4)** as

5) $\exists x \exists y(Fx \,\&\, Fy \,\&\, x = y)$.

This is true upon an interpretation, i, iff at least two things are in $i(F)$. Applying the interpretation rules of (**2.2**) in some detail, **(5)** is true upon an interpretation, i, iff "$\exists y(F\alpha \,\&\, Fy \,\&\, \alpha \neq y)$" is true on some α-variant of i, say i'. This is turn is true on i' iff "$F\alpha \,\&\, F\beta \,\&\, \alpha \neq \beta$" is true on some β-variant of i', say i''. Any such i'' must assign different objects to "α" and "β", hence it must assign to "F" a set having at least two members. But i' and i agree with i'' in their assignment to "F". So **(5)** is true on an interpretation iff it assigns at least two things to "F". Hence **(5)** is an adequate formalization of **(4)**.

We saw (in §4) that "$\exists xFx$" adequately formalizes "At least one thing is F". It is easy to modify **(5)** so that it formalizes "At least three things are F". More generally, the following schema can be seen as giving instructions for formalizing "At least n things are F", for an arbitrary numeral n:

6) $\exists x_1 \ldots \exists x_n(Fx_1 \,\&\ldots\&\, Fx_n \,\&\, x_1 \neq x_2 \,\&\ldots\&\, x_1 \neq x_n \,\&\, x_2 \neq x_3 \,\&\ldots\&\, x_2 \neq x_n \,\&\ldots\ \ldots\&\, x_{n-1} \neq x_n)$.

To apply this recipe, we imagine the variables to be ordered in some way, with "x_1" corresponding to the first variable in the ordering, and so on. The recipe enjoins you to write n instances of the existential quantifier followed by a variable (a different variable each time), followed by n instances of "F" followed by each of the variables, followed by instances of $v \neq v'$ for every pair of distinct variables. Applying it to the case $n = 4$, so that the aim is to formalize "At least four things are F", we get:

7) $\exists x \exists y \exists z \exists x' \ (Fx \,\&\, Fy \,\&\, Fz \,\&\, Fx' \,\&\, x \neq y \,\&\, x \neq z \,\&\, x \neq x' \,\&\, y \neq z \,\&\, y \neq x' \,\&\, z \neq x')$.

Exactly n things are F is true iff (at least n things are F and at most n things are F). So if we can formalize "at most n", it will be easy to formalize "exactly n".

Suppose that

8) At most two things are F

is true. Then, if ever we find objects, o_1, o_2, o_3, which are all F, we know that we must have counted one object twice, which means that either $o_1 = o_2$ or $o_1 = o_3$ or $o_2 = o_3$. This justifies the formalization of **(8)** as

9) $\forall x \ \forall y \ \forall z((Fx \,\&\, Fy \,\&\, Fz) \rightarrow (x = y \vee x = z \vee y = z))$.

Similarly

10) At most one thing is F

is formalized

11) $\forall x \forall y(Fx \,\&\, Fy \rightarrow x = y)$.

In general, at most n things are F is formalized by the recipe:

12) $\forall x_1 \ldots \forall x_{n+1}((Fx_1 \ldots \& \ldots Fx_{n+1}) \rightarrow (x_1 + x_2 \vee \ldots \vee x_1 = x_{n+1} \vee x_2 = x_3 \vee \ldots \vee x_2 = x_{n+1} \vee \ldots\ \ldots \vee x_n = x_{n+1}))$.

To formalize **(3)** – "there are exactly two men at the door" – we can simply conjoin **(5)** and **(9)**. The result is equivalent to the neater:

13) $\exists x\ \exists y(Fx\ \&\ Fy\ \&\ x \neq y\ \&\ \forall z(Fz \rightarrow (z = x \vee z = y)))$.

This says that some x and y are F and are distinct, and that any F thing is one of these. It adequately captures the idea that exactly two things are F. The general recipe on the lines of (13) is

14) $\exists x_1 \ldots \exists x_n(Fx_1\ \&\ \ldots\ \&\ Fx_n\ \&\ x_1 \neq x_2\ \&\ \ldots\ \&\ x_1 \neq x_n\ \&\ x_2 \neq x_3\ \&\ \ldots\ \&\ x_2 \neq x_n\ \&\ \ldots\ \ldots\ \&\ x_{n-1} \neq x_n\ \&\ \forall x_{n+1}(Fx_{n+1} \rightarrow (x_{n+1} = x_1 \vee \ldots \vee x_{n-1} = x_n)))$.

Faced with the task of formalizing an English sentence containing numeral adjectives, the recipes in (6) and (14) may be useful. More insight into the ideas involved is provided by appreciating the possibility of a recursive approach. For example, the notion of "exactly n" (for some positive numeral n) can be abbreviated by a quantifier of the form "$\exists^n x$", and captured by a basis clause (15) and a recursive clause (16):

15) $\exists^1 xFx =_{df} \exists x(Fx\ \&\ \forall y(Fy \rightarrow x = y))$.
16) $\exists^n xFx =_{df} \exists x(Fx\ \&\ \exists^{n-1} y(Fy\ \&\ x \neq y))$.

(15) defines "exactly one thing is F" as "something is F and everything which is F is that thing". (This is how Russell defined "The F exists".) (16) defines "exactly n things are F" as "something is F and there are exactly $n - 1$ other things which are F".

Ex. 4.16 (a) Following the pattern of (9.15) and (9.16), provide general definitions (for arbitrary n) of the "at least" and "at most" quantifiers.

(b) Explain what, if anything, would be wrong with the following in place of (9.16):

$\exists^n Fx =_{df} \exists^1 x(Fx\ \&\ \exists^{n-1} y(Fy\ \&\ x \neq y))$.

(c) Can "There are finitely many foxes" be adequately Q-formalized as

$\exists x(Gx\ \&\ \exists^x y(Fy))$

with "G" corresponding to "is a natural number" and "F" to "is a fox"?

We can use these ideas to capture the plurality involved in sentences in which an existential quantifier attaches to a plural noun. Thus (4.3)

Some elephants are greedy

may mean something like

17) More than one elephant is greedy.

If so, it can be formalized as

18) $\exists x\ \exists y(Fx\ \&\ Fy\ \&\ Gx\ \&\ Gy\ \&\ x \neq y)$

with the intended interpretation assigning to "F" the set of elephants, and to "G" the set of greedy things. If (17) seems more precise than (4.3), it may be that we need to turn to implicature for an explanation. If, looking at a large bag of nails, I exclaim

19) Some of the nails have gone rusty

my audience may well expect there to be more than two rusty ones, but it seems wrong to say that this is actually entailed, rather than implicated, by (19).

The recipes (7), (12) and (14) can be used to formalize sentences in which number words like "one", "two" and "three" qualify nouns or noun phrases (like "men" or "men at the door"). In virtue of this role, number words in these occurrences have been called "numeral *adjectives*". A contrasting role is when the same words are used, apparently, as names of numbers, as in the series "one", "two", "three", or in a sentence like "one plus two is three". In this role, the words are sometimes called, simply, numerals; and in this role, one cannot rely on the recipes to provide formalizations. A sentence like

20) Five plus seven equals twelve

has to be formalized using name-letters to correspond to the numerals.

Ex. 4.17 (a) Formalize the following sentences, showing your correspondence scheme. (You may if you wish use the "exactly" quantifier defined by (9.15) and (9.16), and the quantifiers defined in your answer to Ex. 4.16a.)

(i) Mary kissed two people at once.
(ii) Three kings came to Bethlehem.
(iii) If two persons are present, there is a quorum. There are three of us present. So there is a quorum.
(iv) John is happy and so is Mary. So at least two people are happy.
(v) Bananas and apples are fruits. You have eaten two bananas and three apples. So you have eaten five fruits.
(vi) 2 + 3 = 5.

(b) Is the following consistent?

I don't have two racquets, I have three.

What does your answer show about the correct interpretation of numeral adjectives in English? (Cf. K. Bach [1987], p. 78.)

(c) Formalize the following, noting any problems.

Shem kicked Shaum twice. It happened in the gymnasium.

(d) Does the displayed sentence in (c) above pose any problem for the alleged evidence, presented in Ex. 4.12b, in favour of the event analysis of action sentences? In your discussion, consider also the fact that the following seems ungrammatical, yet arguably should be acceptable on the event analysis:

Three times the red flag was unfurled. *They* happened in the main square.

10 Descriptions

We saw in §7 that there is a case for saying that definite descriptions like "the present Prime Minister of Great Britain" and "the moon" should not be formalized by name-letters. This section explores that case in more detail. First, I present part of the only serious alternative pattern of **Q**-formalization, provided by Bertrand Russell's theory of descriptions.

The theory comes in two parts, one of which deals with sentences like:

1) The moon is cold.

Russell's idea was that such a sentence should be **Q**-formalized as if it were "There is exactly one moon and it is cold":

2) $\exists x(Fx \;\&\; \forall y(Fy \rightarrow x = y) \;\&\; Gx)$.

The second part deals with sentences like

3) The moon exists

which Russell treats as an instance of (**9.14**) for $n = 1$:

4) $\exists x(Fx \;\&\; \forall y(Fy \rightarrow x = y))$.

In both cases, an intended interpretation should assign to "F" the set of moons, and to "G" the set of cold things. (**2**) in effect says that exactly one thing is F and that thing is G. If it correctly formalizes (**1**), then (**1**) is false if there is more than one moon or less than one or if there is exactly one moon which is not cold; and otherwise (**1**) is true. On an intended interpretation, (**4**) says that there is exactly one moon.

We need some relativization, or else (**2**) and (**4**) will be false on an intended interpretation whereas (**1**) and (**3**) are naturally understood as true. Many planets other than the earth have moons. Hence there is more than one moon. Hence (**2**) and (**4**) are false on the intended interpretation. There are various options:

(a) Regard (**1**) and (**4**) as elliptical for something like "The earth's moon is cold", "the earth's moon exists". This may seem acceptable in the present case, but a speaker may utter something similar without having any idea of how to complete it in that kind of way. For example, one who utters "The door is closed" in a room with many doors, may not have troubled to think of any restriction on "door" which only one door satisfies.

(b) Restrict the domain of intended interpretations, or restrict what an intended interpretation assigns to "F". The restriction might be to Earth moons or contextually salient moons. It may be hard to give any general rule whose application would do justice to the meaning of English sentences. Such issues were relatively unimportant to Russell himself, whose initial interest in definite descriptions was in connection with mathematics, in which such contextual sensitivity does not typically arise.

The only serious alternative **Q**-formalization of (**1**) and (**3**) uses a name-letter for "the moon". One advantage of formalizing (**1**) by (**2**), in accordance with Russell's theory of descriptions, is that this captures the validity of some arguments that the name-letter approach cannot capture, for example

5) The moon is cold. Therefore a moon is cold

and

6) The author of *Mein Kampf* is a maniac. Hitler wrote *Mein Kampf*. Therefore Hitler is a maniac.

The name-letter method yields the following:

7) $G\alpha$; $\exists x(Fx \;\&\; Gx)$

(the correspondences being "G" for "is cold", "α" for "the moon", "F" for "is a moon") and

8) $F\alpha$, $G\beta$; $F\beta$

(the correspondences being "F" for "is a maniac", "α" for "the author of *Mein Kampf*", "G" for "wrote *Mein Kampf*" and "β" for "Hitler"). Both of these are patently invalid, whereas (**4**) and (**5**) are valid.

On Russell's theory, by contrast, we get the following **Q**-valid formalizations:

9) $\exists x(Fx \;\&\; \forall y(Fx \rightarrow x = y) \;\&\; Gx)$; $\exists x(Fx \;\&\; Gx)$

(the correspondences being "F" with "is a moon" and "G" with "is cold") and

10) $\exists x(Fx \;\&\; \forall y(Fx \rightarrow x = y) \;\&\; Gx)$, $F\alpha$; $G\alpha$

(the correspondences being "F" with "wrote *Mein Kampf*", "G" with "is a maniac" and "α" with "Hitler"). The **Q**-validity of (**9**) is evident. (**10**) is also **Q**-valid: an interpretation upon which the premises are true will assign a set containing just one object to "F" (on the intended interpretation, "F" will be assigned the set of writers of *Mein Kampf*, that is, just Hitler), and this object will both be assigned to "α" and belong to what is assigned to "G". This ensures that any interpretation upon which the premises are true is one upon which the conclusion is true also.

The fact that Russell's theory of descriptions makes it possible to reflect the validity of some arguments whose validity cannot be captured by the method of name-letters would decisively establish its superiority only if the arguments in question were *formally* valid. Here, however, we find a circle: they are formally valid iff "the" is a logical constant, and, by the present standards, "the" is a logical constant iff Russell's theory of descriptions is correct. Then, and only then, is it definable in terms of the stipulated constants "all", "some", "if" and "is the same as".

Ex. 4.18 Russell gave an argument in favour of his theory of descriptions on the following lines:

> "The King of France is bald" is obviously false, since France is a Republic.
> But then it would seem that "The King of France is not bald" ought to be true.
> My theory can explain how this is possible.

The explanation involved discerning a scope ambiguity in "The King of France is not bald". Use distinct **Q**-formalizations of this sentence (stating the intended correspondences) to bring out the ambiguity. Cf. Russell [1905], p. 53.

If we tried to use name-letters to formalize sentences like (**3**) ("The moon exists") we would represent all of them as true, even

11) The golden mountain exists.

For all would be formalized as "$F\alpha$", which entails

12) $\exists x\; x = \alpha$.

This connects with an argument for Russell's theory of descriptions mentioned in §7: (i) some definite descriptions – we will call them "empty" ones – stand for no object, for example, "the golden mountain"; and (ii) all definite descriptions should be treated in the same way. Since the empty descriptions cannot be adequately formalized by means of a name-letter, (ii) tells us that none should be. Since there are only these two candidates for methods of formalizing, Russell's triumphs.

To reiterate why an empty description cannot be adequately formalized by a name-letter: our standard of adequacy for a formalization is that its truth-upon-an-intended-interpretation conditions should match the truth conditions of the original English. But what could be an intended interpretation when an empty description is formalized by a name-letter? We have to choose some object, since every interpretation assigns objects to all name-letters. Perhaps we should designate an arbitrary object, say the number 0 or the null set, as what an intended interpretation should assign to a name-letter which corresponds to an empty description (cf. Frege [1892b], p. 70). But it seems that whatever object we choose we will get the wrong result. Thus

13) Someone was the unique author of *Principia Mathematica*

is false (it was jointly authored by Russell and A. N. Whitehead), whereas

14) No one was the unique author of *Principia Mathematica*

is true. Formalizing by the method of name-letters would yield, respectively, **(12)** and its negation:

15) $\neg\exists x\ x = \alpha.$

Whatever object we choose as what an intended interpretation assigns to "α", **(12)** will be true upon the interpretation and **(15)** will be false. This gets the truth values just the wrong way about, so there is no adequate formalization to be had by this method.

According to Russell's theory of descriptions, sentences which would naively be classified as identity sentences are formalized not as identity sentences but as existential quantifications. For example, we might naively classify

16) Scott was the author of *Waverley*

as an identity sentence. However, an appropriate **Q**-formalization, using Russell's theory of descriptions, is:

17) $\exists x(Fx\ \&\ \forall y(Fy \rightarrow x = y)\ \&\ x = \alpha)$

where "F" corresponds to "wrote *Waverley*" and "α" to "Scott". This **Q**-sentence, being dominated by an existential quantifier, must be classified as an existential quantification and not as an identity sentence. In particular, something of the form of **(17)** does not yield a premise fit for an application of **(8.5)**, the substitutivity of identicals.[2] Russell [1905] thought he could establish this by the fact that whereas George IV wished to know whether Scott was the author of *Waverley*, he did not wish to know whether Scott was Scott.

Ex. 4.19 George IV perhaps wished to know whether Hesperus was Phosphorus but not whether Hesperus was Hesperus. Evaluate the impact this has upon Russell's argument for the conclusion that **(10.16)** is not an identity sentence.

Russell's view also represents some initially plausible principles as false, for example:

18) The F which is G is F;
The F which is G is G;

19) "The F is not G" is the negation of "The F is G".

The **Q**-formula corresponding **(18)** is not **Q**-valid, and there are two **Q**-formulae corresponding to "The F is not G", only one of which is the negation of a **Q**-formula corresponding to "The F is G" (see Ex. 4.18). Russell exploited the attractiveness of these principles in an argument against a position he attributed to Meinong, according to which there are no really empty definite descriptions, merely some which refer to non-existent things. If this view is combined with **(18)** and **(19)** it leads to contradiction, as we can see by putting "square" for "F" and "not square" for "G".

Ex. 4.20 Show how to derive a contradiction from **(10.18)** and **(10.19)**. How should someone who accepts the view Russell attributed (possibly incorrectly) to Meinong respond?

[2] In *Principia Mathematica*, Russell proved that in restricted contexts you could treat the formal equivalents of English expressions of the form "The F is the G" (for example, "the most powerful man in the world is the President of the World Bank") as if they were identity sentences: see Russell and Whitehead [1910], *14.272.

The principle that all definite descriptions should be treated alike is open to question. If we apply the same sort of principle to names, saying that all names must be treated alike, we find that we should treat "Vulcan" and "Reagan" alike. Since, for reasons already seen in §7, we cannot formalize "Vulcan" by a name-letter we also should not formalize "Reagan" by a name-letter. Russell himself accepted this consequence. Indeed, he held that the only expressions that could be adequately formalized by name-letters were expressions that we do not ordinarily classify as names (but rather as demonstrative pronouns): "this" and "that", as these are used to refer to momentary subjective experiences (these are his so-called "logically proper" names: Russell [1918], pp. 200–1). We will return in §**12** to the question of names and descriptions.

Ex. 4.21 (a) Use the "exactly one" quantifier of (**9.15**) to formalize "The moon is cold".

(b) Using Russell's theory of descriptions where appropriate, formalize the following, giving alternative formalizations in any cases of ambiguity:
(i) The cat sat on the mat.
(ii) The cat did not sit on the mat.
(iii) The cat did not sit on the mat, but on the table.
(iv) The average family has 2.4 children.
(v) Someone was the unique author of *Principia Mathematica*.
(vi) Sally ate some of the cakes.

Russell's own first response to the problem of formalizing descriptions was to introduce a new device, going beyond the syntax of **Q**, the (inverted) iota-operator, "ɿ". This is a variable binding operator, rather like a quantifier, but it forms what at first look like singular terms. Using it, one could write (**1**) as

20) is cold (ɿ*x*)(moon *x*).

This could be expressed in something more like English as "(ɿ*x*)(such that *x* is a moon) is cold" or "the *x* such that *x* is a moon is cold". Russell found that in order to indicate the scope of these apparent singular terms he needed further complications (see *Principia Mathematica*, Russell and Whitehead [1910], ★14.03). He claimed that the iota-expressions were strictly unnecessary: wherever they occur, one can use the methods exemplified by (**2**) and (**4**) to replace them by ordinary Q-formulae. Russell called the iota-expressions, like "(ɿ*x*)(moon *x*)", "incomplete symbols", because they do not need to be "completed" by an object in the world in order to function significantly in language.

11 Existence

(**10.3**) and (**10.13**) raise the question of how we should formalize sentences which affirm or deny existence. The latter could have been rephrased:

1) The unique author of *Principia Mathematica* does not exist.

This is not very idiomatic as compared with (**10.13**), but it seems to mean the same. We are perhaps more familiar with this phrasing in mathematical contexts, for example:

2) The greatest prime number does not exist.

There are also such truths as

3) Vulcan does not exist,

in which a name is used rather than a description.

"∃" is the "existential" quantifier, so one might expect it to have a role to play in formalizing assertions and denials of existence. On the other hand, "exists" is grammatically a predicate, so one might think to formalize it by means of a predicate-letter.

An example of a use of "exists" which cries out for treatment by "∃" is

4) Mad dogs exist.

Though "exists" is grammatically a predicate, its function here is not to predicate existence of each and every mad dog. Rather, its role should be compared with "rare" in

5) Mad dogs are rare.

"Rare" does not predicate rareness of each mad dog and is not appropriately formalized by a predicate-letter. The interpretation of a predicate letter (of degree 1) will assign it a set of objects, of each one of which the predicate-letter is, on the interpretation, true. "Rare" is not an expression which intelligibly applies to a single object; and it is tempting to think that the same goes for "exists". There is certainly a difficulty in understanding "exists" as like an ordinary predicate in denials of existence, for example:

6) Mad dogs do not exist.

It is clear that there is no object in the universe of which it is being said that *that* object does not exist.

Ex. 4.22 What, if anything, would be wrong with the view that an intended interpretation should assign its domain to a predicate-letter used to correspond to "exists"?

Suppose we try to use a predicate-letter to correspond to "exists" in these cases. We might try formalizing (**4**) by

7) $\forall x((Fx \;\&\; Gx) \rightarrow Hx)$

and (**6**) by

8) $\forall x((Fx \;\&\; Gx) \rightarrow \neg Hx)$

with "*F*" corresponding to "is a dog" "*G*" to "is mad" and "*H*" to "exists". If there are no mad dogs, both (**7**) and (**8**) come out true, upon an intended interpretation, whereas (**4**), which (**7**) is supposed to formalize, should certainly be false in this case. If we try instead

9) $\exists x(Fx \;\&\; Gx \;\&\; Hx)$

for (**4**), and

10) $\exists x(Fx \;\&\; Gx \;\&\; \neg Hx)$

for (**6**) (with correspondences as before) we find that both require for their truth upon an intended interpretation that the domain should contain mad dogs which, on the face of it, is just what (**6**) denies.

One attempt to meet this difficulty involves distinguishing between *being* and *existence*. The category of being is the wider, embracing plenty of non-existent things, like Pegasus, the golden mountain, round squares, as well as the existing things like Ronald Reagan and Italy. The existential quantifier, expressed in English by "there is", relates to the category of being, "exists" to the narrower category of existence. Thus

11) There are things which do not exist

expresses a truth, on this theory. It could be formalized, with correspondences as before, as

12) $\exists x \neg Hx$.

Let us call this "Meinong's Theory of Existence" after a famous proponent of a not too dissimilar view. Russell accused the view of showing a "lack of that robust feeling for reality which must be preserved in logical studies" ([1919], p. 170). One may not feel that this observation is enough to refute the view. However, Russell did show that the logical considerations under discussion in this section do not on their own force one to adopt it.

In the case of (4) and (**6**), non-Meinongian formalizations are attained by using the existential quantifier to correspond to "exists", rather than a predicate-letter:

13) $\exists x(Fx \;\&\; Gx)$

14) $\neg\exists x(Fx \;\&\; Gx)$

with "*F*" corresponding to "is a dog" and "*G*" to "is mad". So some sentences containing "exists" can be adequately formalized using "$\exists$" rather than a predicate-letter.

The cases so far discussed have had plural subjects (for example, "mad dogs"). I now turn to singular existential sentences, for example (3):

Vulcan does not exist.

This is a negative sentence, and we should also have an example of a positive one:

15) The third man exists.

Imagine (**15**) to be uttered in connection with some intrigue. There is a debate about how many people were involved. You believe there were three: two well-known villains and a third whose name you do not know. You might utter (**15**) to state your position on the controversy.

We cannot in any straightforward way apply the policy adopted in the plural case to the singular case. In (**4**),

Mad dogs exist,

the plural subject contained predicates for the quantifier to attach to. (**3**) does not on the face of it contain such a predicate. If we try to formalize it by

16) $\neg\exists xFx$,

we need to ask what "F" corresponds to. It seems it must correspond to "Vulcan", but it cannot do that since it has to correspond to a predicate and "Vulcan" is a name. We will return to this question shortly.

In the case of (**15**), we would have exactly the same problem, if we were to think of "the third man" as formalizable by a name-letter. The fact that there are predicates contained in the phrase should encourage the view that they could be made accessible to existential quantification. The theory of descriptions given in §**10** has the merit of taking seriously the fact that definite descriptions contain predicates. If we mechanically apply the theory to (**15**), the result is:

17) $\exists x(Fx \ \& \ \forall y(Fy \rightarrow x = y) \ \& \ Gx)$

where "F" corresponds to "is a third man" and "G" to "exists". However, if we agree with Russell, we will hold that everything exists. In that case, the intended interpretation of "G" will be the domain of interpretation. This means that "& Gx" adds nothing to the truth-upon-an-interpretation conditions of (**17**). So we might as well formalize (**15**) as

18) $\exists x(Fx \ \& \ \forall y(Fy \rightarrow x = y))$

(with "F" as before). The remaining part of Russell's theory of descriptions proposed precisely this. Your opponent's denial of (**15**) will be formalized by prefacing (**18**) with "$\neg$". These formalizations seem adequate, yet avoid the view that there are things that do not exist.

We saw in §**7** that "Vulcan" cannot be adequately formalized by a name-letter. This blocks the formalization of (**3**) as

19) $\neg F\alpha$

with "F" corresponding to "exists" and "α" to "Vulcan". The previous paragraph, however, suggests how one might find existential quantifier formalizations of "exists" as it occurs in (**3**): discern a hidden predicate beneath the occurrence of the name "Vulcan". Two versions of this approach are available.

20) $\exists x \, x = \alpha$

asserts the existence of whatever the interpretation assigns to "α". It does so by saying that something exists which is identical to α, which is just a way of saying that α exists. (This corresponds to the English equivalence between "Napoleon exists" and "There is something which is (identical to) Napoleon".) The name-letter "α" is converted into a predicate "$= \alpha$" – an expression which when coupled with a name or name-letter forms a sentence. This is one way to find a predicate hidden in the occurrence of a name, and it can be applied generally (cf. Quine [1948], Hochberg [1957]). Applying it to (**3**), we could see it as saying that it is not the case that something is identical to Vulcan, and formalize it as:

21) $\neg\exists x \, x = \alpha$

with "α" corresponding to "Vulcan".

The trouble with this suggestion is that (**21**) is not true upon any interpretation, whereas (**3**) is true, so the formalization is inadequate.

(**20**) is **Q**-valid, that is, true upon every interpretation. By (**2.2**), (**20**) is true upon an interpretation i iff "$\beta = \alpha$" is true upon some β-variant interpretation. There is a β-variant of any i, an interpretation which assigns the same object to both "α" and "β". We rely upon the stipulation that every domain contains at least one object; if the empty domain were allowed, there would be no interpretation with respect to it which assigned one and the same object to the two name-letters. (For a relaxation of these stipulations, see §**20**.)

One way to find a predicate hidden in a name, as just described, sees the name, n, as concealing a predicate "$= n$". We have seen that this will not always yield the desired results. The other proposal is more radical. It is that at least some names are "really" definite descriptions, in an abbreviated or truncated form. Perhaps "Vulcan" abbreviates the definite description "The intra-mercurial planet". Then (**3**) is equivalent to

22) The intra-mercurial planet does not exist

and we have already seen how (for (**15**)) one could give an adequate **Q**-formalization of a sentence like this (as (**18**)). If we hold to the view that everything classified as a name in English should be treated alike, and take it that the truth of (**3**) gives us a compelling reason to regard "Vulcan" as really a definite description, then we are committed, as Russell was, to the view that *all* English names are really "abbreviated" or "truncated" descriptions (Russell [1918], pp. 200, 243).

To summarize: in formalizing plural affirmations and denials of existence, like "Mad dogs exist", the quantifier approach gives the right result and the predicate-letter approach the wrong result. Singular assertions and denials of existence superficially divide into two categories: (i) those involving a name, like "Vulcan exists" and (ii) those involving a description, like "The third man exists". Assuming that we adopt Russell's general recipe for formalizing descriptions, the quantifier does all that is needed in regard to (ii). Use of a predicate-letter corresponding to "exists" would be otiose, unless one accepts Meinong's theory. Existential sentences in (i) can only be adequately **Q**-formalized if they are "really" in category (ii): that is, if the names in question are treated as "really" descriptions. Whether this seeming distortion of English can be justified is discussed in the next section, which, as it departs from the question of **Q**-formalization, can be omitted without loss of continuity.

Our policy will be to formalize English names by name-letters, unless, as in (**3**), this leads to an inadequate formalization.

Ex. 4.23 (a) Formalize the following, noting any difficulties:

Not all the characters in fiction exist.

(b) Assess the following argument:

"As regards the actual things there are in the world, there is nothing at all you can say about them that in any way corresponds to this notion of existence. . . . if there were such a thing as this existence of individuals that we talk of, it would be absolutely impossible for it not to apply, and that is the characteristic of a mistake." (Russell [1918], p. 241.)

(c) Assess the following argument:

"Exists" is not a logical predicate; that is to say, there is no corresponding predicate in first-order logic.

12 Are names "really" descriptions?

One of Russell's motivations for holding that English names are definite descriptions sprang from his engagement in a project very similar to that of finding **Q**-formalizations of all English sentences. He was using a richer language than **Q**, the language of *Principia Mathematica*, but its richness has no bearing on the issue we are now discussing, so I shall bracket that difference. Russell believed that some interpretation of **Q** would express anything that could be said or thought in *any* language, and **Q** would have the advantages of clarity, and accessibility to logical manipulations. This belief entails that at least some apparent names (e.g. "Vulcan") are descriptions, and yields a more general conclusion if we assume that all names should be treated alike. This was not Russell's only reason: names, especially in the context of existential sentences, raise problems quite independently of the project of **Q**-formalization. This section briefly introduces some of those problems.

(i) Consider again (**11.3**):

Vulcan does not exist.

This is true (or so we all ordinarily believe). But how does the sentence work? It appears to introduce an entity, viz. Vulcan, and then go on to say of this entity that it does not exist. Unless we are Meinongians, a sentence which does that would be contradictory, and not true. The description theory dissolves this mystery. "Vulcan" does not introduce an entity, but rather a claim to the effect that exactly one thing has a certain property (being a planet between Mercury and the sun), and the rest of the sentence serves to negate that claim.

(ii) Consider how names are learned. It is as often as not by means of a definite description. "Who was Gödel?", you enquire. When I tell you that he was the Austro-Hungarian logician best known for his proof of the incompleteness of arithmetic, you come to be in a position to use the name. A natural hypothesis is that the mechanism here is *definition*: the description defines the name, and thus gives you its meaning.

(iii) Consider those names which have bearers (like "Reagan" and unlike "Vulcan"). Understanding such a name involves knowing who or what its bearer is. How is this knowledge represented in your mind? A natural answer is that it is represented by a definite description: if someone uses a name, *N*, he must be able to answer the question: "who or what do you mean by *N*?", and it seems the only appropriate answer he could make would consist either in pointing to the bearer of "*N*" or in citing a description true of it.

(iv) Consider an identity statement like

1) Lewis Carroll is Charles Dodgson.

This is no trivial truth, but a discovery that became known to a wider and wider circle in the late nineteenth century. Suppose that, instead of the view that names are really descriptions, we hold the view that a name simply stands for an object. Then it seems hard to explain how you could understand both of the names "Lewis Carroll" and "Charles Dodgson" without knowing that (**1**) is true. In understanding each, you must know who or what the name stands for, so how could you fail to know that they stand for the same object? Yet clearly one can understand (**1**) without knowing whether it is true, and this is incontrovertibly allowed for by the description theory: you may associate the names with different descriptions.

I now turn to some criticisms of the description theory, and comment on the above four arguments for it.

(a) It would seem that two people could both understand a name perfectly well, yet associate different descriptions with it. If you are a White House janitor you may associate quite different descriptions with "Reagan" from those associated with the name by an El Salvadorian, perhaps living beyond the reach of television. So do we have to say that the name is ambiguous? This would seem implausible.

Russell was quite well aware of this issue, and explained how it was no objection to his theory (however objectionable it may be to theories sometimes attributed to him, for example by Kripke [1972], p. 27). He allowed that two people could communicate satisfactorily using a name, even though they associated different descriptions with it, provided that the different descriptions were true of the same thing. The definite description which a name "really" is should not be thought of as something that has to be common to speaker and hearer when they communicate. Rather, the relevant definite description varies with speaker and occasion: it is whatever description would make explicit the thought in the speaker or thinker's mind on the occasion in question (Russell [1912], p. 29). There is no one description steadily and universally associated with the name. Hence in one good sense of "meaning", that in which meaning is what is common to speaker and hearer in communication, Russell's theory is not intended as an account of the meaning of names.

Ex. 4.24 (a) Say whether you think Russell's account of the relation between names and descriptions makes it vulnerable to the following line of argument:

> Suppose the description associated with "Reagan" is "The President of the United States (in 1987)". Then the proposition expressed by "Reagan is the President of the United States" is the same as that expressed by "The President of the United States is the President of the United States". But the second of these is trivial and the first is not, so they cannot express the same proposition.

(b) Comment on the following argument, preferably in the light of reading Russell [1912], pp. 29–31:

> The claim that names are descriptions is ambiguous between an ∀∃∀ claim (that for every name, there is a description such that whenever the name is used it is equivalent to the description) and an ∀∀∃ claim (that for every name, every occasion of its use, there is a description to which it is equivalent). The former is wildly implausible; Russell's is the latter.

This very fact, however, may make the logician regard the theory as unsuitable for his purposes. If someone asserts something, the logician will want to know the consequences of what is asserted. These consequences need to derive from something publicly shareable, to derive from something common to speaker and hearer in communication. If someone idiosyncratically thinks of Reagan via the description "The person who always misses the ashtray when stubbing out his cigarettes", we do not want to count as a consequence of this man's assertion that Reagan will give a press release today, that someone who misses the ashtray when stubbing out his cigarettes will give a press release today. Descriptions deriving from these subjective sources are not appropriate for the study of logic (cf. Frege [1892b], p. 59).

(b) The argument of (i) is probably the strongest. Its immediate application is only to names without bearers. Its strength lies largely in the absence, until quite recently, of a plausible non-Meinongian alternative (cf. Evans [1982], ch. 10).

Russell thought one should treat names with and without bearers in the same way on the grounds that one might not know, of some group of names, which have bearers and which do not, even though one understood all the names in the group, so that the use of a name does not make the appropriate division. For example, we can sensibly discuss whether or not Homer existed. In doing so, we use the name "Homer", presumably correctly, without knowing whether it is a name with a bearer or a name without. Logic is supposed to proceed without empirical knowledge, drawing just on the understanding of sentences. So logic ought not to discriminate between names with bearers and names without. This certainly puts the onus on the person wanting to discriminate to find a distinction between the two classes that shows up in the use of language.

(c) The argument of (ii) is not decisive, since the phenomena adduced are consistent with the view that the use of the description in learning is to get the learner to know *who or what the bearer of the name is.* This task can be achieved without giving an expression which means the same as the name (cf. Kripke [1972], pp. 53 ff.).

(d) It is not clear that, as (iii) alleges, we can always give a definite, uniquely identifying, description corresponding to every name we use

with understanding. Suppose we notice someone on our way to work each day whom we inwardly name "Fred". We can recognize him when we see him, but any attempt to describe his features will probably (as the police know only too well) yield a description which fits hundreds of people.

(e) The argument of (iv) is fallacious. There is no valid argument from

$a = b$
He knows what the name "a" stands for
He knows what the name "b" stands for

to

He knows that "a" and "b" stand for the same thing.

Imagine Dodgson's colleagues at Christ Church, who obviously understand the name "Dodgson" and know who Dodgson is. In addition, they are aware of Lewis Carroll's success, and so understand the name "Carroll", and know who Carroll is. They may have no basis for supposing that that Dodgson is Carroll.

(f) On Russell's theory of descriptions, "the" is a kind of quantifier, meaning "there is exactly one . . .". It is plain that we can understand a sentence like

2) The inventor of the jet engine died in poverty

without (by one natural standard) knowing who the inventor was or even if there was one. (Perhaps the jet was the product of team research.) In particular, if there is a unique inventor, we need have no link with him in order to understand the sentence. (**2**), on Russell's theory, is a general statement: no particular object enters into its interpretation.

We tend to think of names otherwise. Understanding

3) Frank Whittle died in poverty

does require knowing who Whittle was. We have to have met him or been told about him or seen traces of him in order to use his name.

It seems that the conditions required for using a name are different from those required for using a definite description, and this counts against Russell's theory that names are descriptions.

13 Structural ambiguity in English

Q has no structural ambiguity. One way to clarify structural ambiguities in English (see chapter **1.12**) is to provide distinct Q-formalizations, one for each distinct reading of the English. Being structurally ambiguous is treated as having more than one logical form.

It is often said that

1) Everyone loves someone

is ambiguous between a (weak) reading upon which everyone is such that there is someone he loves, and a (strong) reading according to which some lucky person is loved by everybody. With "*Fxy*" corresponding to "*x* loves *y*" the unambiguous formalizations are, respectively:

2) $\forall x \exists y Fxy$

and

3) $\exists y \forall x Fxy$.

The strong reading entails the weak:

4) $\exists y \forall x Fxy \models_Q \forall x \exists y Fxy$

but the converse does not hold.

Ex. 4.25 Specify an interpretation upon which $\forall x \exists y Fxy$ is true but $\exists y \forall x Fxy$ is false.

The inference from $\forall\exists$ to $\exists\forall$ is known as the "*quantifier shift fallacy*" and it is commonly attributed to philosophers and others. For example, there is an argument for a foundationalist view of knowledge which, denuded of some of its protective covering, runs as follows:

5) Every justification of a proposition has to end somewhere. Therefore some propositions cannot be justified, but have to be taken for granted.

Ex. 4.26 Formalize (13.5) to show that it is an instance of the quantifier shift fallacy. (The argument of Ex. 4.24b depends upon exposing an example of this fallacy.)

Some of the ambiguity of (**1.12.9**) was of this kind:

> Logic, epistemology and metaphysics are all the philosophical subjects there are. Nicholas has written a book about logic. Nicholas has written a book about epistemology. Nicholas has written a book about metaphysics. Therefore, Nicholas has written a book about every philosophical subject.

The correspondences:

"α" for "logic";
"β" for "epistemology";
"γ" for "metaphysics";
"δ" for "Nicholas";
"F" for "is a philosophical subject";
"Gxy" for "x has written y";
"Hxy" for "x is about y";
"J" for "is a book".

The premises have no structural ambiguity, and can be formalized as follows:

6) $\forall x(Fx \rightarrow (x = \alpha \vee x = \beta \vee x = \gamma))$
$\exists x(Jx \;\&\; G\delta x \;\&\; Hx\alpha)$
$\exists x(Jx \;\&\; G\delta x \;\&\; Hx\beta)$
$\exists x(Jx \;\&\; G\delta x \;\&\; Hx\gamma)$.

The weak version of the conclusion upon which the argument is valid is

7) $\forall x(Fx \rightarrow \exists y(Jy \;\&\; G\delta y \;\&\; Hyx))$.

The strong version of the conclusion upon which the argument is invalid is:

8) $\exists y(Jy \ \& \ G\delta y \ \& \ \forall x(Fx \rightarrow Hyx))$.

(7) is consistent with Nicholas writing various books, perhaps one on each of the three subjects. (8) requires him to have written a compendious book, treating all the subjects at once. In (7) the universal quantifier has wide scope relative to the existential quantifier. In (8) the scopes are reversed, the universal falling in the scope of the existential.

The use of variables in **Q** helps keep track of the application of quantifiers. In (8) it is important which quantifier is applying to the first position in "Hyx" and which to the second. This is shown by the attachment of "x" to both the universal quantifier and the second position, and "y" to both the existential quantifier and the first position.

Pronouns sometimes play a similar role in English, as in

9) Someone called today and *he* brought his wife.

They also play another role, as shorthand for the reapplication of a name, as in

10) Oscar kissed Joan and *he* made *her* cry.

Here "he" and "her" are stylistic variants of the reuse of "Oscar" and "Joan".

Ex. 4.27 (a) Why cannot the "he" in (13.9) be regarded as a stylistic variant of the reuse of "someone"?
(b) Why cannot the "himself" in the following be regarded as a stylistic variant of the reuse of "every American"?
Every American admires himself.
(Cf. McCawley [1981], p. 126.)

Sometimes it is unclear in English which role a pronoun is playing, for example "he" in

11) If Oscar kissed anyone, he will be pleased.

The formalization upon which "he" is equivalent to a reuse of "Oscar" is

12) $\forall x(F\alpha x \rightarrow G\alpha)$

with "Fxy" corresponding to "x kissed y", "G" to "is pleased" and "α" to "Oscar". The formalization upon which "he" marks the application of the quantifier is

13) $\forall x(F\alpha x \rightarrow Gx)$.

Ex. 4.28 Which of (13.12) and (13.13) suggests that Oscar kissed a male person?

The word "only" often gives rise to ambiguity in English.

14) John only eats organically grown vegetables

would normally be interpreted in a way consistent with John eating meat, the claim being that, as far as vegetables go, all the ones he eats are organically grown. This reading is formalized

15) $\forall x((F\alpha x \ \& \ Gx) \rightarrow Hx)$

with "Fxy" corresponding to "x eats y", "G" to "is a vegetable", "H" to "is organically grown", and "α" to "John".

In my view, the more correct reading of (14), the reading that would be favoured by teachers of English, entails that John eats nothing but vegetables. This reading is formalized

16) $\forall x(F\alpha x \rightarrow (Gx \ \& \ Hx))$.

This is not literally a scope difference in **Q**: it is not that (15) and (16) differ only in point of the relative scopes of some pair of operators. The phenomenon is more like the ambiguity that can arise concerning the multiple qualification of noun phrases. For example, in

17) John is a dirty window cleaner

it is unclear whether "dirty" is meant to qualify the complex "window cleaner" or just the word "window". The alternatives are brought out respectively by the very approximate formalizations:

18) $\exists x(Fx \;\&\; Gx \;\&\; H\alpha x)$

19) $\exists x(Fx \;\&\; G\alpha \;\&\; H\alpha x)$

with "F" corresponding to "is a window", "G" to "dirty", "H" to "x cleans y" and "α" to "John". The formalizations are only very approximate, for it takes more (and also perhaps less) than cleaning one or a dozen windows to be a window cleaner.

Ex. 4.29 Which of (**13.18**) and (**13.19**) requires Harry to be dirty?

Representations based on **Q** can clarify structural ambiguities in English sentences, even when those sentences resist **Q**-formalization in any reasonably revealing way. The technique involves mixing English and **Q**. For example,

20) I am trying to buy a house

is ambiguous between the claim that there is a house I have set my eye on and towards which my buying efforts are directed, and a claim which can be true even if there is no such house – even if all I have done is ask the real estate agents to send details. We could represent these claims as follows:

21) $\exists x$(x is a house & I am trying to buy x).

22) I am trying to bring it about that: $\exists x$(x is a house & I buy x).

This is, or is analogous to, a scope distinction: in (**21**), "$\exists$" has wide scope relative to "trying", in (**22**) narrow scope. We sometimes express the first reading in English by saying "I am trying to buy a particular house". All houses are particular houses, so "particular" here is not serving to qualify "house"; it is best seen as effecting a scope distinction.

In discussions of the theory of knowledge, it is often claimed that

23) If you know you can't be wrong

is ambiguous in a way which can be represented as follows:

24) necessarily $\forall x \forall y$(x knows that $y \rightarrow y$ is true).

25) $\forall x \forall y$(x knows that $y \rightarrow$ necessarily y is true).

This is also a difference of relative scope. A familiar view in epistemology is that (**24**) is true, but not terribly interesting, and (**25**), which indeed entails wholesale scepticism about the contingent, is false.

Russell gave a pleasing example of a scope distinction. He argued that

26) I thought your yacht was longer than it is

could be heard as an absurd claim, deserving the reply: "Everything is just as long as itself". With "F" corresponding to "x uniquely numbers in meters the length of y", "Gxy" for "x is greater than y", and "α" for "your yacht" (ignoring, for present purposes, our earlier resolution to treat definite descriptions, like "the yacht which you own", by Russell's theory), the absurd claim could be represented as:

27) I thought that: $\exists x(Fx\alpha \;\&\; Gxx)$.

What the speaker of (**26**) no doubt meant is something more like:

28) $\exists x(Fx\alpha \;\&\; \exists y$((I thought that $Fy\alpha$) & Gyx)).

"Gyx" is not part of the thought I attribute to myself in uttering (**26**). A more long-winded clarification is: the length I thought your yacht was is greater than the length it actually is.

In ethics, people ask whether there can be genuinely incompatible obligations. One way in which the issue might be made more precise is by asking whether it is ever possible for the following both to be true:

29) You ought to do A

30) You ought not to do A

where "A" stands for some type of action. It is unclear whether (**30**) is really the negation of (**29**). Arguably, it is ambiguous between:

31) It is not the case that you ought to do A

and

32) You ought to do not-A.

In (**31**), "not" dominates the sentence, in (**32**) it has narrow scope relative to "ought". Classical logic precludes the joint truth of (**29**) and (**31**), for they have the overall logical form of p and $\neg p$. If there are incompatible obligations in this sense, logic needs to be revised. By contrast, classical logic as such finds nothing problematic in the joint truth of (**29**) and (**32**).

Ex. 4.30 (a) Use distinct **Q**-formalizations to bring out any ambiguities in the following:

(i) Winston is always smoking a big cigar.
(ii) He always carries a large stick.
(iii) Only non-smoking males are eligible for this job.
(iv) I watched the tennis-match in bed.
(v) Only sensible dogs are taken by somebody kind for all their walks.
(vi) Everyone has a problem.

(b) Use mixed English and **Q**-formalization (in the manner of (**13.21**), (**13.22**) etc.) to bring out any ambiguities in the following:

(i) "In the whole wide beautiful world, Aldo Cassidy was the only person who knew where he was." (Le Carré, *The Naive and Sentimental Lover*, p. 8)
(ii) "Most of all I would like to thank my students, who have taught me more than they know." (E. Bach, [1974], p. vi)
(iii) Gerry means everything he says ironically.
(iv) If John is to enter a university, he must pass his examinations.
(v) I thought you were someone else.

14 Q-validity and decision

There is no method like that of truth tables for determining **Q**-validity. There are, however, systematic methods for determining, for any **Q**-valid argument, that it is **Q**-valid. The trouble is that if such a method has still not pronounced an argument valid (say after a hundred or a million steps), we do not know whether the right thing to believe is that the argument is not valid, or whether the right thing to believe is that the argument is valid but the method has not yet managed to show it.

We can get a feel for **Q**-validity, without anything so grand as a systematic method, simply by working on some examples. It is not at first obvious whether the following is **Q**-valid or not:

1) $\forall x\exists y(Fx\rightarrow Gy)$; $\exists y\forall x(Fx\rightarrow Gy)$.

A natural first reaction would be to suppose that it is invalid, by analogy with the invalid

2) $\forall x\exists yFxy$; $\exists y\forall xFxy$.

This reaction would be incorrect. An interpretation, i, upon which the premise of (**1**) is true must either assign the empty set to "F" or some non-empty set to "G". The conclusion is true upon i iff "$\forall x(Fx \rightarrow G\alpha)$" is true upon some α-variant, i'. If $i(F)$, and so $i'(F)$, are empty, then "$\forall x(Fx \rightarrow G\alpha)$" is true upon i' (for reasons spelled out in §**3**); if $i(G)$ is a non-empty set, then *some* α-variant (agreeing with i on "F" and "G") will assign to "α" a member of what it assigns to "G", so again "$\forall x(Fx \rightarrow G\alpha)$" will be true on i. So however i makes the premise true, it will make the conclusion true also.

If an argument is **Q**-invalid, we can establish this if we can find a *counterexample*: an interpretation upon which the premise(s) are true and the conclusion false. For example, an interpretation which assigns to "F" the set of ordered pairs whose first member is smaller than the second, and whose domain is the (positive) integers, is a counterexample to the **Q**-validity of (**2**).

The equivalences between universal and existential quantifiers

3) $\models_Q \exists xFx \leftrightarrow \neg\forall x\neg Fx$

4) $\models_Q \forall xFx \leftrightarrow \neg\exists x\neg Fx$

can be confirmed by reasoning that uses the same equivalences in the English which we use to describe the interpretations. Thus for (**3**) a crucial consideration is that any interpretation, i, upon which "$\exists xFx$" is true is one such that there is an interpretation, agreeing with i on "F", upon which "$F\alpha$" is true; that is, it is not the case that every interpretation agreeing with i on "F" fails to bring out "$F\alpha$" as true; that is, it is not the case that every interpretation agreeing with i on "F" brings out "$\neg F\alpha$" as true; that is, "$\forall x\neg Fx$" is false upon i; that is, "$\neg\forall x\neg Fx$" is true upon i.

Universal quantifications are true only upon interpretations which make the corresponding existential quantification true, for example

5) $\forall xFx \models_Q \exists xFx$.

However,

6) $\exists xFx \nvDash_Q \forall xFx$.

Any interpretation which assigns a non-empty set other than the domain serves as a counterexample.

The following general truth reflects the fact that we require an interpretation to assign an object to every name-letter:

7) $X \vDash_Q \exists \nu X^\star$.

This is to be read: for every sentence X for which there is an appropriate $X^\star$ (one which results from X by replacing one or more occurrences of a name-letter in X by a variable, ν), every interpretation upon which X is true is one upon which the result of prefixing $X^\star$ by "$\exists$" followed by ν is also true.

8) $\forall x(Fx \rightarrow Gx), \exists xFx \vDash_Q \exists xGx$.

An interpretation upon which the second premise is true must assign a non-empty set to "F"; but for the first premise to be true, the conditional "$F\alpha \rightarrow G\alpha$" is true in at least one case in which "$F\alpha$" is true, which means that "G" must be assigned a non-empty set. Such an interpretation is one upon which the conclusion is true.

9) $\forall x(Fx \rightarrow Gx), \exists xGx \nvDash_Q \exists xFx$.

A counterexample is an interpretation which assigns the empty set to "F", and a non-empty set to "G".

We need to distinguish between:

10) $\exists x(Fx \,\&\, Gx) \vDash_Q \exists xFx \,\&\, \exists xGx$

and

11) $\exists xFx \,\&\, \exists xGx \nvDash_Q \exists x(Fx \,\&\, Gx)$.

Ex. 4.31 Establish the truth of **(14.10)** and **(14.11)**, in the latter case by providing a counterexample.

Working through examples like these should give a good feel for **Q**-validity, but what has become of an ideal mentioned earlier: that there be an entirely mechanical test for validity in an artificial language fit for logical purposes?

A *decision procedure for* **Q** is a mechanical method for determining, with respect to an arbitrary **Q**-sentence, and in a finite number of steps, whether or not it is valid. The existence of a decision procedure would indeed satisfy the hankering for mechanical tests. However, it can be proved that there is no decision procedure for **Q**. So that is a hankering which one must simply abandon.

There are systematic procedures which, for every **Q**-valid **Q**-sentence will determine in a finite number of steps that it is valid. As one is putting them through their paces, passing from step to step in accordance with the instructions, there is no point at which one can say: we haven't proved the sentence valid, therefore it is invalid. True, there will be a proof of validity in a finite number of steps, if the sentence is valid, but one does not know what that number is, and any number of steps one has taken may fall just short of the number required for a proof of validity.

I will simply mention three systematic procedures, for the benefit of readers who may already be acquainted with them: axiom systems; systems of natural deduction; and semantic tableaux (or tree) methods. Working with these procedures leads to a sharpened perception of **Q**-validity, and the fact that these procedures exist is, of course, of great importance.

Ex. 4.32 (a) Argue for the truth of the following:
(i) $\forall x(Fx \rightarrow Gx) \vDash_Q \exists x(Fx \rightarrow Gx)$
(ii) $\vDash_Q \exists x(Fx \vee \neg Fx)$
(iii) $\vDash_Q \forall x((Fx \rightarrow Gx) \vee (Gx \rightarrow Fx))$

(b) Devise counterexamples to the following. (A counterexample to an argument is an interpretation upon which the premises are true and the conclusion false. A counterexample to a sentence is an interpretation upon which it is false.)
(i) $\forall x(Fx{\rightarrow}Gx); \exists xFx$
(ii) $\exists x\exists y\, x \neq y$
(iii) $\forall x(Fx{\rightarrow}Gx), \exists x\neg Fx; \exists x\neg Gx$

15 Formalizing arguments

A valid argument which is not **P**-valid:

1) John runs. Therefore someone runs.

There is a **Q**-valid formalization of it:

2) $F\alpha \models_Q \exists xFx$,

with the obvious correspondences. We will for the moment take it for granted that the truth of (**2**) establishes the validity of (**1**).

An old favourite:

3) All men are mortal. Socrates is a man. Therefore Socrates is mortal.

This has a valid **Q**-formalization:

4) $\forall x(Fx \rightarrow Gx), F\alpha \models_Q G\alpha$

with "F" corresponding to "is a man", "G" to "is mortal" and "α" to "Socrates".

One early method of formalizing everyday arguments, Aristotle's syllogistic, had particular trouble with arguments like:

5) All horses are animals. Therefore all heads of horses are heads of animals.

This is **Q**-validly formalizable:

6) $\forall x(Fx \rightarrow Gx) \models_Q \forall x\forall y((Fx \ \& \ Hyx) \rightarrow \exists z(Gz \ \& \ Hyz))$

with "F" corresponding to "is a horse", "G" to "is an animal" and "Hxy" to "x is a head of y". We can argue informally for the truth of (**6**) as follows. Suppose some interpretation, i, verifies the premise. Then every member of $i(F)$ is a member of $i(G)$. To falsify the conclusion, some $\alpha + \beta$-variant of i, $i^\star$, must verify "$F\alpha \ \& \ H\beta\gamma$" and falsify "$\exists z(Gz \ \& \ H\beta z)$", which means that no γ-variant verifies "$G\gamma \ \& \ H\beta\gamma$". But there is such a variant, i', where $i'(\gamma) = i^\star(\alpha)$. This object belongs to $i^\star(F)$ and so to $i'(F)$ and so to $i'(G)$, so i' verifies "$G\gamma$"; and since $<i^\star(\beta), i^\star(\alpha)>$ belongs to $i^\star(H)$, the same goes for $<i'(\beta), i'(\gamma)>$ and $i'(H)$. So i' verifies "$G\gamma \ \& \ H\beta\gamma$". So there can be no counterexample. The idea is that if we have a horse and a head which satisfy the conclusion's antecedent, the premise assures us that we thereby have an animal and a head which satisfy the conclusion's consequent.

Aristotle's syllogistic has been criticized on the grounds that it counts as valid arguments which are not valid. An alleged example is:

7) All unicorns are self-identical. All unicorns are non-existent. Therefore some self-identical things are non-existent.

Aristotelian logic regarded this as an instance of a valid argument-form because it took the truth conditions of a universal quantification, "All Fs are Gs", to require the existence of Fs. Setting aside the question whether the argument is valid, it is certainly the case that:

8) $\forall x(Fx \rightarrow Gx), \forall x(Gx \rightarrow Hx) \not\models_Q \exists x(Gx \ \& \ Hx)$.

An interpretation which assigns the null set to every predicate-letter establishes the truth of (**8**).

Consider

9) Only the brave deserve the fair. Harry is brave and Mary is fair. So Harry deserves Mary.

10) $\forall x\forall y((Gxy \ \& \ Hy) \rightarrow Fx), F\alpha \ \& \ H\beta; G\alpha\beta$

with "Gxy" corresponding to "x deserves y", "H" to "is fair", "F" to "is brave", "α" to "Harry" and "β" to "Mary" yields a **Q**-invalid argument. The invalidity can be seen by considering an interpretation which assigns the set of even numbers to "F", the set of odd numbers to "H", 8 to "α", 3 to "β" and to "G" the set of ordered pairs such that the first member of each pair is greater by one than the second.

A valid argument resembling (**9**) is:

11) Only the brave deserve the fair. Mary is fair but Harry isn't brave. So Harry doesn't deserve Mary.

With correspondences as before, this can be **Q**-validly formalized:

12) $\forall x \forall y((Gxy \,\&\, Hy) \rightarrow Fx),\ G\beta \,\&\, \neg F\alpha;\ \neg G\alpha\beta.$

The **Q**-validity is plain if we reflect that all that matters about the premise is its instance "$(G\alpha\beta \,\&\, H\beta) \rightarrow F\alpha$", and that the following is true:

13) $(p \,\&\, q) \rightarrow r,\ q \,\&\, \neg r \models_P \neg p.$

Now for some examples of arguments involving identity, definite descriptions and numeral adjectives.

14) Hesperus is a planet. Hesperus is identical to Phosphorus. So Phosphorus is a planet.

With "α" corresponding to "Hesperus", "β" to "Phosphorus" and "F" to "is a planet", we can formalize **Q**-validly:

15) $F\alpha,\ \alpha = \beta \models_Q F\beta.$

Any interpretation upon which the premises are true assigns the same object to "α" and "β" and that object to the set it assigns to "F"; so it verifies the conclusion.

Compare:

16) John believes that Hesperus is a planet. Hesperus is identical to Phosphorus. So John believes that Phosphorus is a planet.

If we could make "Fx" correspond to "John believes that x is a planet", then (**16**) would be formalizable by (**15**), but, intuitively, (**16**) is invalid. Suppose John does not realize that Hesperus is identical with Phosphorus. He uses "Hesperus" of a heavenly body he sees in the evening. He uses "Phosphorus" of a heavenly body he sees in the morning (never suspecting that these are one and the same). He believes that Hesperus is a planet, but believes that Phosphorus is not a planet but a star. (When you ask "Is Phosphorus a planet?" he replies, firmly, "No".) So for this case the premises are true and the conclusion false, so (**16**) is not valid.

The moral is that "John believes that x is a planet" should not be allowed to count as a predicate. If a predicate is something adequately formalizable by a predicate-letter, the justice of this ruling can be shown not just by the example considered, but more generally. A predicate-letter is, on any interpretation, assigned a set of things of which the predicate-letter is true upon the interpretation. But there is no set of things of which "John believes that x is a planet" is true. This is shown by the fact that Hesperus ought to be both a member of and not a member of any such set, and this is impossible.

Consider

17) Only the fastest walker will reach London. John walks faster than Mary. So Mary will not reach London.

We might offer the formalization

18) $\exists x(\forall y(x \neq y \rightarrow Fxy) \,\&\, Gx \,\&\, \forall z(Gz \rightarrow z = x)),\ F\alpha\beta;\ \neg G\beta$

with "Fxy" corresponding to "x walks faster than y", "G" to "will reach London", "α" to "John" and "β" to "Mary". The idea is to treat the first premise of (**17**) as saying that someone walks faster than anyone else and will reach London, and no one else will reach London. However, (**18**) as it stands is not **Q**-valid. We need to add to the premises "$\alpha \neq \beta$", and replace "$F\alpha\beta$" by "$\neg F\beta\alpha$"; thus amended, the argument is **Q**-valid. We have interpreted "only" in such a way that "Only α is F" entails "α is F". We have not needed to use Russell's theory of descriptions to formalize "the fastest walker", and hence we have not included the uniqueness, which "the" imparts, in the formalization. If "F" is assigned the set of ordered pairs such that the first loves the second, then "$\forall y(\alpha \neq y \rightarrow F\alpha y)$" can be true on interpretations differing only in what they assign to "α". That is, more than one person can satisfy the condition of loving everyone else. If no more than one person can walk faster than everyone else, that is to do with the nature of the *faster than* relation rather than with the logical form of (**18**).

Ex. 4.33 Are there occurrences of "Only F" which are more plausibly interpreted as not entailing that there are Fs? (Cf. McCawley [1981], pp. 180–2.)

The following argument requires the uniqueness to be shown in the formalization, if the formalization is to be valid:

19) Only the fastest walker will reach London. John will reach London. So only John walks faster than anyone else.

Using the style of (**18**), and the same correspondences, we would get:

20) $\exists x(\forall y(x \neq y \rightarrow Fxy) \,\&\, Gx \,\&\, \forall z(Gz \rightarrow z = x)), G\alpha;$
$\forall x(\forall y(x \neq y \rightarrow Fxy) \rightarrow x = \alpha) \,\&\, \forall y(\alpha \neq y \rightarrow F\alpha y).$

This is not **Q**-valid, though (**19**) is valid. (**20**) fails to capture the validity through failing to formalize the uniqueness implied by "the" in the premise. This can be captured by:

21) $\exists x(\forall z(x \neq z \rightarrow Fxz) \,\&\, \forall y(\forall z(x \neq z \rightarrow Fyz) \rightarrow x = y) \,\&\, Gx \,\&$
$\forall z(Gz \rightarrow z = x)), G\alpha;$
$\forall x(\forall y(x \neq y \rightarrow Fxy) \rightarrow x = \alpha) \,\&\, \forall y(\alpha \neq y \rightarrow F\alpha y).$

The first premise of (**21**) is easier to read if we see the first part of it as an instance of the familiar:

$\exists x(Fx \,\&\, \forall y(Fy \rightarrow x = y))$

with "F" replaced by "$\forall z(x \neq z \rightarrow Fxz)$".

Ex. 4.34 Provide a counterexample to (**4.20**).

Consider

22) Every man has two hands. Every hand has a thumb. So every man has two thumbs.

This may strike one as valid, even reading the "two" in the conclusion as "exactly two", but only by making explicit a number of presuppositions can it be formalized as **Q**-valid, with this reading of the conclusion. First, we assume that "two" in the first premise is intended as "exactly two". Secondly, we assume that "a thumb" in the second premise is intended as "exactly one thumb". Thirdly, we assume, and make explicit as a third premise, that the relation of having, as obtaining between a person and his bodily parts, is transitive, so that if the person has a hand, and the hand a thumb, then the person has a thumb.

"F" corresponds to "is a man", "Gxy" to "x has y", "H" to "is a hand" and "J" to "is a thumb".

23) $\forall x(Fx \rightarrow \exists y\, \exists z(y \neq z \,\&\, Gxy \,\&\, Gxz \,\&\, Hy \,\&\, Hz \,\&\, \forall w((Gxw \,\&\, Hw) \rightarrow w = y \vee w = z))),$
$\forall x(Hx \rightarrow \exists y(Jy \,\&\, Gxy \,\&\, \forall z((Gxz \,\&\, Jz) \rightarrow z = y))),$
$\forall x\, \forall y\, \forall z((Gxy \,\&\, Gyz) \rightarrow Gxz);$
$\forall x(Fx \rightarrow \exists y\, \exists z(y \neq z \,\&\, Gxy \,\&\, Gxz \,\&\, Jy \,\&\, Jz \,\&\, \forall w((Gxw \,\&\, Jw) \rightarrow w = y \vee w = z))).$

The example shows how formalization can bring to light hidden assumptions in an argument.

Finally, a valid argument that is not formalizable as **Q**-valid:

24) Necessarily, if there is a first moment in time, the history of the universe up to now is finite. Therefore if there had to be a first moment in time, the history of the universe up to now has to be finite.

In a **Q**-formalization, we cannot reach beyond the non-truth functional sentence connective "necessarily". The deepest **Q**-formalization of the premise which is not inadequate is "p", and it is obvious that however hard we try with the premise we cannot find an adequate **Q**-valid argument.

Ex. 4.35 (a) Formalize the following arguments and determine the validity of the formalizations:
(i) Harry loves anyone who loves him. Mary loves Harry. So Harry loves Mary.
(ii) Only Harry loves himself. So only Harry loves Harry.
(iii) No one but Harry lives in the house and Harry is not mad. So the inhabitant of the house is not mad.

(b) This book is silent on the problems of time and tense. Some appreciation of the problems can be obtained by attempting to formalize the following in such a way that the formalizations are **Q**-valid iff the arguments are intuitively valid (cf. Lacey [1971]):
(i) George is marrying Mary. Mary is an orphan. So George is marrying an orphan.
(ii) George married Mary. Mary is a widow. So George married a widow.

16 Attitudes

We saw from (**15.16**) that an adequate formalization of "John believes that Hesperus is a planet" cannot match "John believes that x is a planet" with a predicate-letter. It might seem that the task of providing **Q**-formalizations of sentences of this kind is hopeless. But one should not despair too quickly, since there is a proposal, due to Donald Davidson, for (in some respects adequately) **Q**-formalizing any sentence of the form "John believes that . . .", provided that what fills the dots is itself **Q**-formalizable. Indeed, the essence of the proposal applies more widely, to include also sentences of the forms: "John *knows* that . . .", "John *said* that . . .", John *wonders* whether . . .". The italicized expressions are called *verbs of propositional attitude.*

Davidson called his proposal "paratactic", on the grounds that it sees the relevant sentences as really pairs of sentences. Thus

1) John believes that Hesperus is a planet

is held to consist of

2) John believes that. Hesperus is a planet.

"That" is held to be a demonstrative pronoun, referring forward to the subsequent "Hesperus is a planet". The **Q**-formalization is thus:

Ex. 4.36 On Davidson's own account ([1969], pp. 105–6) the word "that", as used in reporting a propositional attitude, refers to the utterance token which follows. (For the distinction between *token* and *type*: if you are obey a request to say something twice, you may make two utterance tokens of a single utterance type.)

(a) Evaluate the suggestion that the existence of beliefs that never have been and never will be expressed argues for regarding "that", in sentences like (**16.2**), as referring forward to an utterance type rather than a token.

(b) Evaluate the suggestion that the truth of a sentence like

John believes that she is happy,

uttered in circumstances which make plain the referent of "she", argues for regarding "that", in sentences like (**2**), as referring forward to an utterance token rather than a type.

3) $F\alpha\beta$, $G\gamma$

with "Fxy" corresponding to "x believes y", "G" to "is a planet", "α" to "John", "β" to "that" and "γ" to "Hesperus". An intended interpretation of (**3**) will assign to "β" a certain sentence, namely the second sentence in (**2**).

This proposal formalizes (**15.16**):

John believes that Hesperus is a planet. Hesperus is identical to Phosphorus. So John believes that Phosphorus is a planet.

as follows

4) $F\alpha\beta$, $G\gamma$, $\gamma = \delta$; $F\alpha\varepsilon$, $G\delta$

with "F", "α", "β", "γ" and "G" as before, "δ" corresponding to "Phosphorus" and "ε" corresponding to the second "that" (the one referring to "Phosphorus is a planet"). Since the demonstrative pronouns refer to different things, they must be formalized by different name-letters.

Q-validity is strictly speaking undefined for (**4**), since there are two sentences in the conclusion. A standard proposal is to say that for such an argument to be valid, the truth of at least one of the sentences in the conclusion must be guaranteed by the truth of the premises, and so it is in (**4**). However, that would clearly not satisfy someone who wanted to use (**15.16**) in reasoning, for what would matter would be what corresponds to $F\alpha\varepsilon$ rather than what corresponds to $G\delta$. It is clear that the desired truth is not guaranteed by the truth of the premises in virtue of the **Q**-logical forms of (**4**). Given the invalidity of (**15.16**), this is a point in favour of Davidson's proposal.

The proposal has the merit of giving **Q**-valid **Q**-formalizations to intuitively valid English arguments involving propositional attitudes. The valid

5) John believes that Hesperus is a planet. That is true. Therefore John believes something true

can be given a valid **Q**-logical form as follows:

6) $F\alpha\beta$, $G\gamma$, $H\beta \models_{Q} \exists x(F\alpha x \,\&\, Hx)$

with "H" corresponding to "is true", and other correspondences as before. We have to regard both occurrences of "that" as referring to the same thing, and thus as formalizable by the same name-letter. The fact that English contains a sentence like the conclusion of **(5)**, which seems naturally formalizable by the conclusion of **(6)**, supports Davidson's proposal by suggesting that "believes" is at least sometimes a predicate of degree two. And if sometimes, why not always?

Ex. 4.37 Using Davidson's paratactic proposal, formalize (making explicit any assumptions you need to make):

John believes that Hesperus is a planet. "Hesperus is a planet" is true. Therefore John believes something true.

A problem with the proposal is that we can also derive "$G\gamma$", which corresponds to "Hesperus is a planet". This happens to be true, but it ought not to follow from "John believes that Hesperus is a planet", and in other cases we would move from truth to falsehood. If we formalize

7) John believes that the earth is flat

as

8) $F\alpha\beta$, $G\gamma$

(with "α", "β" and "F" as before, "G" corresponding to "is flat" and "γ" to "the earth") we would formalize an argument with **(7)** as premise and "the earth is flat" as conclusion as **Q**-valid, which is obviously unsatisfactory. Davidson's proposal involves a contrast between what one asserts, and what one says without asserting. In an assertive utterance of "John believes that the earth is flat", Davidson's theory has it that "John believes that" is asserted, but "the earth is flat" is not asserted. Hence Davidson's view is not subject to the difficulty just noted. But such distinctions are not accommodated within **Q**.

Ex. 4.38 (a) Using Davidson's paratactic proposal, formalize:

Galileo said that the earth moves, and Newton said the same. So there is something that they both said

(b) Can Davidson's proposal be modified so as to provide an adequate formalization of, for example,

Referring to Jones' murderer, John said that he was innocent?

See Platts [1979], ch. 5, §5; Hornsby [1977].

(c) Can Davidson's proposal give an adequate account of the truth conditions of, for example,

John believes something which no one has ever or will ever express in an utterance?

See Schiffer [1987], ch. 5.

17 Binary quantifiers

There are quantifier expressions with no correlates in **Q**, for example "most" and "few". Let us see whether we can formalize English sentences containing these quantifiers by just adding them to **Q**, to form a new language, call it **Q+**.

We will write the new quantifiers "T" (for "most") and "W" (for "few") and add the syntactic rule that all **Q**-sentences are **Q+** sentences, and if X is a **Q+**-sentence and $X^\star$ results from X by replacing a name-letter by some variable, v, not in X, then the following are **Q+**-sentences:

$\mathrm{T}vX^\star$, $\mathrm{W}vX^\star$.

(We need to add that the modes of combination, e.g. by "&", which form **Q**-sentences also form **Q+**-sentences.) Using these quantifiers to formalize sentences which are (in a respect to be made more precise shortly) like "Everything is physical", there are no problems.

1) Most things are physical

and

2) Few things are physical

are **Q+**-formalizable as

3) $\mathrm{T}xFx$

and

4) $\mathrm{W}xFx$.

Thinking just of such cases, appropriate rules for interpretations could be modelled on the rule (**2.2**vi), replacing talk of all interpretations by talk of most or a few. However, quantifiers thus specified cannot adequately formalize sentences like

5) Most men lead lives of quiet desperation

(or the corresponding optimistic sentence with "few" for "most"). If we modelled our attempt on the method used to formalize similar sentences starting with "all" we would write:

6) $\mathrm{T}x(Fx \rightarrow Gx)$

with "F" corresponding to "is a man" and "G" to "leads a life of quiet desperation". However, this formalizes not (**5**) but rather:

7) Most things are such that: if they are men then they lead lives of quiet desperation.

(**7**) is true if most things are not men, for it is true iff most things in the universe have this property: that if they are men then they lead lives of quiet desperation. One way in which this could be true, at least as formalized by (**6**), is for most things not to be men. Then most things vacuously have the property in question. But this is not a condition upon which (**5**) is true. Hence (**6**) is not an adequate formalization of (**5**).

Perhaps only the details of the strategy were wrong, and "→" simply the wrong connective to use. But we also cannot use "&", as we do when formalizing similar existentially quantified sentences. For, with correspondences as before,

8) $\mathrm{T}x(Fx \ \& \ Gx)$

is true only upon an interpretation which assigns most things to the intersection of what it assigns to "F" and "G". In other words, formalizing (**5**) by (**8**) misrepresents the former as requiring for its truth that most things in the universe are both men, and also leaders of quietly desperate lives. No truth functional sentence connective can be inserted in the place marked ¢ in

9) $\mathrm{T}x(Fx ¢ Gx)$

in such a way as to yield an adequate formalization of (**5**).

Ex. 4.39 To show that no truth fractional connective can replace "¢" in (**10.9**) so as to produce an adequate formalization of "most", one has to proceed by an enumeration of cases. The ones considered in the text show that the falsity of the antecedent to the supposed connective must not be sufficient for the truth of the compound, and that the truth of both antecedent and consequent must not be sufficient for the truth of the compound. Establish the result for the remaining cases.

This is one motivation for a somewhat different conception of quantification, which I shall now introduce. Let us say that an *open sentence* is what results from a **Q+** or **Q**-sentence by replacing a name-letter by a variable not already contained in the sentence. (In this terminology, an open sentence is a kind of non-sentence.) We can characterize the quantifiers of **Q** as *unary quantifiers* because they take just one open sentence to make a sentence. This is why the formalization of sentences like "Everything is physical" was so straightforward, whereas the formalization of sentences like "All men are happy" was not. This last sentence contains two predicative expressions, "man" and "happy", welded into a sentence by the quantifier and the copula "are". To **Q**-formalize it we have to find a *single* open sentence for the quantifier to attach to. What seems approximately to do the trick is "$Fx \rightarrow Gx$". The problem with formalizing sentences like (**5**) was precisely that, with the available resources, we could not find a suitable single open sentence for the "T" quantifier to apply to.

A natural response is to introduce a quantifier which takes *two* open sentences to make a sentence: a *binary quantifier*. Let us add binary quantifiers "μ" (for "most") and "φ" (for "few") to **Q**, to create the language **QB**, by the stipulation that every **Q**-sentence is to be a **QB**-sentence, and if $X^\star$ and $Y^\star$ result from **QB**-sentences X and Y, by replacing in both some occurrence(s) of a name-letter by a variable, ν, which occurs in neither, then the following are also **QB**-sentences:

$\mu\nu(X^\star : Y^\star)$, $\varphi\nu(X^\star : Y^\star)$.

These are to be read: "most (few) *X*s are *Y*s". The mark ":" functions merely as punctuation.

Applying this to (**5**) yields:

10) $\mu x(Fx : Gx)$

with correspondences as for (**6**). On the appropriate interpretation, this will say that most things which are men lead lives of quiet desperation.

A rule of interpretation for "μ" could be phrased on the following lines (after (**2.2**vi)):

11) $\mu\nu(X: Y)$ is true upon an interpretation iff most n–variants upon which $X\frac{n}{\nu}$ is true are interpretations upon which $Y\frac{n}{\nu}$ is also true.

The English "most" used in the statement of the rule is naturally seen as a binary quantifier.

Ex. 4.40 (a) Would it make any difference to replace "most" in (17.11) by "more than half")

(b) Give an account of "Most natural numbers are non-prime" upon which it is false.

(c) Give an account of "Most natural numbers are non-prime" upon which it is true. Would the account also give the right result when applied to finite cases? (For the claim that "most" and "few" are not formalizable in first order logic, see Van Benthem and Doets [1983], p. 277.)

It is hard to resist the thought that "every" and "most" belong to the same linguistic category, and so should be treated in the same way. Any sentence containing "most" is still grammatical if that quantifier is replaced by "every"; and conversely. So it is natural to conclude that

if "most" is a binary quantifier, so is "every". Since we have the antecedent of this conditional, its conclusion comes naturally.

It is straightforward to add a binary universal quantifier to **QB**, say "λ". With the obvious correspondences, we could formalize "all men are happy" as

12) $\lambda x(Fx : Gx)$

which looks somewhat more like the English. On the formalization by a unary universal quantifier, an "$\rightarrow$" appeared which was invisible in the English, and there is no such distortion in (**12**).

Sentences like "everything is physical" appeared well adapted to the unary treatment. A closer look suggests that a binary treatment is closer to English even in this case. We cannot say "every is physical". "Thing" appears to function precisely as a first term to the quantifier, so that the appropriate formalization is still (**12**), but now with "F" corresponding to "is a thing" and "G" to "is physical". An intended interpretation will assign every thing in its domain to "F".

We saw earlier that there is room for doubt concerning whether English universal quantifications imply the existence of something corresponding to the first term (of John's children, for example, in (**3.4**) "All John's children are asleep"). The treatment of quantifiers as binary is neutral on this issue. The obvious rule of interpretation for "λ" is

13) $\lambda\nu(X: Y)$ is true on an interpretation iff all n-variants upon which $X\frac{n}{\nu}$ is true are interpretations upon which $Y\frac{n}{\nu}$ is also true.

The question of whether "$\lambda x(Fx: Gx)$" is true on an interpretation, i, which assigns the empty set to "F" becomes the question of whether it is true that "all α-variants upon which '$F\alpha$' is true are interpretations upon which '$G\alpha$' is also true", given that there are no n-variants upon which "$F\alpha$" is true. Presumably, those who are most struck by the suggestion that "All John's children are asleep" is not true if John has no children will answer the last question negatively; whereas those who are impressed by the truth of (**3.5**) (All bodies acted on by no forces continue in a uniform state of rest or motion) will incline to answer it positively, laying themselves open to the charge that perhaps in that case they ought also to assign truth to (**3.6**) (All bodies acted on by no forces undergo random changes of velocity). Binary quantification as such does not speak to the debate.

Ex. 4.41 (a) Define a binary quantifier (by giving the syntax and an appropriate rule of interpretation) suitable for formalizing "the". Illustrate by examples of formalizations. Do the truth conditions your formalizations attribute differ from those attributed by Russell's Theory of Descriptions?

(b) Define a binary quantifier (by giving syntax and an appropriate rule of interpretation) suitable for formalizing the quantifier "no". Formalize the following pair, using binary quantifiers in both cases:
(i) Every student will pass if he works.
(ii) No student will pass if he works.

Compare this treatment of (i) and (ii) with that awarded by unary quantifiers (i.e. their Q-formalizations). (For discussion, see Higginbotham [1986].)

There is another approach to quantifiers which is motivated by similar considerations and yields an essentially similar language. On this approach, a quantifier attaches to a predicate to form a "restricted quantifier", like "all men", which is then fit to attach to a second predicate, say "mortal", to form a sentence ("all men are mortal").

Either of these approaches offers an insightful way of specifying the contribution which quantifiers make to truth conditions. "All *F*s are *G*s" is true iff the set of *F*s is a subset of the set of *G*s; "Most *F*s are *G*s" is true iff the *cardinality* of the set of *F*s which are *G*s exceeds the cardinality of the set of *F*s which are not *G*s; "The *F* is *G*" is true iff the cardinality of the set of *F*s is 1 and the set is a subset of the set of *G*s. (The cardinality of a set is the number which says how many members it has.) Using $|\sigma|$ to express the cardinality of a set σ, and with i ranging over interpretations, the truth conditions of various binary quantifications can be listed:

14) $\lambda x(Fx: Gx)$ is true upon i iff $i(F) \supseteq i(G)$
some $x(Fx: Gx)$ is true upon i iff $|i(F) \cup i(G)| > 0$
$\mu x(Fx: Gx)$ is true upon i iff $|i(F) \cap i(G)| > |i(F) - i(G)|$
$\varphi x(Fx: Gx)$ is true upon i iff $|i(F) - i(G)| > |i(F) \cap i(G)|$
the $x(Fx: Gx)$ is true upon i iff $|i(F)| = 1$ & $i(F) \supseteq i(G)$.

For example, the last line says that a sentence of the form "the $x(Fx: Gx)$" is true upon i iff the cardinality of the set i assigns to F is 1, and everything which belongs to what i assigns to F belongs to what i assigns to G. This is the familiar Russellian truth condition in new notation.

18 Substitutional quantifiers

The Q-quantifiers are called "*objectual*": whether or not a quantification is true upon an interpretation depends on how things are with the *objects* in the domain of interpretation. For example, the rule for "$\exists$" entails that "$\exists xFx$" is true on an interpretation, i, iff some object in the domain of i is a member of what i assigns to "F".

An alternative style of quantifier is called "*substitutional*". The rule for such a quantifier makes whether a quantification is true upon an interpretation depend on whether sentences resulting from the quantification by deleting the quantifier and *substituting* a name for the variable of quantification are true. For an existential substitutional quantifier, written "E", the rule might be:

1) E$\nu X^\star$ is true upon an interpretation, i, iff for some name, N, $X\frac{N}{\nu}$ is true upon i.

Analogously, for a universal substitutional quantifier, written "A", the rule might be:

2) A$\nu X^\star$ is true upon an interpretation, i, iff for all names, N, $X\frac{N}{\nu}$ is true upon i.

Here, $X^\star$ results from a sentence by replacing a name-letter by some variable, ν, not already in the sentence, and $X\frac{N}{\nu}$ results from $X^\star$ by replacing each occurrence of ν by some name, N. We do not require that N be absent from $X^\star$.

Ex. 4.42 Show how requiring the absence of "*N*" from $X\frac{N}{\nu}$ would lead to unwanted truth-upon-an-interpretation conditions for sentences like "E*x*E*y*F*xy*".

Q, as defined so far, contains no names, but only name-letters. Names should be related to name-letters as predicates to predicate-letters. Whereas a predicate is assigned the same set on every interpretation (modulo differences of domain), a predicate-letter is assigned different sets upon different interpretations. Likewise, a name, as opposed to a name-letter, will be assigned the same object in every interpretation. Let us call a language "**QS**" if it adds to **Q** the two substitutional quantifiers recently mentioned, and also adds names, specifying, for each name, the object which any interpretation must assign to it.

QS is still underspecified, since we have not said what names it contains. Let us suppose, first, that it contains just the name "Reagan", and we stipulate that in every interpretation this is to be assigned Ronald Reagan, the famous American political film star. In conjunction with (**1**) and (**2**), this would ensure that "AxFx" and "ExFx" would alike be true upon an interpretation, *i*, iff Ronald Reagan is the member of what *i* assigns to "*F*". This is not incoherent, but has no discernible utility.

Could we specify **QS** in such a way that a substitutional quantification in **QS** is true upon an interpretation iff the corresponding objectual quantification in **Q** is true upon the corresponding interpretation? A necessary condition is that **QS** contain a name for every object in the domain of interpretation. If it did not, it would be "easier" for "AxFx" than for "$\forall xFx$" to be true upon an interpretation. A further necessary condition is that **QS** should not contain a name for some object not in the domain of interpretation.

The two necessary conditions are jointly sufficient for coincidence of objectual quantifiers of **Q** and the substitutional quantifiers of **QS**. If we held to these conditions, there would be no interest in substitutional quantification. There are two ways of modifying **QS** which would make substitutional quantification of interest. One is to allow **QS** to contain empty names; the other is to allow it to contain opaque contexts. I will discuss only the latter. (§**20** below contains some material on empty names.)

An *opaque context with respect to names* is one in which there is no guarantee that two co-referring names can be substituted *salva veritate*; that is, it is a context in which the substitutivity of identicals – (**8.5**) – fails; that is, the context ". . . . . ." is opaque with respect to names iff there is no guarantee that ". . . N_1 . . ." and ". . . N_2 . . ." have the

same truth value, despite the fact that "N_1" and "N_2" have the same bearer. The context

3) John believes that . . . is a student of Christ Church

is opaque with respect to names, since it may be that the first but not the second of the following is true

4) John believes that Charles Dodgson is a student of Christ Church

5) John believes that Lewis Carroll is a student of Christ Church

even though Dodgson is Carroll.

We cannot regard (**3**) as a predicate, for reasons already noted in §**15** (see (**15.16**)). Let us instead call it a "quasi-predicate". To formalize sentences like (**4**) and (**5**) we need what I shall call "quasi-predicate-letters". These letters cannot be interpreted by being assigned a set of objects. We will not consider how they are to be interpreted, but let us assume that somehow or other they can be, so that there are **QS**-formalizations of (**4**) and (**5**) which, upon an intended interpretation, are true and false respectively. Then by the interpretation rules for the substitutional quantifiers, both of the following are true upon an intended interpretation (where "ψ" corresponds to the quasi-predicate (**3**)):

6) E$x\psi x$

7) E$x\neg\psi x$.

These quantifiers are not objectual: the truth of (**6**) and (**7**) on an intended interpretation does not turn on how things are with some object, for the same object is involved in the verification of both, yet no one object can be both ψ and not-ψ.

Substitutional quantification is intelligible in contexts in which objectual quantification would not be. Consider the quasi-predicate "was so-called because of his size". The objectual quantification

8) $\exists x$(x was so-called because of his size)

is nonsense. What does the "so" refer back to? On the other hand, if a formalization has a quasi-predicate-letter, say ψ, corresponding to the quasi-predicate, and, moreover, the formalization of

9) Giorgione was so-called because of his size

is true upon the interpretation, and if, finally, "Giorgione" (or its formal equivalent) is included in the substitution class of names with respect to which the quantifier is defined, then the substitutional quantification

10) $Ex\psi x$

is also true upon the interpretation.

Substitutional quantifiers are perfectly intelligible. It is unclear whether any English quantifiers are substitutional. The best candidates are substitutional quantifiers whose variables occupy *predicate position* (see §**19**). If there are substitutional quantifiers in English whose variables occupy name position, then what (**10**) formalizes would be expressible in English, and true. But anything like

11) Something was so-called because of his size

would seem to have all the unintelligibility of (**8**).

Ex. 4.43 "Snow is white" is true iff snow is white.
This appears to hold not just for the sentence "snow is white", but quite generally. Attempt to formalize an appropriate generalization both in **Q** and in **QS**, and comment on the success of your efforts.

19 Predicate quantifiers and second order logic

Second order logic involves quantification over properties or sets. Such quantification can also be effected in the first order logic of **Q**. Hence we need to add some further specification if we are to say what second order logic is. I will begin by dwelling upon the way in which **Q** can be used to formalize English sentences apparently involving quantification over properties.

We seem to be able to generalize from a sentence like

1) Reagan and Thatcher are both powerful

in two kinds of way. One way is familiar:

2) Someone, x, and someone, y, are such that x is powerful and y is powerful

which is easily formalized in **Q**. The other way is:

3) There is something which both Reagan and Thatcher are.

The "something" is a property, and the generality in (**3**) is inferred from the fact that *powerful* is something both Reagan and Thatcher are, a fact expressed (in a slightly different way) by (**1**). Pressed to Q-formalize, we might offer

4) $\exists x(H\alpha x \;\&\; H\beta x)$

with "α" and "β" corresponding (on an intended interpretation) to the two leaders and "H" to the relation of *having*, the relation that holds between an object and a property when the object *has* the property. Nothing in **Q** restricts the domains of interpretation, so there is no reason not to include properties, on the assumption that there are such things. (Sometimes the word "individuals" is used in a technical sense, to mean all the entities in the interpretation domains of some first order language like **Q**. Then one must say, with only superficial awkwardness, that properties may be among the individuals.) So there is no special problem about "quantifying over properties" in **Q**. (If there are no properties, as nominalists maintain, then no matter what one's linguistic or logical resources one will not be able to quantify over them.)

Adopting this approach threatens to obscure some logical relations. (**3**) is supposed to follow from (**1**), but (**4**) will not be the conclusion of a Q-valid argument with a straightforward formalization of (**1**) as premise. So one might offer a non-straightforward formalization of (1) as:

5) $H\alpha\pi \;\&\; H\beta\pi$

with "H", "α" and β" as before and "π" corresponding to the property of being powerful, regarded as an object fit to belong to a domain of Q-interpretation. (**4**) and (**5**) represent the argument whose premise is (**1**) and whose conclusion is (**3**) as **Q**-valid. The formalization of (**1**) as (**5**) may seem unnatural, but it has a point, though it would be misguided to attempt to formalize every predicate by a name.

Ex. 4.44 Why would it be misguided to attempt to formalize every predicate by a name?

The **Q**-quantifiers are "*name quantifiers*": their variables occupy the kind of position that names can occupy. We could extend **Q** by adding *predicate quantifiers*, quantifiers whose variables occupy the kind of position that predicates can occupy. Using "∇" for the universal quantifier and "Δ" for the existential, the syntactic rule could be:

6) If X is a **Q**-sentence then so are $\nabla f X^\star$ and $\Delta f X^\star$, where $X^\star$ results from X by replacing one or more occurrences of a predicate or predicate-letter in X by f.

Instead of (**4**), we could formalize (**2**) as

7) $\Delta f(f\alpha \; \& \; f\beta)$.

The proposal can be intended in more than one way.

(a) It merely provides notational abbreviations for the kinds of **Q**-formulae we have already, in particular the kind used in (**4**) and (**5**). No new semantic ideas are invoked: "∇" and "Δ" are used in essentials like "∀" and "∃", except for the suggestion that an intended interpretation will see an expression of the form "fx" as abbreviating something like "Hfx" and will assign to "H" a set of ordered pairs whose first member is a property and whose second member is a possessor of that property. This conservative way will not treat as logical truths some expressions which arguably are logical truths, for example

8) Everything has some property,

which emerges as

9) $\forall x \Delta f(fx)$.

If this is as an abbreviation of "$\forall x \exists y(Hxy)$" (containing special guidance about intended interpretations) it is not **Q**-valid.

(b) We can restrict the interpretation of the new quantifiers in specific ways. Given the examples considered so far, the natural restriction would be to properties. The effect of this stipulation would be unclear in two ways. First, it is unclear what the general ("logical"?) truths about properties are. For example, must every property have an instance? Second, it is unclear why this would achieve anything significantly different from what is achieved by (a). It would seem that we could have simply included properties in the domain anyway, in which case the truth of sentences containing the existing quantifiers would have been sensitive to how things are with properties.

These unclarities are resolved by an alternative proposal, which gives the core of the standard form of second order logic: the new quantifiers are to range over subsets of the domain of interpretation. This allows set theory to supply definite answers to some questions. For example, not every subset of the domain must have a member, but everything in the domain is a member of some subset of the domain, thus showing that (**9**) is valid on these semantics. Moreover, the proposal gives a clear answer to why the new entities could not have been simply members of the domain. The Russian-born mathematician Cantor proved that the subsets of a set are more numerous than the set; so the cardinality of the set of entities the predicate quantifiers are sensitive to is greater than the cardinality of the set of entities the name quantifiers are sensitive to. The effect of this proposal could not be achieved in the style of (a).

On this approach, second order logic stands in some intimate relation to set theory, for set-theoretical truths are instrumental in determining which sentences are valid (as we just saw with (**9**)). This is not to say that second order logic is set theory. For example, "There is a set which contains everything as a member" is not true on standard set theories (to suppose it to be true would quickly lead to paradox), whereas

10) $\Delta f \forall x(fx)$

is valid in the proposed second order logic. This disparity is explained by the fact that (**10**) does not say that there is a set which contains everything as a member in some absolute way. Rather, it is valid because for every interpretation, there is a set, namely the domain itself, to which everything in the domain belongs.

(c) Second order logic, as developed along the lines of (b), does not treat the relation between predicate variables and predicates on a par with that between name variables and names. We can replace a name variable by an expression fit to refer to some entity to which name quantifiers are sensitive, that is, some entity in the domain of interpretation, and the result is an intelligible sentence; but we cannot replace a predicate variable by an expression fit to refer to some entity to which predicate quantifiers are sensitive, that is, some subset of the domain of interpretation, with the guarantee that what results will be an intelligible sentence. We can lop off the "$\exists x$" in "$\exists x\ x = x$" and replace the variable by a name and end up with the fully intelligible "Reagan = Reagan". But if we perform a similar operation upon, for example, (**7**) ($\Delta f(f\alpha \ \&\ f\beta)$) the result is not strictly intelligible: (the set of men)α & (the set of men)β. What is missing is the use of the predicate to ascribe something to something, to say something about something. This is what makes the juxtaposition of a name and a predicate more than a mere list, and is the feature of predication which Frege [1892a] referred to as "unsaturatedness". Not suggesting a philosophical account of this matter is of no significance whatever for the purposes for which second order logic was introduced: a juxtaposition like "(the set of men)α" can be regarded as an abbreviation of "α belongs to the set of men". But it is of some significance in connection with the attempt to understand what appear to be predicate quantifiers in English, as in (**3**) (There is something which both Reagan and Thatcher are). This differs both from "There is something to which both Reagan and Thatcher belong – viz., a set" and from "There is something which both Reagan and Thatcher have – viz., a property". We can naturally add a *videlicet* clause to (**3**) as follows:

11) There is something which both Reagan and Thatcher are – viz., powerful.

"Powerful" is an adjective, and not a noun or noun phrase (like "power" or "being powerful") fit to refer to a property. (Compare "There is something which both Reagan and Thatcher are – viz., happy". "Happy" does not, or does not just, refer to happiness: "happiness" is the closest word which does that, but we cannot intelligibly conclude the sentence just quoted with "– viz., happiness".)

Quantification is naturally thought of as quantification over entities: over individuals, or properties or sets. Yet no entity seems to capture the predicative character of predicates. Genuine predicate quantification would be quantification into genuinely predicate position, that is, position in which predicative character is retained. Hence there is a tension in the very notion of predicate quantification: there is pressure to think of it as *over entities*, and pressure to think of it as involving more than entities. In standard second order logic, quantification is over entities, and there is no attempt or need to explain "unsaturatedness".

A type of quantification more apt to retain the predicative character of predicates would be substitutional quantification into predicate position. "$\nabla f(\ldots f \ldots)$" would be true iff the result of replacing the variable in "$(\ldots f \ldots)$" by a predicate (from some specified class of predicates) is true. This would give a natural understanding of the *videlicet* clause in (**11**): it supplies a predicate which verifies the quantification. Though this may shed some light on some English idioms, I do not know that it has any significance for logic.

20 Free logics

A "free logic" is one which rejects the following assumption:

1) Every name (or name-letter) refers to something.

A "universally free logic" is one which rejects the following assumption:

2) Only non-empty domains feature in the definition of validity.

The assumptions are independent: either can be rejected without the other. **Q** is committed to both assumptions, as the following facts make plain:

3) $\models_{\mathbf{Q}} \exists x\ x = \alpha$.

4) $\models_{\mathbf{Q}} \exists x(Fx \lor \neg Fx)$.

In this section, I shall mostly consider various ways in which (**1**) can be rejected, and some philosophical motivations. I close the section with some brief remarks about (**2**).

Motivations for rejecting (**1**):

(a) Our language seems to contain names with no bearers, like "Vulcan", and an adequate logic should be able to deal with them; so **Q** falls short of its aspiration to formalize English sentences.

(b) Logic is apriori, and so are logical relations like validity. We cannot tell apriori whether any of our names have bearers. In the case of many names, like "London", we are confident that they do have bearers, but the confidence is based on non-apriori empirical knowledge; in other cases, like "Vulcan", we are confident that they do not have bearers, but this emerged from an astronomical discovery (there was observed to be no planet between Mercury and the sun), not from apriori reflection; in yet other cases, the issue is in dispute (perhaps experts still differ on the question whether there was really any such person as Homer). If an English argument guaranteed the truth of its conclusion only on the assumption that the names occurring in it had bearers, we could not tell apriori whether, if the premises were true, so would the conclusion be. To formalize these names with name-letters would wrongly suggest that we can tell apriori that they have bearers, for name-letters are by stipulation assigned objects (bearers).

(c) Some **Q**-valid arguments formalize invalid English arguments, if there actually are empty names; since there are, **Q** is incorrect. For example, the English argument

5) Everything is just as heavy as itself; so Vulcan is just as heavy as Vulcan

has a premise which is true but a conclusion which is not, yet if we use **Q**-validity as our guide to validity we might be tempted to classify (**5**) as valid (since $\forall x\ Rxx \models_Q R\alpha\alpha$).

Ex. 4.45 Can there be an "intended interpretation" of $R\alpha\alpha$, regarded as a formalization of "Vulcan is just as heavy as Vulcan"? Use your answer to evaluate the claim that the practice of **Q**-formalizing would lead us to classify as valid arguments with premises which are true and conclusions which are not.

(d) Some **Q**-valid arguments formalize invalid English arguments, if some actual entities might not have existed. For in that case

6) Everything is perishable; so Socrates is perishable,

though supposedly an example of $\forall xFx \models_Q F\alpha$, is not valid: since Socrates might not have existed, it might have been that the premise is true and the conclusion is not (Socrates could not be perishable without existing). This point does not depend on the supposition that there are any empty names.

Ex. 4.46 Show how a formalization of (**20.6**) using binary quantifiers would undermine this argument for free logic.

Those who find one or more of these arguments convincing will reject (**1**). There is general agreement that doing this requires changes in the quantifier rules. Using "$\forall x(\ldots x \ldots)$" to represent an arbitrary universal quantification, **Q** says:

7) $\forall x(\ldots x \ldots) \models_Q (\ldots \alpha \ldots)$.

Using "**QF**" for free logic, the closest correct **QF** claim is

8) $\forall x(\ldots x \ldots), \exists x\ x = \alpha \models_{QF} (\ldots \alpha \ldots)$.

The logical rule of "universal quantifier elimination" or "specification" must be restricted so that only a name with a bearer can replace the quantified variable. **QF** requires a similar modification to the existential quantifier.

Ex. 4.47 Using an example, explain why the argument in (**20.7**) ought not to be **QF**-valid.

These generally agreed changes in quantifier rules do not resolve what to go on to say about sentences with empty names.

Negative free logic: All atoms containing a bearerless name are false. So "Vulcan is a planet" is false; since the negation of a falsehood is a truth, "It is not the case that Vulcan is a planet" is true (Burge [1974]; Bostock [1997], pp. 356ff.).

Positive free logic: Some atoms containing a bearerless name are true. A likely example would be "Vulcan is Vulcan" given that this "follows from the unexceptionable identity principle '$x = x$'" (Lambert [1991b], p. 25). Sensible versions of this view will draw back from holding that all are true; not, for example, "Vulcan is a noted logician".

Fregean free logic: all sentences containing a bearerless name lack truth value. This seems to have been Frege's view. If S is neither true nor false, "not-S" (assuming "not" to introduce the standard kind of negation) is also neither true nor false: not true, for that would require S to be false, and not false, for that would require S to be true. In general, the idea is that however complex the sentence in which the bearerless name occurs, it will lack a truth value. (Frege [1892b]; Lehman [1994].)

I will review some motivations for these various options.

(1) Predication

One of the most basic acts in thought or speech would appear to be that of predicating something of something, for example, predicating being happy of John. The result is true iff there is something *of* which something is predicated, and that object is as it is said to be. Hence a predication which fails to be *of* anything cannot be true.

If this is accepted, as I think it should be, positive free logic is excluded, but the point is consistent with both Fregean and negative free logic. The argument can be developed into one for the Fregean view if we add that to predicate falsely is to say *of* something that it is other than it is: no object, no falsehood.

(2) Fiction

We discriminate among fictional sentences, holding, for example, that

9) Holmes was a detective

is true whereas

10) Holmes was a farmer

is false. Since there was no such person as Holmes, we must treat "Holmes" as an empty (bearerless) name. This supposedly shows that free logic should be positive.

The argument has more than one weakness. First, it is not uncontroversial that "Holmes" is an empty name, for arguably it is a name of a fictional character. One ought to be specially inclined to this view if one accepts without qualification that **(9)** and **(10)** are literally true and false respectively.

Secondly, it is unclear that one ought to accept this. To check that **(10)** is true we go to the works of Conan Doyle, and it is enough that this sentence follows from what Doyle wrote. This strongly suggests that either **(9)** is elliptical for something like

11) According to the stories, Holmes was a detective

or else that in believing that **(9)** is *true* we are really believing that it is *true-in-the-stories*.[3] If the former, then **(9)** and **(10)** are not strictly speaking fictional sentences after all; if the latter, **(9)** is not strictly true.

Perhaps the examples can still serve the cause of positive free logic; for even after reinterpretation, we have cases of both true and false sentences containing a name which is arguably empty. This suggestion needs to be set within a positive account of the kinds of context generated by the envisaged kind of operator ("according to the stories"). Suppose, for example, that "according to the stories (p)" meant something like "among the sentences making up the stories, there occurs 'p'". Then "Holmes" would not really be being *used* in a sentence like **(9)**, understood as elliptical for **(11)**: it would merely be being *mentioned*, and so there is no inference to the conclusion that an empty name can be used in a true sentence.

In any case, this conclusion would tell at most only against Fregean free logic. It could be accepted by both negative and positive versions. The negative free logician says only that all *atoms* containing an empty name are false, and **(11)** is clearly not atomic.

(3) Logic

If we have limitless respect for "laws of logic" like "$p \vee \neg p$" or "$\alpha = \alpha$", we may incline to suppose that they hold quite unrestrictedly; hence that we must have truths like

[3] We certainly need an operator like the one envisaged in **(11)** to do justice to the thought that fictions are sometimes woven around real objects, a thought which might have a form along the lines:

$\exists x$ (according to some story) $(\ldots x \ldots)$.

Positive free logic: Some atoms containing a bearerless name are true. A likely example would be "Vulcan is Vulcan" given that this "follows from the unexceptionable identity principle '$x = x$'" (Lambert [1991b], p. 25). Sensible versions of this view will draw back from holding that all are true; not, for example, "Vulcan is a noted logician".

Fregean free logic: all sentences containing a bearerless name lack truth value. This seems to have been Frege's view. If S is neither true nor false, "not-S" (assuming "not" to introduce the standard kind of negation) is also neither true nor false: not true, for that would require S to be false, and not false, for that would require S to be true. In general, the idea is that however complex the sentence in which the bearerless name occurs, it will lack a truth value. (Frege [1892b]; Lehman [1994].)

I will review some motivations for these various options.

(1) Predication

One of the most basic acts in thought or speech would appear to be that of predicating something of something, for example, predicating being happy of John. The result is true iff there is something *of* which something is predicated, and that object is as it is said to be. Hence a predication which fails to be *of* anything cannot be true.

If this is accepted, as I think it should be, positive free logic is excluded, but the point is consistent with both Fregean and negative free logic. The argument can be developed into one for the Fregean view if we add that to predicate falsely is to say *of* something that it is other than it is: no object, no falsehood.

(2) Fiction

We discriminate among fictional sentences, holding, for example, that

9) Holmes was a detective

is true whereas

10) Holmes was a farmer

is false. Since there was no such person as Holmes, we must treat "Holmes" as an empty (bearerless) name. This supposedly shows that free logic should be positive.

The argument has more than one weakness. First, it is not uncontroversial that "Holmes" is an empty name, for arguably it is a name of a fictional character. One ought to be specially inclined to this view if one accepts without qualification that **(9)** and **(10)** are literally true and false respectively.

Secondly, it is unclear that one ought to accept this. To check that **(10)** is true we go to the works of Conan Doyle, and it is enough that this sentence follows from what Doyle wrote. This strongly suggests that either **(9)** is elliptical for something like

11) According to the stories, Holmes was a detective

or else that in believing that **(9)** is *true* we are really believing that it is *true-in-the-stories*.[3] If the former, then **(9)** and **(10)** are not strictly speaking fictional sentences after all; if the latter, **(9)** is not strictly true.

Perhaps the examples can still serve the cause of positive free logic; for even after reinterpretation, we have cases of both true and false sentences containing a name which is arguably empty. This suggestion needs to be set within a positive account of the kinds of context generated by the envisaged kind of operator ("according to the stories"). Suppose, for example, that "according to the stories (p)" meant something like "among the sentences making up the stories, there occurs 'p'". Then "Holmes" would not really be being *used* in a sentence like **(9)**, understood as elliptical for **(11)**: it would merely be being *mentioned*, and so there is no inference to the conclusion that an empty name can be used in a true sentence.

In any case, this conclusion would tell at most only against Fregean free logic. It could be accepted by both negative and positive versions. The negative free logician says only that all *atoms* containing an empty name are false, and **(11)** is clearly not atomic.

(3) Logic

If we have limitless respect for "laws of logic" like "$p \vee \neg p$" or "$\alpha = \alpha$", we may incline to suppose that they hold quite unrestrictedly; hence that we must have truths like

[3] We certainly need an operator like the one envisaged in **(11)** to do justice to the thought that fictions are sometimes woven around real objects, a thought which might have a form along the lines:

$\exists x$ (according to some story) $(\ldots x \ldots)$.

12) Vulcan is a planet or Vulcan is not a planet

13) Vulcan = Vulcan.

On second thoughts, logic does not hold for quite all sentences, for we exclude questions, commands, nonsense, etc. Frege [1892b] seems to have believed that fiction was a species of nonsense: there are no genuine thoughts in fiction, but only mock thoughts; we do not really make assertions, true or false, but merely play at making them. He would presumably have assimilated sentences like (**12**) and (**13**) to unwitting examples of this sort of nonsense.

The view is especially implausible when extended in the way just envisaged beyond the sphere of fiction. Whereas fiction is an activity engaged in wittingly, the possibility of unwitting fiction makes which sentences are "nonsense" and which are not a matter not available apriori, so it would be inconsistent with the supposed apriori status of logic.

One could imagine (**12**) having been used in some kind of *reductio* proof of Vulcan's non-existence. Something similar is used in a classical proof that there is no greatest prime ("either the greatest prime is odd, or it is even"). So there are non-negligible reasons to reject the Fregean version of free logic.

Negative and positive free logicians agree that (**12**) is true: they are separated by (**13**). Both agree that it does not follow from "$\forall x\ x = x$" (cf. (**8**)). The positive free logician may well adopt a semantics which allows for the truth of (**13**) by positing an additional "null" entity, lying outside the domain, and which serves as the referent of the empty names; more exactly, being assigned the null entity is a reflection of the fact that the name is empty (cf. Scott [1967]). Though this method may not be essential to positive free logic, it has the unfortunate consequence (from the present point of view) of also validating:

14) Vulcan = Holmes.

The negative free logician may offer the following points in favour of not treating (**13**) as true: (i) to do so would be inconsistent with our intuitions about predication (a true predication must be a predication *of* some object); (ii) to do so is unnecessary, since all we need from the logic of identity can be obtained without allowing

that there are true instances of "$\alpha = \alpha$" with "α" empty (cf. Burge [1974]).

(4) Denial and scope

The topic is horses. There are mares and foals, Arabs and Palaminos, greys and chestnuts. You say: "And winged and wingless. After all, Pegasus was a winged horse." I disagree, and so wish to deny what you said. I say

15) No: Pegasus was not a winged horse.

This sounds unnatural. This is consonant with Fregean free logic, which sees (**15**) as neither true nor false, and so as incapable of expressing a truth fit to disabuse someone who has been deceived by a myth. The unnaturalness of (**15**) raises a prima facie problem for negative free logic, according to which "Pegasus is a winged horse" should be false and so its negation should be straightforwardly true. Yet (**15**) does not strike many people as a straightforward truth.

For the negative free logician, a sentence containing a name and a sign for negation raises the possibility of significant scope distinctions. (**15**) is true only if the "not" has wide scope. If it had narrow scope, (**15**) would be attempting to make a negative predication of Pegasus, to ascribe to him the property of being not a winged horse. This cannot be true, on the negative free logical view, for a true predication needs an object to be a predication of. Perhaps, then, the problem with (15) is that it doesn't clearly deliver the wide-scope reading: it is not clearly a denial of "Pegasus was a winged horse". Yet however much we try to make the wide-scope reading salient (for example, by saying "It is not the case that Pegasus was a winged horse"), we do not get a sentence which seems straightforwardly true. The negative free logician requires an explanation of this.

I think the explanation is that special conventions govern the ways in which we should disabuse those we take to have been seduced by myth or fiction. We should say something like: "Pegasus is just a mythical creature". That's why sentences like (**15**), which are indeed true according to negative free logic, don't sound natural or appropriate.

More distinctive data favouring negative free logic come from cases in which fact and fiction are mixed. Utterances of the following (in obvious kinds of factual context) seem unequivocally true:

That man's not Odysseus.
That foal wasn't sired by Pegasus.

There may be more than one way to systematize such data, but Fregean free logic is not a serious starter, and negative free logic handles everything fine, including scope distinctions. Using square brackets in an obvious way to indicate scope, candidate formalizations include:

[a] [b] ¬ Rab
[a] ¬ [b] Rab
¬ [a] [b] Rab.

Presumably the second is the most plausible; and this explains the less satisfactory status of

Odysseus isn't that man.
Pegasus didn't sire that foal.

These sound unnatural, at the very least. If they contain occurrences of empty names with wide scopes relative to the negation sign (as marked by the formalization [a] [b] ¬ Rab), they are false according to negative free logic, and this would explain our disinclination to assert them.

Logic has various purposes, like developing a foundation for mathematics, or providing resources for computer science. To the extent that logic aims to reflect ordinary reasoning, and the ordinary language in which we reason, some form of free logic seems to me preferable; and of the versions discussed, negative free logic seems closest to ordinary language.

Universally free logic rejects (**2**); that is, its definition of validity makes room for the empty domain. The primary motivation is that the interpretations involved in the definition of validity are supposed to represent all the logical possibilities, but it does not appear to be logically impossible that there should be nothing. Hence one does not do justice to genuine validity if one defines it in terms of "all" interpretations, yet excludes interpretations whose domain is empty.

The allegedly unsatisfactory view would show up in connection with such truths as:

16) $\forall xFx \models_{\mathbf{Q}} \exists xFx$.

With respect to an empty domain, one expects the premise to be true, given the equivalence between $\forall$ and $\neg\exists\neg$: there is no counterexample to the universal quantification. Yet, intuitively, the conclusion should be false with respect to the empty domain, for there is nothing in it which is any way at all, let alone something which is *F*.

As things stand, there are, by explicit stipulation, no **Q**-interpretations whose domain is empty, so it would be nonsense to say that (**16**) was falsified by the empty domain. However, let us adjust the definitions in **Q** in the direction of the present line of thought. Once we remove the stipulation that an interpretation requires a non-empty domain we must adjust the rules of interpretation. For example, (**2.1**) requires that any interpretation assign to each name-letter an object in its domain. This is impossible if the domain is empty. One global and systematic option is to rephrase to something like: "any interpretation assigns to each letter either nothing or else an object from its domain". This means that we no longer think of interpretations as (complete) functions (in the mathematical sense): it may be that nothing is identical to $i(\alpha)$.

On a natural reading of the **Q**-clause for the truth-upon-an-interpretation of atoms, it will produce unwanted results (**2.2**v):

> an atom is true upon an interpretation whose domain is *D* iff the sequence of the object(s) from *D* assigned to the name-letter(s) by the interpretation belongs to the set it assigns to the predicate or predicate-letter; it is false iff the sequence does not belong to the set.

An interpretation over the empty domain will not assign anything to a name-letter; in this case, a natural way to read "the sequence of the object(s) assigned to the name-letter(s) by the interpretation" is as denoting the null sequence, the sequence with no members. Since this belongs to every set of sequences, the upshot would be that any atom would be true upon any interpretation over the empty domain. No theorist would want this, and it is the antithesis of the result desired by a negative free logician.

We could revise (**2.2v**) as follows:

17) an atom is true upon an interpretation over D iff it assigns an object from D to each name-letter in the atom, and the sequence of the object(s) assigned to the name-letter(s) belongs to the set the interpretation assigns to the predicate or predicate-letter; otherwise it is false.

This makes the conclusion of (**16**) false upon an interpretation over the empty domain. However, given the existing **Q**-rule for "∀", it also has the unwanted result of making the premise false: its truth would require the truth of some atom, say "$F\alpha$" upon all α-variants, but these interpretations must also be over the empty domain, so they will not assign anything to α, so they will none of them meet the first part of the truth-upon-an-interpretation condition of (**16**). This result is unwanted, for intuitively we wanted there to be a counterexample to the validity of an argument from ∀ to ∃, and this means an interpretation (over some domain) which makes the ∀-sentence true and the ∃-sentence false, and this we have not yet found.

We can work our way towards appropriate adjustments of **Q** by reflecting on the intuitive reason for thinking of an ∀-sentence, for example $\forall xFx$, as true with respect to the empty domain: there is no counterexample. This means that there is no object (in the domain) to falsify $\forall xFx$ by verifying $\neg F\alpha$. This suggests the following modification of the quantifier rule (**2.2vi**):

18) $\forall vX$ is true upon an interpretation, i, whose domain is D iff there is no object z, in D and no n-variant i' of i such that $i'(n) = z$ and $i'(X\frac{n}{v}) = $ F.

(Cf. Bostock [1997], p. 368.) On this rule, being an n-variant is not a reflexive relation: the n-variants of an interpretation which assigns nothing to n will not include the interpretation itself. When coupled with an analogous rule for existential quantification, (**18**) ensures the failure of the validity of the argument from ∀ to ∃, for it ensures that ∀-sentences are true upon interpretations over empty domains.

Ex. 4.48 Provide an existential quantifier rule that is analogous to (**20.18**).

Even though there are free logics which allow the empty domain without allowing for interpretations which assign nothing to some or all name-letters, the way in which the empty domain has been allowed for here *does* allow for interpretations which assign nothing to name-letters. It therefore provides interpretations which would count as intended interpretations of English sentences containing empty names (at least if these are known to be empty). (For an argument that if one allows the empty domain one should allow empty names, see Bostock [1997], pp. 348–51.)

One moral to be drawn from the discussion of free logics is the reminder that there are coherent alternatives to **Q**. Almost every supposedly inviolable "logical law" has been challenged: dialetheic logic holds that there are true contradictions; some logics repudiate modus ponens; intuitionistic logic does not affirm double negation (the inference from "not not-A" to "A"); some logics deny the validity of the inference from not-A, together with A or B, to B. The choice among the alternatives is not to be made on the grounds that "logic" demands, for example, the validity of "Something exists". Rather, it has to be made upon the basis of philosophical arguments: we look to these to say what a correct logic should validate, and what it should not. Some departures from the classical logic of **Q** are larger than others, and some are better motivated than others. Free logic is a small departure and scores highly on motivating reasons.

21 Depth

A **Q**-formalization may be adequate, yet not reveal all the logical structure of the English. Thus

1) $F\alpha$

is an adequate **Q**-formalization of

2) John is not happy

(with "F" corresponding to "is not happy"). The truth conditions of (**2**) coincide with the truth-upon-an-intended-interpretation conditions of (**1**). Yet (**1**) does not reflect the fact that (**2**) contains an occurrence of a logical constant. We shall say that, relative to (**2**), (**1**) is *a less deep* **Q**-*formalization* than

3) $\neg F\alpha$

with "F" corresponding to "is happy".

Sometimes lack of depth in a formalization may prevent the validity of an argument becoming manifest. For example, the following is valid:

4) John will enjoy any book about cosmology. *The First Three Seconds* is about cosmology. Therefore John will enjoy *The First Three Seconds*.

The following meets the standard for being an adequate **Q**-formalization (truth-upon-an-intended-interpretation conditions match truth conditions):

5) $F\alpha\beta$, $G\gamma\beta$; $H\alpha\gamma$

with "Fxy" corresponding to "x will enjoy any book about y", "Gxy" to "x is a book about y", "Hxy" to "x will enjoy y", "α" to "John", "β" to "cosmology" and "γ" to "*The First Three Seconds*". Yet (**5**) is plainly **Q**-invalid.

Lack of depth may not prevent the manifestation of validity. The following is **Q**-valid and an adequate formalization of (**4**):

6) $\forall x(Kx \rightarrow H\alpha x)$; $K\gamma$; $H\alpha\gamma$

with "K" corresponding to "is a book about cosmology" and the other correspondences as before. The following is also adequate but deeper:

7) $\forall x(Gx\beta \rightarrow H\alpha x)$, $G\gamma\beta$; $H\alpha\gamma$

(correspondences as before). If X and Y are adequate formalizations of A, and Y is deeper than X, and X is **Q**-valid, then Y is also **Q**-valid.

How far should depth go? One extreme idea, to be explored in chapter **6.2**, is that wherever there is validity in the English, then some adequate **Q**-formalization is valid. Putting this idea into practice would require one, for example, to formalize the argument

8) Tom is a bachelor. Therefore Tom is not married

along the lines of

9) $F\alpha$ & $\neg G\alpha$; $\neg G\alpha$

(with "F" corresponding to "is a man" and "G" to "is married"). And it would require one to formalize

10) Socrates is mortal. Therefore he is or will be dead

along the lines

11) $\exists x(Fx$ & $G\alpha x)$; $\exists x(Fx$ & $G\alpha x)$

(with "F" corresponding to "is a time", "Gxy" to "x dies at y" and "α" to "Socrates").

As in the case of **P**, that an adequate formalization of an argument is not **Q**-valid does not in itself enable us to infer anything about the validity of the argument. Perhaps a deeper **Q**-formalization would reveal it as valid – and valid in virtue of its **Q**-logical form. Or perhaps it is valid, but not in virtue of its **Q**-logical form (*pace* the proponents of the extreme idea just mooted).

Standard practice shrinks not at all from introducing in a **Q**-formalization logical constants that do not visibly appear in the English. Examples are the formalizations of two-term universal and existential quantifications (e.g. "All (some) men are happy"), the former seen as introducing an occurrence of "→", the latter an occurrence of "&". Indeed, it is not unknown for practitioners to go further, and introduce predicate-letters not corresponding to predicates visible in the English. Some versions of Davidson's theory of adverbs, for example, see a sentence like "John runs" as formalizable by a **Q**-sentence containing a predicate-letter which, on an intended interpretation, will be assigned the set of runs. No such predicate is visible in the English, for "runs" appears to be true of people rather than of runs.

22 From Q-validity to validity

If an adequate formalization of an English argument is not **Q**-valid, what can be inferred about the validity of the English? Answer (given in the previous section): Nothing.

Suppose, however, that an English argument is adequately formalized by a **Q**-valid one. The exercise would have been pointless unless we could infer that the English argument is valid – valid, as we shall say, in virtue of its **Q**-logical form. How can this inference be justified?

We have to show is that if a **Q**-argument ϕ is **Q**-valid and is an *adequate* formalization of an English argument ψ, then ψ is valid.

The adequacy of the formalization ensures that, where ϕ' is the argument recovered from ϕ by applying the relevant correspondences, ϕ' says the same as ψ. The notion of "saying the same" that is required here is that the sentences of ϕ' should have the same truth conditions as the corresponding sentences in ψ. Validity can be defined in terms of truth conditions, and if two arguments are related as ϕ and ϕ' then, necessarily, both or neither are valid. So it will be enough if we can show that if ϕ is **Q**-valid then ϕ' is valid.

A necessary condition for there being any adequate formalizations is that the **Q**-operators make the same contribution to truth conditions as the corresponding English expressions. Let us suspend any doubts on this score for the moment. We will use a phrase like "corresponding name", "corresponding predicate", etc., to refer to a name in ϕ' to which, by the correspondence schema, a name-letter in ϕ corresponds; likewise for the other cases. The following argument establishes what is needed:

1) (i) Suppose ϕ is **Q**-valid
(ii) Then every interpretation upon which the premises of ϕ are true is one upon which the conclusion is true.
(iii) Hence whatever the corresponding names in ϕ' refer to, whatever the corresponding predicates in ϕ' are true of, and whatever the truth values of the corresponding sentences, if the premises of ϕ' are true, so is the conclusion.
(iv) Hence, necessarily, if the premises are true so is the conclusion.

The step from (ii) to (iii) requires that every English expression in ϕ' corresponding to a constant in ϕ express the same as the **Q**-constant.

The step from (iii) to (iv) requires justification, for the former makes no explicit mention of logical necessity, yet the latter does. (iv) could be re-expressed as the claim that all logical possibilities which verify the premises also verify the conclusion. To reach this from (iii) we need to suppose that the various **Q**-interpretations, by taking into account every possible assignment to the letters, run through all the logical possibilities. This in turn requires the following (to be refined in chapter **5.1** into the conception of extensionality):

2) the truth of a sentence which is adequately **Q**-formalizable of necessity depends on nothing except the reference of any corresponding names it contains, what any corresponding predicates are true of, and whether any corresponding sentences are true or false.

If all logical possibilities for what the corresponding names refer to, what the corresponding predicates are true of, and what truth value the corresponding sentences have, are ones upon which the conclusion is true if the premises are, as the **Q**-validity of ϕ in effect assures us is the case, then we can legitimately infer that, necessarily, if the premises are true so is the conclusion.

$\models_Q$ shares the formal properties of $\models$. Were this not so, its claim to give even a partial representation of $\models$ would be undermined.

Ex. 4.49 Using the definition of a **Q-interpretation**, establish the following:

(i) If $[X_1, \ldots X_n \models_Q Z]$ then $[X_1, \ldots X_n, Y \models_Q Z]$ whatever Y may be. (Compare (**1.6.5**).)

(ii) If $[X_1, \ldots X_n \models_Q Z]$ and $[Y_1, \ldots Y_k, Z \models_Q W]$, then $[X_1, \ldots X_n, Y_1, \ldots Y_k \models W]$. (Compare (**1.6.6**).)

(iii) If Z is among the $X_1, \ldots X_n$, then $[X_1, \ldots X_n \models_Q Z]$. (Compare (**1.6.7**).)

(iv) If $[(X_1, \ldots X_n) \models_Q]$ then $[X_1, \ldots X_n \models_Q Y]$, whatever Y may be. (Compare (**1.6.8**).)

(v) If $[\models_Q X]$ then $[Y_1, \ldots Y_n \models_Q X]$, whatever $Y_1, \ldots Y_n$ may be. (Compare (**1.6.9**).)

Bibliographical notes

§§1–4

Gamut [1991] (vol. 1) treats most of the topics of this and the previous chapter from the perspective of understanding natural language. Strawson [1952], chs 5 and 6, provides useful comparisons between Q-quantifiers and English; see also McCawley [1981]. For a claim that Q-quantifiers cannot represent everything we want to say in English by the corresponding quantifiers, see Quine [1970], pp. 89–91 (branched quantifiers). For the claim that other quantifiers (e.g. "most") cannot be represented in the style of the Q-quantifiers, see below, §**17**.

§5

For more detailed discussions of adjectives, see Platts [1979], ch. 7, and Kamp [1975].

§6

For Davidson's proposal see his [1967b] and [1970a]. For discussion, see Platts [1979], pp. 190–201 and Taylor [1985]. For a different approach, see Clark [1970] and Parsons [1972]. For a proposal which accepts much of the spirit of Davidson's, while rejecting the full appropriateness of Q-formalization, see Wiggins [1985].

§§7 and 12

Mill [1879], book 1, held that names and descriptions differ. Frege [1892b], p. 59, footnote, suggests that the sense of a name might be given by a definite description, but this is not a consequence of his general doctrines on the subject (see Dummett [1973], pp. 97–8: "there is nothing in what [Frege] says to warrant the conclusion that the sense of a proper name is always the sense of some complex description"), and might even be inconsistent with them (Evans [1982], pp. 22–38). None the less, the theory that names are descriptions is frequently attributed to Frege, for example by Kripke [1972], esp. p. 27, and Searle [1958]. Russell thought there are two kinds of name: "logically" proper names formalizable by name-letters, but in natural language consisting in only "this" and "that"; and "ordinary" proper names, like "Aristotle", which

are "really" abbreviations for descriptions. See Russell [1912], ch. 5, and [1918], pp. 200 ff. Russell does not intend this as a doctrine about the meaning of names, if meaning is what is in common between speaker and hearer in communication, but rather as contributing to an account of "the thought in the mind of a person using a proper name correctly" ([1912], p. 29); cf. Sainsbury [1993].

A gentle introduction to opposing views is Searle [1958]. Kripke [1972] attacks description theories of names, and his arguments have been widely discussed, for example by Dummett [1973], appendix to ch. 5. Some subtle distinctions are brought to bear by McDowell [1977]. See also Platts [1979], ch. 6, Linsky [1977], Davies [1981], esp. pp. 103 ff., Pollock [1982], chs 2–3, and Evans [1982], chs 1–3 and 11.

§8

See Quine [1960], §24.

§9

This standard treatment of numeral adjectives derives from Frege [1884], esp. §55 ff.

§10

Russell's theory of descriptions was first presented in Russell [1905], along with a general theory of quantification. A much clearer presentation, detached from the general theory of quantification, is in Russell [1919], ch. 16. A classic criticism is by Strawson (see his [1950], and his [1952], pp. 184–90), and a source of much debate is Donnellan [1966]. For further discussion, see Peacocke [1975], Sainsbury [1979], ch. 3, Davies [1981], ch. 7, and Pollock [1982], ch. 4. The best single source is Neale [1990].

§11

For Russell's account, together with his attack on Meinong, see Russell [1905] and [1918], lecture 5. For Meinong's theory see Meinong [1904]. For a philosophical discussion, see Lambert [1983], ch. 5. For a formal defence of a Meinongian position, see Parsons [1974]. For a philosophical plea for non-existents, see Yourgrau [1987].

For a brilliant account of negative singular existential truths which treats "exists" as a predicate, see Evans [1982], ch. 10.

§16

For Davidson's proposal, see Davidson [1969]. For discussion, see Platts [1979], ch. 5.

§17

See Platts [1979], pp. 100–6, Davies [1981], esp. pp. 123 ff., and Wiggins [1980a].

§18

A classic text on substitutional quantification is Kripke [1976]. A gentler introduction is by Quine [1969a]. See also Davies [1981], pp. 142 ff.

§19

For philosophical discussion, see Strawson [1974], Davies [1981], pp. 136 ff, and, for a sceptical view, Quine [1953b] and [1970], pp. 66–8. Predicate quantification may not be well motivated by considering natural English idiom, but there is certainly a case for it in the formalization of mathematics. Both Russell [1908] and Frege [1879], in their early attempts at formalizing mathematics, found it quite natural to use predicate quantifiers. A justification is provided by Boolos [1975], and Boolos and Jeffrey [1974 (1989)] provide a classic account of second order logic.

§20

A brief history of free logic is given in Lambert [1991a], editor's introduction, and the collection itself contains most of the classic papers. A longer history is provided by Bencivenga [1986]. My introduction to the topic was through Burge [1974], who argues for negative free logic in a truth theoretic setting. See also Schock [1968], Lambert [1965] and [1983], esp. ch. 5, and references therein. Bostock [1997] introduces the subject briefly, clearly and with philosophical insight. For an interesting application, see Evans [1979]. For the empty domain, see Quine [1952], p. 98, and Bostock [1997]. For a philosophical discussion of whether "Something exists" is a logical truth, see Cohen [1962], §33.

5

Necessity

If you are a physicalist, you may wish to assert not merely that everything is physical – which is consistent with this important fact being accidental – but also, more strongly:

1) Necessarily, everything is physical.

This chapter explores how one might augment **Q** so as to capture the distinctive contribution of English expressions like "necessarily". Superficially, at least, (**1**) appears to be composed of "everything is physical", dominated by the non-truth functional sentence operator "necessarily". One option is to take appearances at face value, and add suitable non-truth functional sentence operators to **Q** (stipulating that these be counted among the logical constants). This augmented language will be called **QN**.

(**1**) appears to be equivalent to

2) It is necessarily the case that everything is physical.

This seems to ascribe something – the property of being necessarily so – to a proposition. Possibility and necessity are interdefinable. A proposition is possible iff it is not impossible, iff its negation is not necessary. Possibility and necessity are called *modal notions*, and studying them is part of the study of *modality*.

Modal notions surface in idioms other than "necessarily" and "possibly", for example in "has to" and "must" as used in sentences like the following:

3) You have to make adequate financial provision for your children.

4) You must leave now or you'll miss your plane.
5) If you press down on one end of a rigid rod, freely balanced at its centre, the other end must go up.
6) This just has to be a herring.

These examples exploit different standards or criteria for the necessity they invoke: (3) is naturally heard as invoking moral necessity, (4) prudential necessity, (5) natural necessity and (6) epistemic necessity. We will discuss necessity of the broadest kind, so-called logical or metaphysical necessity. We make no attempt to give any account of what that is. It is exemplified in (1), and there will be many subsequent examples.

1 Adding "□" to P

We approach **QN** by an intermediate stage, a language, to be called **PN**, obtained from **P** by adding a new sentence operator, "□", pronounced "box", with the intended meaning of "necessarily". **PN** is called a "propositional modal language".

Syntactically, "□" is just like "¬": if X is a **PN**-sentence, so is $\Box X$. A sentence of the form $\Box X$ is called a "necessitation".

Intuitively, we want $\Box A$ to be true iff A has to be true, iff A is true however things might have been or may be, iff A is true in all possibilities, iff A is true at every possible world. A possible world is, as David Lewis [1973b] put it, a way our world could have been. Possible worlds are "maximal" in the following sense: for any possible world, every sentence is either true at it or false at it. To say that snow could have been black is to say that snow *is* black at some possible world. All modal auxiliaries (like "could" and "must") and modal adverbs (like "possibly" and "necessarily") are removed from the main body of sentences (in the example just given, "could" becomes plain "is"), and their contribution is regimented into locutions like "at some possible world".

Ex. 5.1 Assuming that the sentences to which "□" can be applied are as rich as the sentences of English, show that "□" is a non-truth functional sentence connective.

The maximality of possible worlds makes them different from possibilities as we ordinarily conceive them. If we think about the possibility of rain, nothing in the possibility determines whether, in the possibility in question, kangaroos have tails. But at each possible world either "kangaroos have tails" is true, or else it is false.

We can use possible worlds to give semantics for **PN**. An *interpretation of PN* specifies, for each **PN**-letter, at which of the possible worlds it is true and at which it is false. This corresponds to the intuitive idea that the meaning of a sentence determines, among other things, in what circumstances it would be true, and in what circumstances it would be false. The interpretation rules likewise require that truth be relativized to a world as well as to an interpretation. Our rules from chapter **2.1** have to be rewritten; one way to do so is as follows:

1) For any set of possible worlds, W, any world, w, in W, and any interpretation i of **PN**,
   $\neg X$ is true at w upon i iff X is false at w upon i;
   $(X \,\&\, Y)$ is true at w upon i iff X is true at w upon i and Y is true at w upon i;
   $(X \vee Y)$ is true at w upon i iff X is true at w upon i or Y is true at w upon i;
   $(X \rightarrow Y)$ is true at w upon i iff X is false at w upon i or Y is true at w upon i;
   $(X \leftrightarrow Y)$ is true at w upon i if either both X and Y are true at w upon i or both X and Y are false at w upon i.
   $\Box X$ is true at w upon i iff for every world w' in W, X is true at w' upon i.

We can add a symbol for "possibly" by the following definition:

2) $\Diamond X =_{df} \neg\Box\neg X$.

"◊" is pronounced "diamond". Validity is defined as follows:

3) $X_1, \ldots X_n \models_{PN} Y$ iff: for all interpretations, i, and all sets of worlds, W, if, for any world w in W, all of $X_1, \ldots X_n$ are true at w upon i, Y is true at w upon i.

These semantics determine a system of propositional modal logic known as S5 (the terminology derives from Lewis and Langford [1932]). In the rest of this section we will explore its properties. Towards the end of the section I will briefly indicate how the semantics can be varied to give alternative propositional modal logics.

Consider the following argument:

4) Possibly, no one will come. Possibly, many people will come. Therefore it's possible that both no one and many people will come.

The argument is plainly invalid, rather in the way that

5) Some numbers are odd and some are even, therefore some numbers are both odd and even

is invalid.

With "p" corresponding to "no one will come" and "q" to "many people will come", the obvious **PN**-formalization of (4) is

6) $\Diamond p, \Diamond q; \Diamond(p \ \& \ q)$.

To show that (6) is **PN**-invalid, we need to find an interpretation and a set of worlds such that the premises are true at a world in the set upon the interpretation, whereas the conclusion is false at that world upon the interpretation. The definition of "$\Diamond$", together with the interpretation rule for "$\Box$", gives the derived interpretation rule for "$\Diamond$":

7) $\Diamond X$ is true at w upon i iff for some world w' in W, X is true at w' upon i.

Ex. 5.2* Show how (1.7) is derived from the interpretation rule for "$\Box$" together with the definition of "$\neg$".

Let i assign truth at w_1 to "p", falsehood at w_1 to "q", falsehood at w_2 to "p", truth at w_2 to "q"; and falsehood at all other worlds in W to both "p" and "q". This can be set out as in table 5.2, in which the assignment of falsehood to a letter is represented as the assignment of truth to its negation. "$\Diamond p$" is true upon i, since there is a world, viz. w_1, at which "p" is true upon i; and at w_1, "$\Diamond q$" is true upon i, since there is a world, viz. w_2, at which "q" is true upon i; but at w_1, "$\Diamond(p \ \& \ q)$" is false upon i, since there is no world at which "$p \ \& \ q$" is true

Table 5.1 Interpretation i

w_1	w_2	w_3	...
p	$\neg p$	$\neg p$	$\neg p$
$\neg q$	q	$\neg q$	$\neg q$

Table 5.2 Further features of interpretation i

w_1	w_2	w_3	...
p	$\neg p$	$\neg p$	$\neg p$
$\neg q$	q	$\neg q$	$\neg q$
$\Diamond p$	$\neg(p \ \& \ q)$	$\neg(p \ \& \ q)$	$\neg(p \ \& \ q)$
$\Diamond q$			
$\neg(p \ \& \ q)$			
$\neg\Diamond(p \ \& \ q)$			

upon i. Hence i is a counterexample to the validity of (**6**). The essential point is structurally like what needs to be said to identify the fallacy of (**5**): that there is a world at which "p" is true and one at which "q" is true does not entail that there is one at which both are true. Table 5.1 can be extended to show the relevant further features of i, as determined by (2) and (7) (see table 5.2). The justification for adding "$\Diamond p$" to the w_1 column is that "p" occurs elsewhere in the table; the justification for adding "$\neg\Diamond(p \ \& \ q)$" is that "$(p \ \& \ q)$" occurs nowhere in the table.

Ex. 5.3 (a) Draw a diagram illustrating a set of worlds and an interpretation which establish the falsehood of:

$\Diamond p, \Diamond(p \rightarrow q) \models_{PN} \Diamond q$.

(b) Say whether the following is true, and argue informally for your view:

$\Box p, \Diamond(p \rightarrow q) \models_{PN} \Diamond q$.

The following is also invalid:

8) $p, \Box(p \rightarrow q); \Box q$.

Table 5.3

w_1	w_2
p	$\neg p$
q	$\neg q$
$p \rightarrow q$	$p \rightarrow q$
$\Box(p \rightarrow q)$	
$\neg\Box q$	

Ex. 5.4 Give an English argument of the form of (**1.8**) in which the premises are plainly true and the conclusion plainly false.

This can be established by appealing to a set of worlds, W, containing just w_1 and w_2, as shown in table 5.3. The justification for adding "$\neg\Box q$" to w_1 is that "q" does not appear in every world in the table, corresponding to the fact that at some world "q" is false upon the interpretation represented.

In contrast to (8), the following is a cardinal principle of modal logic:

9) $\Box X, \Box(X \rightarrow Y) \models_{\mathbf{PN}} \Box Y.$

The truth of (**9**) can be established informally by reflecting that an interpretation upon which both premises are true at some world in some arbitrary set of worlds, W, assigns truth at *every* world in W to both X and $X \rightarrow Y$, and so, by the interpretation rule for "$\rightarrow$", must also assign truth to Y at every world in W. Hence for any set of worlds, W, and any world in W, any interpretation upon which the premises are true at that world is one upon which the conclusion is true at that world.

Since the interpretation rules of **P** are mirrored in those for **PN**, we have

10) If $[\models_{\mathbf{P}} X]$, then $[\models_{\mathbf{PN}} X]$.

In other words, any **P**-valid sentence is also a **PN**-valid sentence. We also have the stronger

11) If $[\models_{\mathbf{P}} X]$, then $[\models_{\mathbf{PN}} \Box X]$.

This says that the result of prefixing a **P**-valid sentence by box is **PN**-valid. It reflects the thought that a valid **P**-sentence corresponds to a necessary truth. If $\models_{\mathbf{P}}$ is enough for $\models$, in the way argued in chapter **2.10**, and box corresponds to "it is logically necessary that", then (**11**) must hold. For "$\models A$" is equivalent to "it is logically necessary that A" (cf. chapter **1.6**).

The intuitively natural thought that what is necessarily true is true is verified by the interpretation rule for "□". It can be expressed by the truth of the generalization

12) $\Box X \models_{\mathbf{PN}} X.$

This yields the stronger

13) $\models_{\mathbf{PN}} \Box X \rightarrow X.$

Ex. 5.5 The move from (**1.12**) to (**1.13**) applies the "Deduction Theorem" to **PN**. We saw that an analogue of it for "if . . . then" and $\models$ can be put to a controversial use (**1.8.14**). The present application is not controversial, as you can establish by using the interpretation rules for **PN** to show:
If $[X, Y \models_{\mathbf{PN}} Z]$ then $[X, \models_{\mathbf{PN}} Y \rightarrow Z]$.

By the definition of "◊", this delivers

14) $\models_{\mathbf{PN}} \neg X \rightarrow \Diamond \neg X$

and (after some intermediate steps)

15) $\models_{\mathbf{PN}} X \rightarrow \Diamond X.$

Ex. 5.6 Use (**1.14**) to establish (**1.15**). Remember that "X" and "Y" are metalinguistic variables, so you can replace them by any **PN**-formula.

This accords with the intuition that anything actual is possible: what is in fact true can be true.

More controversially, **PN** has it that

16) $\models_{PN} \Box X \rightarrow \Box\Box X.$

This says that any conditional is valid in **PN** if its antecedent starts with an occurrence of "□" and its consequent consists of the antecedent prefixed by an occurrence of "□". This corresponds to the claim that anything that is necessarily true is of necessity necessarily true. It is no accident which truths are necessary.

It is not entirely uncontroversial that (**16**) accurately reflects our intuitive views about necessity. Suppose, for example, that necessity is not an objective feature of the world, but is a product of human thinking. Suppose, further, that our patterns of thinking are determined by evolutionary pressures, but in a way that is not necessary (random mutations might be involved). Then we do not *necessarily* think as we do, so something which is in fact a necessary truth might not have been.

There is also in **PN** a stronger version of (**15**):

17) $\models_{PN} X \rightarrow \Box\Diamond X.$

An interpretation upon which X is true at a world, w, is one upon which $\Diamond X$ is true at that world and every other world, and so is one upon which $\Box\Diamond X$ is true at that world.

It is not uncontroversial that this correctly reflects our modal notions. You might agree that the fact that something is true guarantees that it is in fact possible, but not agree that it had to be possible. This is certainly correct for some restricted notions of necessity. For example, it is true (let us suppose) that you will catch the train, and this shows that it is possible for you to catch it. But it did not have to be possible: something could have happened to delay you, and then it would not have been possible for you to catch it (even though you will in fact catch it). So we have both "You will catch the train" and "It is not necessarily possible for you to catch the train", contrary to (**17**). The modality here is related to time, and it would seem that the claim that it did not have to be possible for you to catch the train relates to an earlier time than the claims that you will catch it and that it is possible for you to catch it.

A further fact about **PN** which might be regarded as only dubiously appropriate to our intuitions about modality is that

18) $\models_{PN} \Diamond\Diamond X \rightarrow X.$

This holds because for any world w, any interpretation upon which the antecedent is true at w is one upon which $\Diamond X$ is true at some world, and thus one upon which X is true at some world; this last condition is enough to ensure that $\Diamond X$ is true at w.

Do we intuitively believe that if something is possibly possible it is actually possible? A case for a negative answer is as follows (cf. Salmon [1981]). This table on which I am writing could not have been made out of entirely different parts, for a table made out of entirely different parts would not have been this table. However, it could have been made out of slightly different parts. We can put this as follows. Let α_1 be the collection of parts out of which my table was in fact made, and let α_2 be a slightly different collection. Then it seems plausible to say that my table could have been made out of α_2. Let α_3 be a collection of table parts differing slightly from α_2 but significantly from α_1. It seems plausible to say that *if* my table had been made of α_2, as I allow is possible, then it could have been made from α_3, but that as things are it could not have been made from α_3 since these parts are too different from the actual parts. We might express this by the combination of claims, inconsistent with (**18**):

19) $\Diamond\Diamond$ (my table is constructed out of α_3);
$\neg\Diamond$ (my table is constructed out of α_3).

PN has an interesting property, called *modal collapse*: given a non-modal sentence, X, that is, one containing no boxes or diamonds, there are only two non-equivalent fresh sentences you can form from it just by adding modal operators. They are, simply, $\Box X$ and $\Diamond X$. However many other boxes and diamonds you may stick in front of X, in whatever order, the result will be equivalent to one of these two sentences. To take two examples: (**12**) and (**16**) together ensure that

20) $\models_{PN} \Box\Box X \leftrightarrow \Box X$

which means that any pair of boxes can be collapsed to a single box. In addition,

Table 5.4 Failure of transitivity

w_1	$\Rightarrow$	w_2	$\Rightarrow$	w_3
p		p		$\neg p$
$\Box p$		$\neg\Box p$		
$\neg\Box\Box p$				

21) $\models_{\mathbf{PN}} \Box\Diamond X \leftrightarrow \Diamond X$,

which shows that box followed by diamond can be collapsed to just a diamond. Those who think that some **PN**-validities are inappropriate to our ordinary conception of modality can devise weaker modal languages by restricting in various ways the worlds which are relevant to truth upon an interpretation. The standard way to do this is to introduce a relation between worlds: the *accessibility* relation, call it R. Truth at a world depends at most upon how things are at R-related worlds. So, for example, the last clause of (**1.1**) will be revised to

22) $\Box X$ is true at w upon i iff X is true upon i at every w' such that $Rw'w$.

The original rule for box is equivalent to letting R hold universally, between arbitrary pairs of worlds. Restrictions on R reduce the class of **PN**-valid sentences. For example, (**13**) ($\models_{\mathbf{PN}} \Box X \rightarrow X$) holds only if R is reflexive (that is, only if every world is R-related to itself), so anyone wishing not to have this formula as valid could stipulate that R is to be non-reflexive (that is, that there is at least one world, w, such that not-Rww).

Instances of (**16**) ($\Box X \rightarrow \Box\Box X$) hold only if R is transitive (that is, only if for arbitrary worlds, if Rw_1w_2 and Rw_2w_3 then Rw_1w_3). Using open arrows to indicate the R-relation between worlds, table 5.4 shows a non-transitive R and a counterexample to (**16**). If w_1 were R-related to w_3, "$\Box p$" would not be true at w_1. In the case of purely logical modality, it is doubtful whether we have any independent intuitions about what sort of relation R should be. Rather, we shape R to capture our intuitions about what logical principles should hold. However, for some restricted conceptions of necessity, it may be that we can have

intuitions which relate directly to R. For instance, it might be suggested that what is morally necessary is what obtains in all morally perfect worlds, worlds in which what is morally required obtains. Then a natural account of what it would be for Rw_1w_2 to hold is that the moral requirements that obtain at w_1 are satisfied at w_2. It is not likely that this would be a transitive relation.

Ex. 5.7 Say whether you think that the accessibility relation for moral necessity would be transitive, and justify your view.

2 Non-indicative and counterfactual conditionals

A conditional apparently involving the subjunctive mood, like (**2.4.35**) (If Oswald hadn't shot Kennedy, someone else would have), appears to involve something modal. It seems to suggest some kind of necessary connection between Oswald not shooting and someone else shooting. It invites us to consider a possible world in which Oswald did not shoot Kennedy and its truth seems to require that in such a world someone else shot Kennedy. In this section, I consider how far we might go in understanding such conditionals using the modal notions we now have to hand: box, diamond, and possible worlds and accessibility relations. The first issue is taxonomy: how should we characterize these "subjunctive" or "counterfactual" conditionals?

In chapter **2** a preliminary distinction was drawn between indicative conditionals and others, based on contrasting sentences like (**2.4.35**) with ones like (**2.4.34**) (If Oswald didn't shoot Kennedy, someone else did). An initial guide was that indicative conditionals are expressed by using the indicative mood, and the other kind by using the subjunctive mood. However, we mentioned that this might be a less than reliable guide, and that (**2.4.39**) (If John dies before Joan, she will inherit the lot) might best be classified as "subjunctive" rather than "indicative" even though standard grammar would count both verbs as indicative. We now need to re-examine this classification.

One way to divide conditionals into two classes is to contrast the "counterfactual" ones with the rest, where a counterfactual conditional is defined by this feature: one who asserts it thereby represents its antecedent as false (or, in the case in which the antecedent is subjunctive, and so does not have a truth value, represents the corresponding indicative as false). For many people, (**2.4.35**) is a conditional which counts as counterfactual by this test. However, some conditionals which resemble (**2.4.35**) – they have "had" in the antecedent and "would" in the consequent – can properly be used both by one wishing to represent the antecedent as false, and by one with no such wish. Suppose John has been accused of taking the bribe in the form of a large bundle of banknotes; suppose John is innocent and he and his lawyer believe, falsely, that the banknotes were discovered in the cleft of a certain apple tree. His lawyer can reasonably assert

1) If John had taken the bribe, he wouldn't have put the money in the apple tree,

representing the antecedent (or rather its corresponding indicative) as false: John did not take the bribe. But one of the detectives, thinking John to be guilty, and knowing that the money was not put in the apple tree, may also affirm (1) with the aim of showing that the fact that the money was not put in the apple tree does not count against John's guilt. The detective thus uses (1) not representing the antecedent to be false, but rather as part of a strategy designed to show that it is true.

Edgington ([1997], p. 99) puts the point in a strikingly general context. It is a general feature of empirical reasoning concerning some hypothesis, H, that we try to determine its truth by finding out how things *would* observably be it if *were* true. Such an enterprise cannot represent H as false in advance of the enquiry. We say "if H were true, we would observe such-and-such" with no commitment to the falsehood (or, for that matter, the truth) of H. There does not appear to be a category of conditionals the correct use of which requires representing the antecedent as false. This way of marking the contrast between the two "Oswald"-conditionals does not seem to yield a good general classification.

Ex. 5.8 Would the counterfactual/non-counterfactual contrast (as developed in the text) separate the two Oswald conditionals?

I shall use a more grammatical criterion, which starts by regarding the presence of the subjunctive in the antecedent as sufficient condition for belonging to one category, which I call that of "non-indicative" conditionals. This places (**2.4.35**) but not (**2.4.34**) in the category of non-indicatives, and also places in this category

2) If I were braver, I would stand up to her.

In this context, "were" is subjunctive. But, in English, the subjunctive is not always easy to identify and its use is not wholly systematic. It would be hard to believe that (**2**) deserves to be treated differently from

3) If I was braver, I would stand up to her.

One may think that "was" should be classified as a past tense rather than a subjunctive. Even if this is right, I would still wish to adopt a criterion which includes (**3**) among the non-indicatives. One way to achieve this is to stipulate that anything equivalent to a non-indicative is to count also as non-indicative. Another way is to say that having "would" in the consequent is a further sufficient condition for belonging to this category. I make both stipulations. There is a distinct reason for counting (**3**) as non-indicative. There is no genuine pastness about the use of "was" in (**3**), any more than there is in such a sentence as

4) I wish I was braver.

This gives us a reason for not counting this occurrence of "was" as a straightforward indicative past tense, and so for including (**3**) among the non-indicatives. (In some languages, words used like "was" in (**3**) are counted as belonging to the "conditional" tense.)

Ex. 5.9 What words would you use to express a wish about your bravery that related to some occasion now in the past?

There are some reasons to think that conditionals like

5) If John attends tomorrow, he will vote for the motion

should not be classified with indicative ones. Suppose that Mary uttered (**5**) on Monday, the meeting takes places on Tuesday, and on Wednesday I am asked to report what Mary told me. The following seems appropriate:

6) Mary said that if John had attended yesterday, he would have voted for the motion.

In my report I use a non-indicative conditional to report what was said by an apparently indicative one. The switch seems to be of the same kind as the switch from "tomorrow" to "yesterday". In further support of classifying (**5**) as non-indicative, it can be pointed out that "John attends tomorrow" is no ordinary present tense indicative. Used assertively on its own, it would not be strictly intelligible (unless, perhaps, it issued from the mouth of an almighty creator fixing the future behaviour of his creatures, a use which has no bearing on (**5**) – cf. Edgington [1997], p. 98). This kind of bogus present tense seems to be widespread in English conditionals in which "will" governs the main verb in the consequent. If these points are accepted, (**2.4.39**) (If John dies before Joan, she will inherit the lot) would get classified with non-indicatives.

The suggestion has been disputed. One's willingness to make the switch exemplified in (**5**) and (**6**) seems to vary with the kind of reason one has for accepting the seemingly indicative conditional. The switch is most natural if I believe (**5**) for the more obvious kinds of reason: I know John's views on the matter, he has made sincere expressions of intention, etc., but there may be some uncertainty about whether he will be able to be present. Contrast this with a less usual scenario: John is hostile to the motion and its proponents are planning to prevent him voting against it: they will either bribe him to vote for the motion (plan A), or, if he won't be bought, they will kidnap him and so prevent him from attending (plan B). Knowing all this, Mary can affirm (**5**) before the meeting: if John attends, then plan A has worked, so he will vote for the motion. But afterwards, if Mary knows John didn't attend, she will not be disposed to believe that if he had attended he would have voted for the motion: if he didn't attend, the likely explanation is that the villains adopted plan B, given which, if John had attended, if would be because he had escaped from his kidnappers, in which case he would have voted against. (Cf. Edgington [1991], Bennett [1995].)

For present purposes, this issue does not need to be resolved: it is enough that we have clear cases of non-indicatives, marked by the subjunctive or by the modal auxiliaries like "would" (in the consequent) and "had" (in the antecedent), exemplified by (**2.4.35**) (if Oswald hadn't . . .), and clear cases of indicatives like (**2.4.34**) (if Oswald didn't . . .). We can be neutral about the classification of conditionals like (**5**).

To say that there is a contrast between indicative and non-indicative conditionals is not to say that "if" is ambiguous. It could be that "if" functions in the same way in both kinds of conditional, and that the contrast is due to the contrasting indicative or non-indicative character of the verbs (cf. Woods [1997], p. 10).

With this rough account of our subject matter as non-indicative conditionals, we can return to the main question of whether we can give an account of them using our current resources (**P** supplemented by boxes and diamonds, and semantics using possible worlds and accessibility relations). The non-indicative "Oswald"-sentence (**2.4.35**) (If Oswald hadn't shot Kennedy, someone else would have) carries a suggestion of necessity, so an obvious first thought is to formalize it:

7) $\Box(\neg p \rightarrow q)$

with "p" corresponding to "Oswald shot Kennedy" and "q" to "Someone else shot Kennedy". (We will ignore the fact that **PN** obviously cannot express the quantificational structure.)

In favour of the formalization, it might be argued that all possible worlds divide into two classes: those in which Oswald did shoot Kennedy, and those in which he did not. "$\neg p \rightarrow q$" is true (upon an intended interpretation) at all members of the first class of worlds, which will include the actual world (on the assumption of Oswald's guilt), in virtue of the falsity of the antecedent; and if (**7**) is true, there is some appeal in the thought that "$\neg p \rightarrow q$" is true at every world in the second class as well. (**7**) may seem to say that any world in which Oswald didn't shoot Kennedy is one in which someone else did.

However, this is much too demanding relative to the English, which is not falsified by a world in which Kennedy never existed, nor by a world in which he never came before the public eye, and so was not a target for assassination, nor by a world in which the most stringent security precautions were invariably taken.

The defect may seem easy to remedy. Let us restrict the worlds that are relevant to the truth of a formalization of (**2.4.35**) to worlds that are similar, in certain contextually determined respects, to the actual one. For (**2.4.35**) the relevant worlds are restricted at least to those in which Kennedy exists, became President, and was not always subject to the most stringent security precautions.

No **PN**-operator carries this restriction to similar worlds. However, let us use "□→" to formalize non-indicative conditionals, introduce a special interpretation rule for them, and use "**PNS**" to stand for the result of adding this symbol and this rule to **PN**.

The syntax of "□→" (pronounced "box arrow") is that it takes two indicative sentences to form a sentence. (This formalization carries a substantive commitment: it presupposes that the non-indicative aspects of the conditional can be bundled into the box arrow, and that the overall content is fixed by *indicative* contents that are not explicitly expressed.) With the correspondences of (7), (**2.4.35**) will be formalized

8) $p \,\Box\!\!\rightarrow q$.

An interpretation rule for "□→" which would reflect the considerations so far adduced is

9) $X \,\Box\!\!\rightarrow Y$ is true at w upon i iff for every world w' in W which is *similar* to w, $X \rightarrow Y$ is true at w' upon i.

This rule will validate intuitively invalid inferences using non-indicative conditionals. For example, we have

10) $X \,\Box\!\!\rightarrow Y =_{PNS} (X \,\&\, Z) \,\Box\!\!\rightarrow Y$,

for arbitrary Z. The premise is true at a world, w, upon an interpretation, i, iff $X \rightarrow Y$ is true at all worlds in W which are similar to w. But any world at which $X \rightarrow Y$ is true upon i is one at which $(X \,\&\, Z) \rightarrow Y$ is true upon i. So the conclusion is true at w upon i. However, we do not accept that arguments like the following are valid (compare (**2.5.28**)):

11) If I had put sugar in this cup of coffee, it would have tasted good. So if I had put sugar and diesel oil in this cup of coffee, it would have tasted good.

Ex. 5.10 Assess the following argument:

The antecedent of (**2.11**) *is* false, taken quite literally, though what is intended is true, and could be more properly expressed along the lines: "If I had added sugar to this cup of coffee and everything else had remained as much the same as possible, then it would have tasted good."

Cf. Urbach [1988], p. 197.

This sort of case suggests that there cannot be a fixed standard of similarity of worlds, applicable to all non-indicative conditionals. For example, consider evaluations with respect to the actual world, and suppose that I put sugar but not diesel into my cup: if we set the standard of similarity high, then "(I put sugar and diesel oil in this cup of coffee) → (it tastes good)" is true at every similar world, since every world which is *very* similar to ours is one in which the antecedent is false. This high standard would formalize the conclusion of (**11**) as true upon an intended interpretation, which is not what we want. If we set the standard low, then "(I put sugar in this cup of coffee) → (it tastes good)" is false at some similar world, for example a world in which diesel is added as well as sugar. This low standard would formalize the premise of (**11**) as false upon an intended interpretation, which is not what we want. What is needed, therefore, is variability in how great the similarity must be, as a function of the content of the antecedent. The worlds we need to consider are not all those which are similar to the actual one, but only those at which the antecedent is true. I shall call such accounts of conditionals "*A*-world" accounts, and will consider two quite similar versions, one by Lewis and one by Stalnaker.

I start by showing that the desirability of an *A*-world account can be arrived at from a quite different direction. We can start by considering how we tell whether or not to accept a conditional. Following a discussion by Frank Ramsey [1929], Stalnaker proposed the following approach:

> First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency . . . finally consider whether or not the consequent is then true. (Stalnaker [1968], p. 44)

Although this speaks to the conditions under which we should *believe* a conditional, Stalnaker argued that we can use the same basic idea to arrive at *truth conditions* for conditionals:

Table 5.5

	Less similar ⟶	
World of evaluation, $w^\star$	w_1	w_2
$\neg X$	X	X
$\neg Y$	Y	$\neg Y$
$\neg Z$	$\neg Z$	Z

> Consider a possible world in which A is true, and which otherwise differs minimally from the actual world. *"If A, then B" is true (false) just in case B is true (false) in that possible world.* (p. 45)

This is an A-world account, for it implies that if a world is relevant to the truth of a conditional it is a world at which the antecedent is true. Converting the proposal to the present approach and terminology, we can express it thus:

12) $X \Box\!\!\rightarrow Y$ is true at w upon i iff Y is true upon i at a world maximally similar to w at which X is true upon i. (If there is no such world, that is, if X is impossible, $X \Box\!\!\rightarrow Y$ is true upon i.)[1]

The account is appealing for its claim to mirror our actual conditional thinking. It correctly classifies as invalid the pattern of reasoning exemplified by (**11**):

13) $X \Box\!\!\rightarrow Y; (X \,\&\, Z) \Box\!\!\rightarrow Y.$

Table 5.5 shows an interpretation in a structure of worlds upon which the premise is true and the conclusion false. In this structure, there are just three worlds: $w^\star$ (which we may think of as the actual world), which is the world of evaluation, and two others, both X-worlds (worlds at which X is true). Of these two, the one at which Y holds is more similar to $w^\star$ than is the other, so the structure verifies $X \Box\!\!\rightarrow Y$. However, it does not verify $(X \,\&\, Z) \Box\!\!\rightarrow Y$, for the most similar $(X \,\&\, Z)$-world to $w^\star$ is not a Y-world.

[1] This does not do justice to Stalnaker, who wants his account to apply to conditionals generally, rather than just to non-indicative ones.

Table 5.6

	Less similar ⟶	
World of evaluation, $w^\star$	w_1	w_2
$\neg X$	$\neg X$	X
$\neg Y$	Y	Y
$\neg Z$	Z	$\neg Z$

These semantics for $\Box\!\!\rightarrow$ make it non-transitive (in the sense of (**2.5.21**): the following argument pattern is not valid:

14) $X \Box\!\!\rightarrow Y, Y \Box\!\!\rightarrow Z; X \Box\!\!\rightarrow Z.$

This will be welcome if we find arguments like the following invalid:

15) If Hoover had been born a Russian, he would have been a communist.
If he had been a communist, he would have been a traitor.
So if he had been born a Russian, he would have been a traitor.

> **Ex. 5.11** What happens if the two premises of (**2.15**) are conjoined into a single premise? For a defence of transitivity for $\Box\!\!\rightarrow$ (the principle of which (**2.15**) is an instance), see Urbach [1988], p. 198.

Table 5.6 serves as counterexample. It makes vivid that only A-worlds are relevant to the truth of a $\Box\!\!\rightarrow$ sentence; and among A-worlds, only ones more similar to the world of evaluation.

The logic which results from these semantics coincides (in terms of what it says is valid) with the probabilistic logic of chapter **3**.

Can we assume that for each world, and in particular for the actual world, there is a unique "maximally similar" world? For example, should we say that a world in which Verdi and Bizet were both French is more similar or less similar to the actual world than one in which they were both Italian? The question would appear to need an answer (on Stalnaker's semantics) if we are to evaluate conditionals like

16) If Bizet and Verdi had been compatriots, Bizet would have been Italian.

17) If Bizet and Verdi had been compatriots, Verdi would have been French.

If a "both Italian" world is closer than a "both French" world then (16) is true and (17) false; and vice versa. If both worlds are equally close, then there is no such thing as "the" A-world closest to actuality.

Stalnaker's approach, in more detail, uses a selection function which, for every antecedent–world pair, $\langle A, w \rangle$, delivers a unique A-world "most similar" to w. Formally, then, there is no problem about non-uniqueness. However, Stalnaker proposes that, in interpreting a conditional, context may determine which of various distinct selection functions is appropriate. Suppose that (15) is being considered against the background of some fantastic tale in which both musicians are born with no nationality, but are rescued from an international orphanage, and given name and nation. The parents who ended up adopting Verdi were Italian (so the story goes), and very nearly adopted Bizet as well, whereas the French parents who in fact adopted Bizet had no intention of adopting more than one child. Arguably, in this background, it will be natural, in interpreting (15), to prefer a selection function which assigns to the pair consisting of the world of the story and "Bizet and Verdi are compatriots" a world in which both are Italian.

There may also be cases, and as things are (15) and (16) are examples, in which we have no reason to prefer one selection function rather than another. In this case, there is no determinate assignment of truth value.[2]

Stalnaker's account has been criticized for rendering all sentences of the following form valid:

18) $(X \Box\!\!\rightarrow Y) \vee (X \Box\!\!\rightarrow \neg Y)$.

There is something to be said on behalf of its validity: to deny a non-indicative conditional of the form $X \Box\!\!\rightarrow Y$, a good thing to say is $X \Box\!\!\rightarrow \neg Y$. Arguably, either you or your protagonist must be right, in which case (18) should be valid.

[2] Among the ways different selection functions can be made salient is mood. It is on this basis that Stalnaker [1975] (who is offering an account of conditionals in general, and not just non-indicatives) can explain the distinction between indicative and non-indicative conditionals: they suggest different selection functions.

Its validity on Stalnaker's semantics follows from the facts that (i) Stalnaker's worlds are classical, in that the law of excluded middle $(X \vee \neg X)$ holds at each; and (ii) any selection function will choose a unique X-world. Since either Y or $\neg Y$ will hold at the selected world, so will one or other disjunct of (18).

There appear to be counterexamples to the validity of (18), for example

19) Either, if it were not 30° outside it would be 40°, or if it were not 30° it would not be 40°.

If it is not 30° it might or might not be 40°, but it seems wrong to affirm, under this supposition, that it would be 40° and wrong to affirm that it would not (cf. Pollock [1976], p. 16).

Logical space contains a number of A-worlds accounts. One of the most discussed has been offered by Lewis [1973b]. Adapting it to present purposes and terminology, it is this:

20) $X \Box\!\!\rightarrow Y$ is true at w upon i iff some world at which X & Y is true upon i is more similar to w than any world at which X & $\neg Y$ is true upon i, if there are any worlds at which X is true (if there are none, $X \Box\!\!\rightarrow Y$ is true).

This does not require a unique X-world or a unique (X & Y)-world: there can be any number. But it does require that if an (X & Y)-world ties for most similar with an (X & $\neg Y$)-world, the conditional is false. This makes formalizations of (16) and (17) false upon an intended interpretation (assuming that both-Italian worlds and both-French worlds are equally similar to actuality). It also withholds validity from (18). Consider the instance

Ex. 5.12 What is the truth value upon an intended interpretation of an adequate **PNS**-formalization of the following (applying (**2.20**))?

If Bizet and Verdi had been compatriots, they would have either both been Italian or else both French.

What impact does your answer have upon the correctness of Lewis's proposal?

21) (If the coin had been tossed it would have landed heads) or (if the coin had been tossed it would not have landed heads).

If the coin is never actually tossed, a head-landing and a non-head-landing world seem not to differ in degree of similarity to ours; hence no tossing-world in which it lands heads is *more* similar to actuality than any in which it doesn't; nor is any tossing-world in which it doesn't land heads more similar to actuality that any in which it does. So both disjuncts are false. As the example suggests, it is not entirely clear that this is a desirable result. Even if one has doubts about the validity of (**18**), one might be keen that it have some true instances – and (**21**) is, arguably, true.

Lewis says that the notion of overall similarity involved in his account is "extremely vague". He counts this as a virtue, for the non-indicative conditionals under discussion "are both vague and various. Different resolutions of overall similarity are appropriate in different contexts" ([1979], p. 56). Let us see how he applies these ideas by considering an objection. Intuitively, the following is true:

22) If Nixon had pressed the button, there would have been a nuclear holocaust. (Cf. Fine [1975]; Lewis [1979], in Jackson [1991], p. 59)

But it might seem to come out false on Lewis's semantics. A world in which Nixon pressed the button but through some fault nothing happened (and soon thereafter he changed his mind) would seem much more similar to our own world than one in there was no fault and so there was a holocaust. We don't think a fault was at all likely: we think that if the button had been pressed, the ICBMs would have taken to the skies. But we think this would have made for a world very unlike ours, much more unlike it than one in which there was a fault. For we also hold true the following:

23) If Nixon had pressed the button, things would have been radically different.

Lewis does not think that we have an entirely independent conception of similarity upon which we can simply draw to effect an analysis of conditionals. Rather, he thinks we can allow an explication

of the relevant conception of similarity to be shaped and guided by our judgements about non-indicative conditionals. The shaping, he suggests, involves weighting similarities. In evaluating conditionals like (**22**), we implicitly operate with a notion of similarity among worlds which gives greatest weight to avoiding "big, widespread, diverse violations of law"; next greatest to maximizing "the spatio-temporal region throughout which perfect match of particular fact prevails"; next greatest to avoiding "even small, localized, simple violations of law"; and least importance to securing "approximate similarity of particular fact" (cf. Lewis [1979], in Jackson [1991], p. 64). A world in which Nixon presses the button is somewhat unlike ours; to match ours perfectly where possible, we take everything to be in order with the wiring and the rockets. If we add that the holocaust occurs, we need not suppose any difference in laws. There are two ways to develop the supposition that the signal never gets to the rockets. Upon the first, the signal failure is the only difference in law. So the light waves carrying the image of Nixon pressing the button continue on their journey through space, Nixon's memories are different, and so on. In that case, we lose perfect match at that point, which is a grade 2 point of dissimilarity (we have approximate match, but this is only grade 4.) The idea is that this dissimilarity should outweigh the similarities. Alternatively, we secure reconvergence by a host of nomic differences (the light waves mysteriously peter out, etc.), but now we are involved in widespread violations of law, a grade 1 dissimilarity. Arguably, therefore, with appropriate weighting of similarities Lewis's semantics for conditionals assigns truth to (**22**).

There are some rather different difficulties about incremental similarities. Intuitively, a world in which I am taller than I actually am by being 6 ft 6 in is more similar to the actual world than one in which I am taller by being 6 ft 7 in. This leads to the strange consequence that

24) If I were taller, I would only be a tiny bit taller

is true. Any world in which I am taller by a given amount is less similar to actuality than one in which I am taller by a smaller amount. We might again see this case as calling for an adjustment in the concept of similarity. The original intuition, shared by Lewis and Stalnaker, was, in Lewis's words this time, that a conditional of the kind in question is true "iff it takes *less of a departure from actuality* to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent" (Lewis [1973c], p. 164, my emphasis). However, it is plausible to suggest that one world can by some standard be more similar to actuality than another while involving no less of a departure from actuality (cf. Pollock [1976], p. 21). I actually live in London; a world in which I live in Slough departs from actuality and so does a world in which I live in Oxford. There is nothing to choose between these departures from the point of view of evaluating any non-indicatives I can think of. However, there is a standard of similarity by which the first world is more similar to actuality than the second, for Slough is nearer to London than is Oxford. This difference is not to be taken seriously in evaluating a conditional like "If I didn't live in London I would live in Slough". It would be absurd to suppose that if I didn't live in London I would live near it: on the contrary, if I didn't live in London, I would live in the heart of the country.

I am actually 6 ft 4 in. Both being 6 ft 6 in and being 6 ft 7 in are departures from actuality. But when we are asked, by Lewis or Stalnaker, to consider *A*-worlds which involve "minimal departures" from actuality, the relevant minimality means worlds in which there are no changes other than something required to make the world an *A*-world: no "gratuitous" changes, as Lewis sometimes puts it. This gives a basis for a standard of similarity in which, relative to the antecedent of **(24)**, 6 ft 6 in worlds are no smaller departures than are 6 ft 7 in worlds.

The discussion is designed to show that the resources of possible worlds accounts are rich: if non-indicative conditionals have truth conditions at all (an issue raised for indicative conditionals in chapter **3.5**) it would be surprising if they could not be stated with more or less the resources specified here.

Stalnaker thought that just the same semantic clause would do for conditionals of every kind, the differences between indicatives and non-indicatives emerging simply at the "pragmatic" level in the choice of selection function. This is an extreme version of the view that conditionals need a unified treatment. Lewis, by contrast, holds that indicative conditionals are truth functional. It is open to him to regard his treatment of indicative and non-indicative as, in a sense, unified. Lewis's semantics for non-indicatives in effect require the truth of the corresponding truth functional conditional at a subset of possible worlds. He could regard "if" as contributing the truth functional conditional, and

departures from the indicative as contributing the rather complex modal material, so that there is no question of "if" having more than one meaning.

Supposing appears to be a mental act crucial in understanding conditionals. It is one thing to suppose that Oswald didn't kill Kennedy and another to suppose that he hadn't killed Kennedy; supposing is a kind of mental act which is normally involved in both non-indicative and indicative cases. This should encourage the view that there is a deep unity between these kinds of conditional.

3 Adding "□" to Q

QN results from adding box to **Q**, treating it as having the syntactical properties of tilde. This gives rise not only to sentences like

1) $\Box\forall x(Fx \rightarrow Fx)$

which, with "*F*" corresponding to "is a mathematician", seems a reasonable formalization of

2) Necessarily, all mathematicians are mathematicians;

but also to sentences like

3) $\forall x(Fx \rightarrow \Box Fx)$

which, with correspondences as before, would formalize

4) All mathematicians are necessarily mathematicians.

The reading of **(4)** formalized by **(3)** intuitively strikes one as false. People who become mathematicians don't have to turn to mathematics: there is an element of choice.

The interpretation rules for **PN** are silent on the question of how quantifiers interact with modal operators. Intuitively we want an interpretation with respect to the actual world to falsify **(3)**, for a mathematician who didn't *have* to be a mathematician would be a counterexample to the conditional. In the obvious way of understanding this idea, this means that some mathematician needs to occupy two worlds, one at which he is a mathematician and one at which he is not. The latter fact ensures the falsity of the consequent of the conditional (as instanced on this person), and the former the truth of its antecedent. Not all theorists allow that we can make sense of objects which exist at more than one world (see §**9** below). Even if one can make sense of it, we need to say whether interpretations have the same domain relative to each world, or whether the domain varies from world to world. Choices on these issues are controversial, and have an impact on validity. In this section I will adopt one mainstream view. It has an unrestricted accessibility relation (so this will not be mentioned); it allows objects to exist at more than one world; and it allows interpretations to have different domains with respect to different worlds.

We stipulate that for each world w there is a non-empty domain D^w (the entities in w). In chapter **4**, we used expressions like "$i(F)$" to designate the set assigned by a **Q**-interpretation, i, to the predicate letter "F". Here we extend that notation, using expressions of the form "$i_w(X)$" to designate what a **QN**-interpretation, i, assigns to an expression, X, with respect to a world, w.

5) For any set of worlds, W, any world, w, in W, and any interpretation i of **QN**:

for each sentence-letter P, $i_w(P)$ is a truth value, either T or F;

for each name-letter n, for some world w', $i_{w'}(n)$ is an object, α, in $D^{w'}$ and for every world, w'', if α belongs to $D^{w''}$, $i_{w''}(n)$ is α and if α does not belong to $D^{w''}$ there is no $i_{w''}(n)$ (e.g. "α" might be assigned Ronald Reagan with respect to any world at which Reagan exists, and assigned nothing with respect to the other worlds);

for each predicate-letter ϕ of degree n, $i_w(\phi)$ is a set of n-tuples (n-membered sequences) all of whose members belong to D^w (for example, $i_w(F)$ might be a set of ordered pairs in w such that the first member of the pair loves the second);

$i_w(=)$ is the set of ordered pairs of members of D^w such that in each pair the first object is the same thing as the second.

An intended interpretation of a **QN**-sentence will assign to each name-letter in the sentence, with respect to each world w, the actual bearer

of the corresponding name, if that thing exists at w, and otherwise will assign nothing to the letter with respect to w; it will assign to each n-ary predicate-letter in the sentence, with respect to each world w, those n-tuples of members of D^w which possess the property associated with the corresponding predicate.

6) For any set of worlds, W, any world, w, in W, and any interpretation i of **QN**,

- (i) $i_w(\neg X)$ is T iff $i_w(X)$ is F.
- (ii) $i_w(X \,\&\, Y)$ is T iff $i_w(X)$ is T and $i_w(Y)$ is T.
- (iii) $i_w(X \vee Y)$ is T iff $i_w(X)$ is T or $i_w(Y)$ is T.
- (iv) $i_w(X \rightarrow Y)$ is true iff $i_w(X)$ is F or $i_w(Y)$ is T.
- (v) $i_w(\phi n_1 \ldots n_k)$ is T iff all of $i_w(n_1) \ldots i_w(n_k)$ belong to D^w and $<i_w(n_1) \ldots i_w(n_k)>$ belongs to $i_w(\phi)$; otherwise $i_w(\phi n_1 \ldots n_k)$ is F.[3]
- (vi) $i_w(\forall v X)$ is T iff $X\frac{n}{v}$ is T upon every n-variant of i which assigns something to n at w.
- (vii) $i_w(\exists v X)$ is T iff $X\frac{n}{v}$ is T upon some n-variant of i which assigns something to n at w.
- (viii) $i_w(\Box X)$ is T iff for all worlds w' in W, $i_{w'}(X)$ is T.

QN-validity can be defined exactly like **PN**-validity. What makes the difference is the richer notion of interpretation for **QN** as compared with **PN**.

An intended interpretation is one which assigns entities to the various simple expressions in such a way as not to depart from their actual meaning. Such an interpretation should deliver adequate formalizations: ones whose truth-upon-an-intended interpretation conditions match the truth conditions of the English. Applying our interpretation rules to (**3**), an intended interpretation, i, will, for each of its domains, assign the set of mathematicians in the domain to "F". (When a domain contains no mathematicians, $i(F)$ is the empty set.) This interpretation will assign truth to (**3**) with respect to the actual

[3] (**5**) and (**6**) between them assume what Forbes calls the "Falsehood Principle": an atom is false with respect to a world if a name-letter in it is not assigned a referent with respect to that world (see Forbes [1985], p. 29–31). There are other ways of proceeding, which would lead to different truth-upon-an-interpretation conditions. One motivation for my choice is to keep close to the semantics of **Q**, which required every name-letter to be assigned an object.

world, w★, iff α-variants of it which assign something to α at w★ assign truth to "$F\alpha \rightarrow \Box F\alpha$" with respect to w★. Some α-variant i' of i will assign Cantor to α at w★. He was a mathematician, but he didn't have to be one. So $i'_{w\star}(F\alpha)$ is T, but there is a world w other than w★ such that $i'_{w}(F\alpha)$ is F, so $i'_{w\star}(\Box F\alpha)$ is F, so $i'_{w\star}(F\alpha \rightarrow \Box F\alpha)$ is F, so $i_{w\star}(\forall x(Fx \rightarrow \Box Fx))$ is F.

Whether (**6**viii) is appropriate to the ordinary notion of necessity raises a number of issues, for example those relating to whether or not the accessibility relation between worlds needs to be restricted (cf. §1 above). Here we raise just one issue, concerning existence and a distinction between "*strong*" and *"weak" necessity*.

We have given an example in which a sentence of the form "$\Box F\alpha$", "Necessarily, Cantor was a mathematician", is false. Are there any examples of truths of this form? One candidate is

7) Necessarily, Socrates is human.

You might well think that this is true: anything non-human – a stone or a crocodile – simply could not be Socrates. Even if there is room for doubt on this point, it would seem that an appropriate language for necessity should not preclude the truth of a sentence like (**7**). However, the **QN**-formalization of (**7**) as "$\Box F\alpha$" will not be true upon an intended interpretation, i. With respect to each world, w, $i_w(F)$ is the set of all humans in D^w; and $i_w(\alpha)$ is Socrates with respect to every world at which Socrates exists. Any reasonably rich set of worlds will contain one whose domain does not include Socrates – a world at which Socrates does not exist. (We all agree that Socrates might not have existed, and would not have done if the world had ended ten years before the year in which he was born.) With respect to such a world, w, $i_w(F\alpha)$ is F, by (**6**v), so $i_{w\star}(\Box F\alpha)$ is false. An intended interpretation assigns F to a **QN**-sentence which is supposed to formalize a truth (or at least something whose truth is plausible); and this is a sign that something is wrong.

There are at least two ways of responding to this difficulty. One involves weakening the interpretation rule for $\Box$ so that $\Box X$ is true upon an interpretation i iff X is true upon i at every world at which all the objects which i assigns to any name-letters in X exist. Then we could formalize (**7**) straightforwardly, as "$\Box F\alpha$", and it would not be false upon an intended interpretation merely in virtue of the fact that Socrates might not have existed.

Alternatively, we could keep to the original interpretation rule for $\Box$, and formalize sentences like (**7**) as

8) $\Box((\exists x\, x = \alpha) \rightarrow F\alpha)$

with "F" corresponding to "is human". The semantics for **QN** do not preclude (**8**) being true upon an intended interpretation. Worlds at which Socrates does not exist will be worlds at which "$(\exists x\, x = \alpha) \rightarrow F\alpha$" is true upon an intended interpretation in virtue of the falsity of the antecedent at such worlds.

The first response introduces the so-called "weak" interpretation of necessity; the interpretation rule (**6**viii) expresses the "strong" interpretation

Necessitated existential sentences themselves provide a reason for preferring the strong interpretation, and thus for adopting the more complex formalizations, in the style of (**8**), of sentences like (**7**). We might debate whether the following is true:

9) Necessarily, the number 7 exists.

We could formalize this by

10) $\Box\exists x\, x = \alpha$

with "α" corresponding to "the number 7". Keeping to our rule (**6**viii), that is, treating the box as expressing strong necessity, the question of the truth of (**10**) upon an intended interpretation is not foreclosed by the formalization. If we interpret box as expressing weak necessity, (**10**) will express a trivial truth upon an intended interpretation: it will in effect say that α exists at every world at which it exists. This would make it a poor candidate for formalizing (**9**), and a hopeless candidate for formalizing "necessarily, Socrates exists".

On the strong interpretation of box, (**10**) is false upon some interpretations at some worlds and so is not trivial. Consider an interpretation assigning Socrates to "α". "$\exists x\, x = \alpha$" will be false upon this interpretation at a world at which Socrates does not exist, so upon this interpretation (**10**) will be false (at every world).

This virtue of the strong interpretation carries with it a worry. The fact that (**10**) is not **QN**-valid means that we cannot accept for **QN** the correlate of (**1.11**) for **PN**. That is, the following is false:

11) If [$\models_Q X$], then [$\models_{QN} \Box X$].

We have $\models_Q \exists x\, x = \alpha$, but not $\models_{QN} \Box\exists x\, x = \alpha$. Given the philosophical motivation for (1.**11**), this should be genuinely disturbing. It would not be satisfactory to try to reinstate (**11**) by insisting that every name-letter is always assigned something by every interpretation with respect to every world, for then, intuitively, a name-letter cannot adequately formalize a name for a contingent being, one which actually exists but which might not have done.

Ex. 5.13 Assess the following argument:

True, valid sentences should correspond to necessary truths. The problem raised by (**3.10**), however, is a problem with **Q**, not with strong necessity. For **Q** invites us to formalize the doubtful sentence "Homer exists" by the **Q**-valid $\exists x\, x = \alpha$. Modify **Q** so that it is a free logic, and (**3.11**) will be, as it should be, true.

This raises deep issues about the role of names.[4] At the more superficial level, the strong version of necessity has the advantage noted of giving natural formalizations both of claims like (**7**) and claims like (**9**), so I shall persist with it.

Any sentence formalizable as "$\Box F\alpha$" can be read as ascribing a property to an object: it ascribes to α the property of being necessarily F. Sentences like (**3**) speak generally of things being necessarily thus-and-so. For a number of different reasons, philosophers have held that such ways of talking are illegitimate. If they are right, then it is bootless to investigate much further the properties of **QN**, since it is committed to this supposedly illegitimate way of talking.

Ex. 5.14 Formalize the following in **QN** (omitting parenthetical material), commenting on the validity of your formalizations of (ii) and (iii):

(i) Causation is a necessary relation.

[4] The interpretation rules for **QN** are designed to mirror a view of names proposed informally by Kripke [1972]: that they are rigid designators. This view has been the subject of considerable debate: see e.g. McCulloch [1989] and bibliographical references associated with chapter **4.7** and **4.12** above. In recent work, Kripke's views about the modal profile of proper names have become entangled with theoretically distinct questions about the epistemology of names. See e.g. Recanati [1993].

(ii) A necessary condition for the possibility of experience is that there are causally related events. So the existence of our experience establishes the necessity of the causal relation.

(iii) It is possible for my body to exist when I do not (e.g. after my death). Therefore I am not my body.

(iv) What is known must be true.

(v) A married man must be married *to* someone.

4 Necessity de re and de dicto

A sentence expresses "*necessity de re*" iff it is adequately **QN**-formalizable by a sentence in which there is a name-letter within the scope of some occurrence of box or if there is an occurrence of box within the scope of a quantifier. Let us say that a sentence expresses "*necessity de dicto*" just on condition that it expresses necessity but does not express necessity de re. The doubts alluded to at the end of the last section are doubts concerning the coherence of the notion of de re necessity.

As I have defined the difference between de re and de dicto necessity, it is simply a matter of scope. The definitions do not introduce two concepts of necessity. We could with our available resources make an exactly parallel distinction between "negation de re" and "negation de dicto", but the distinction would give no support to the thought that there are two distinct concepts of negation.

Ex. 5.15 Give a pair of examples to contrast "negation de re" with "negation de dicto".

The distinction between de re and dicto necessity is sometimes associated with some or all of the following further (and independent) theses:

1) In ascribing de re necessity we attribute a property to a non-linguistic object; in ascribing de dicto necessity we attribute a property to a sentence.

2) A de re necessary truth records how things are in the world; a de dicto necessary truth records only linguistic facts.

3) Whether a de re necessary sentence is true depends on which objects exist at which worlds.

One who believes that there are de re necessary truths is called an "essentialist". Optionally, he may also subscribe to some or all of the above theses. Attacks on the intelligibility of de re necessity are sometimes expressed as attacks on essentialism. In the following three sections, I shall consider three such attacks.

5 The number of the planets

Quine has advanced two related arguments against essentialism (e.g. [1960], §41). First, he asked us to consider the following argument:

1) 9 is necessarily greater than 7;
9 = the number of the planets.
Therefore, the number of the planets is necessarily greater than 7.

Quine claimed that the argument has true premises and a false conclusion, yet that it ought to be valid if essentialism is true. If essentialism is true, the first premise will constitute the ascription to an object (the number 9) of a property, that of being necessarily greater than 7, and the conclusion will constitute the ascription of the same property to the same object.

Quine does not use the distinction between de re and de dicto given in §4. Rather, he identifies a de re statement by a claim of (**4.1**): it is one which ascribes a property to an object. However, a sentence judged by this last standard to be one of de re necessity is also de re by our official standard. If the premise of (**1**) ascribes the property of being necessarily greater than 7 to the number 9, then it is appropriately **QN**-formalizable as:

2) $\Box F\alpha$.

Here the name-letter "α" falls within the scope of box; so we must follow Quine in counting it, and hence the first premise of (**1**), de re.

We can also agree with Quine that an essentialist *may* view (**1**) as sound. The following formalization is **QN**-valid:

3) $\Box F\alpha,\ \alpha = \beta;\ \Box F\beta$.

(The correspondence scheme is: "F" to "is greater than 7", "α" to "9" and "β" to "the number of the planets".) The essentialist will allow that the premises are true upon an intended interpretation. So he must allow that the conclusion is also true upon an intended interpretation. So he should hold that the conclusion of (**1**) is true. Why does Quine say that it is false?

His ground is that there might not have been nine planets. If there had been only 5, then the number which numbers them would not have been greater than 7, let alone necessarily greater than 7. Although this is correct, it is not inconsistent with the truth upon an intended interpretation of the conclusion of (**3**). Let us see what an intended interpretation, i, assigns to "β", "F" and "$F\beta$" with respect to a world, w, at which there are just 5 planets. First, $i_w(\beta)$ is the very object that i assigns to "β" with respect to the actual world, viz. the number 9 (assuming it belongs to D^w);[5] $i_w(F)$ is the set of members of D^w which, at w, are greater than 7. Thus 9 will be a member of $i_w(F)$, and so $i_w(F\beta)$ is T. That there are worlds in which there are only 5 planets is irrelevant to the truth of the conclusion of (**1**), as it will be understood by one who takes the argument to be formalizable by (**3**), and so valid.

The conclusion of (**1**) is, perhaps, ambiguous. If (**1**) is a valid argument, its conclusion must be read in the de re fashion brought out by the formalization (**3**). The de dicto reading of the conclusion might be formalized:

4) $\Box\exists x(Gx\ \&\ \forall y(Gy \rightarrow x = y)\ \&\ Fx)$

with "G" corresponding to "numbers the planets". For the argument to have a chance of being valid, we must make an analogous adjustment to the formalization of the phrase "the number of the planets" as it occurs in the premise, so that the overall formalization would be:

5) $\Box F\alpha,\ \exists x(Gx\ \&\ \forall y(Gy \rightarrow x = y)\ \&\ x = \alpha)$;
$\Box\exists x(Gx\ \&\ \forall y(Gy \rightarrow x = y)\ \&\ Fx)$.

(**5**) is invalid: the premises do not ensure even that every world has a unique G, let alone one which is F. The essentialist has no more

[5] It is crucial to this point that, contrary to an official policy of chapter 4, "the number of the planets" be formalized by a name-letter, as in (**3**) above.

reason than anyone else to suppose that (5) is valid or has a true conclusion.

The failure of the argument against essentialism can be summarized as follows: if the conclusion of (1) is read in such a way that it is false, the essentialist has no more reason than the rest of us for supposing the argument to be valid. If (1) is read in such a way that the argument is valid, then no reason has been given to suppose that the conclusion is false.

Quine used another argument to make a similar point, and a similar criticism applies.[6] He says that essentialists will subscribe to the following:

6) All bachelors are necessarily unmarried but not necessarily army officers

7) All majors are necessarily army officers but not necessarily unmarried.

Consider Major Smith, who is both a bachelor and a major: qua bachelor, he is not necessarily an army officer, but qua major he is necessarily one; qua bachelor he is necessarily unmarried, but qua major he is not necessarily unmarried. So there is no question of an object possessing a property necessarily or contingently: what matters is how you refer to the object. Hence essentialism – the claim that necessity attaches to objects – is false.

Appropriate **QN**-formalizations of (6) and (7) are all de dicto:

8) $\Box\forall x(Fx \rightarrow Gx)\ \&\ \neg\Box\forall x(Fx \rightarrow Hx)$

9) $\Box\forall x(Jx \rightarrow Hx)\ \&\ \neg\Box\forall x(Jx \rightarrow Gx)$

with "F" corresponding to "is a bachelor", "G" to "is unmarried", "H" to "is an army officer" and "J" to "is a major". No conclusions of the de re form "$\Box G\alpha$" or "$\neg\Box H\alpha$" follow from (8) or (9) or their conjunction. These are not universal quantifications, so we cannot infer to an instance. The essentialist should say that Major Smith is neither necessarily unmarried nor necessarily an army officer. The crucial point is that the following is consistent – that is, true upon at least one interpretation:

[6] See Quine [1960], §41, the example of the mathematical cyclists. I have made trivial changes in the details, to avoid irrelevant objections.

10) $\Box\forall x(Fx \rightarrow Gx)\ \&\ F\alpha\ \&\ \neg\Box G\alpha$.

Bachelors don't have to be bachelors (denial of de re necessity), even though, necessarily, everyone who is one is unmarried (affirmation of related de dicto necessity).

Quine's arguments do indeed show that not all necessity is de re. However, no one should deny this. Essentialism is not the claim that all necessity is de re, but only that some is.

Ex. 5.16 Formalize the following:

Not every belief can be justified, so there are beliefs which cannot be justified.

Is the argument as formalized **QN**-valid? What does your answer show about the relation between de dicto and de re necessities? (Cf. Anscombe [1959], pp. 138–9.)

6 "Frege's argument"

An argument has been attributed to Frege which, if sound, would establish the absurdity of de re necessity. Without committing myself to the accuracy of this attribution, I shall refer to it as "Frege's argument".[7]

De re necessities formalizable as "$\Box F\alpha$" satisfy the substitutivity of identicals: in conjunction with a premise formalizable as "$\alpha = \beta$" you validly get a conclusion formalizable as "$\Box F\beta$". Moreover, "$\Box$" is plainly a non-truth functional sentence connective. The truth value of $\Box X$ is not determined by the truth value of X. The conclusion of Frege's argument is that these features are incompatible: where the substitutivity of identicals holds, there you must have truth functionality. I will introduce some terminology which will place this point in a wider setting, and facilitate its discussion.

Let us say that the *extension* of a sentence is its truth value, the extension of a name, its bearer, and the extension of a predicate of degree n the set of n-tuples of which it is true. In other words, extensions of

[7] Compare Quine [1960], pp. 148–9; Davidson [1967c], pp. 153; Davies [1981], pp. 210–11; Neale [1990], [1995]. The source is supposed to be Frege [1892b].

expressions are the sorts of object a **Q**-interpretation assigns to corresponding letters.

A sentence is *extensional with respect to a position for an expression (sentence, name or predicate)* iff replacing the expression in that position with any other expression having the same extension leaves the truth value of the whole sentence unchanged.

A sentence is *extensional, tout court,* iff it is extensional with respect to all the positions for sentences, names and predicates it contains.

A language is extensional iff all its sentences are extensional.

Truth functionality is a special case of extensionality: to say that a language is truth functional is just to say that all its sentences are extensional with respect to all positions for sentences.

Lewis Carroll, the author of the Alice books, was a mathematician whose "real" name was Charles Dodgson. This is well known to one of his colleagues, John. So the following is true:

1) John believes that Charles Dodgson is a mathematician.

This sentence is not extensional with respect to the position of the sentence "Charles Dodgson is a mathematician". If it were extensional with respect to that position, then the result of inserting any truth into that position would be a truth. We know that this is not so, since John, being of only finite intellect, does not believe every truth.

(**1**) is also not extensional with respect to the position occupied by the name "Charles Dodgson".[8] I can without inconsistency stipulate that John does not believe that Lewis Carroll is a mathematician, despite the fact that

2) John believes that Lewis Carroll is a mathematician

results from (**1**) by replacing a name by another name having the same extension.

Finally, (**1**) is not extensional with respect to the position occupied by the predicate "is a mathematician". For, let us suppose, this predicate is true of just the same things as those of which the predicate "can rattle off a proof that there is no greatest prime" is true. Yet I can without inconsistency stipulate that

[8] Compare chapter **4.18**. An opaque position for a name is simply a non-extensional one.

3) John believes that Charles Dodgson can rattle off a proof that there is no greatest prime

is false, despite the fact that it results from (**1**) by replacing a predicate by another predicate having the same extension.[9]

The non-extensionality of (**1**) with respect to the position occupied by "Charles Dodgson" does not entail that it is non-extensional with respect to the position occupied by "John". Moreover, if the same name (or predicate or sentence) occurs twice, the sentence may be extensional with respect to the position of one occurrence but not extensional with respect to the position of another. For example,

4) Charles Dodgson believes that Charles Dodgson is a mathematician

is extensional with respect to the first but not the second position occupied by "Charles Dodgson". To illustrate the non-extensionality, imagine that some other name, say N, for Dodgson is in general currency, though Dodgson wrongly thinks that N refers to someone else, a botanist, and so does not believe that N is a mathematician.

Modal contexts uncontroversially give rise to non-extensionality with respect to sentence-positions (that is just to say that modal sentence connectives are not truth functional). They also give rise to non-extensionality with respect to predicate-positions.

5) Necessarily, all humans are humans

is true, but presumably

6) Necessarily, all featherless bipeds are humans

is false, since it is only contingently true that there are no non-human bipeds without feathers. Yet (**6**) results from (**5**) by replacing a predicate occupying the relevant position by another with the same extension.

We can extend the definition of extensionality to the language **Q** by relativizing to an interpretation throughout, and replacing talk of

[9] For scepticism about some of these claims of non-extensionality in English, see Bach [1987], pp. 206–14.

name-position by talk of name-letter-position. Thus defined, **Q** is extensional, a fact ensured by the interpretation rules which attend only to the extensions (upon an interpretation) of the various expressions. Any English sentences adequately **Q**-formalizable are extensional with respect to the positions of expressions corresponding to **Q**-letters.

By specifying *intensions* for sentences, names and predicates we can arrive, by analogous definitions, at various notions of intensionality. We stipulate that the intension of an expression is the set of all ordered pairs whose first member is a possible world and whose second member is the extension of that expression with respect to that world. A **QN**-interpretation thus fixes an intension for every category of **QN**-letter.

QN is non-extensional, but it is *intensional*. This is because the interpretation rules in effect attend just to the intensions of the various expressions. Some English sentences dominated by "necessarily" are certainly intensional with respect to some contained positions for sentences and predicates. Obviously, an English sentence "necessarily, *A*" is intensional with respect to the *A*-position: if *A* and *B* have the same intension, that is, are true in just the same worlds, "necessarily *A*" will have the same truth value as "necessarily *B*".

English sentences of the form "necessarily, all *F*s are *G*s" are intensional with respect to the positions of "*F*" and "*G*".

Some English sentences contain positions which are neither intensional nor extensional (so it is a bad idea to abbreviate "non-extensional" as "intensional"). (**1**) is an example:

John believes that Charles Dodgson is a mathematician.

"Charles Dodgson is a mathematician" has the same intension as "Either Charles Dodgson is a mathematician or there is a largest prime", but one may consistently suppose that John believes the first and not the second.

Ex. 5.17 Explain why "Charles Dodgson is a mathematician" has the same extension as "Either Charles Dodgson is a mathematician or there is a largest prime".

"Charles Dodgson" and "Lewis Carroll" have the same intension: with respect to each world, they name the same person. Yet we have already seen that their substitution in (**1**) may fail to preserve truth value.

Frege's argument claims that if a sentence is extensional with respect to all its name-positions, then it is extensional with respect to all its sentence-positions; that is, it is truth functional. Sentences formalizable in **QN** are extensional with respect to name-positions, as will shortly be made plain. Hence it should follow that **QN** is truth functional, which it evidently is not. If one is convinced of the soundness of Frege's argument, one can but conclude that there is something incoherent about the semantic notions which animate **QN**.

Built in to the interpretation rules for **QN** is the requirement that a name-letter be assigned, with respect to every world *w*, the same object, providing that the object exists at *w*. An intended interpretation of an English sentence was said to be one which assigned to a name-letter in the formalization the object which bears the corresponding name. So if English is adequately **QN**-formalizable, an English name which in fact names *o* must also name *o* with respect to each world at which *o* exists. This in turn ensures that names having the same extension also have the same intension. The intensionality with respect to name-positions of a language formalizable in **QN** ensures the extensionality with respect to name-positions of that language.

To take an example. The fact that "Lewis Carroll" and "Charles Dodgson" have the same extension is enough to ensure that the following sentences have the same truth value:

7) Necessarily, Charles Dodgson is human.

8) Necessarily, Lewis Carroll is human.

We thus have the most important premise that Frege's argument requires: **QN** is extensional with respect to all its name-positions.

The argument requires a further uncontroversial assumption:

9) If any sentence adequately **QN**-formalizable as $\Box(A \leftrightarrow B)$ is true, then the truth value of any larger **QN**-formalizable sentence containing *A* is the same as the truth value of the result of replacing *A* by *B*.

We know that (**9**) would be false without the restriction to adequately **QN**-formalizable sentences containing *A* and *B*; but its truth with respect to **QN**-formalizable sentences is secured by the intensionality of **QN**, and thus by the intensionality of the sentence-position occupied by A.

Ex. 5.18 Show that **(6.9)** would be false without the restriction to QN-formalizable sentences.

Now for the argument itself. We suppose that "p" and "q" are two sentences with the same truth value, and we use "δ" as a device to form from a sentence an expression able to occupy name-position, with the stipulation that for any sentence, s, "$\delta s = 1$" is true iff s is true, and "$\delta s = 0$" is true iff s is false. Since this is constitutive of the meaning of "δ", we can infer that

10) $\Box(\delta p = 1 \leftrightarrow p)$, and $\Box(\delta q = 1 \leftrightarrow q)$.

The claim is that the following series of transformations preserves truth value:

11) (i) $\ldots p \ldots$
(ii) $\ldots \delta p = 1 \ldots$
(iii) $\ldots \delta q = 1 \ldots$
(iv) $\ldots q \ldots$

(9) and **(10)** together ensure that **(11**i**)** and **(11**ii**)** have the same truth value and so do **(11**iii**)** and **(11**iv**)**. The assumption of extensionality with respect to name position, together with the obvious fact that the co-extensiveness of "p" and "q" ensures the co-extensiveness of "δp" and "δq", establishes that **(11**ii**)** and **(11**iii**)** have the same truth value. Hence a context which is extensional with respect to positions for names must also be extensional with respect to positions for sentences, that is, truth functional.

The argument appears to be valid and to have true premises. Must we, therefore, conclude that **QN** is incoherent?

We need to distinguish a broad and a narrow conception of a "name", corresponding to which there are broad and narrow conceptions of extensionality with respect to name-positions. On the broad conception, Frege's argument is valid but not sound, since it is not the case that a language formalizable in **QN** is broadly name-extensional. On the narrow conception, Frege's argument is not valid, since the crucial expressions, those which stand to the left of the identity sign in **(11**ii**)** and **(11**iii**)**, are not, on this conception, names.

A name, broadly conceived, is any expression in some language or other which is supposed to refer to a particular object and which can combine with an n-ary predicate and n-1 other names to form a sentence. (The circularity will not matter for our purposes.) Names in the broad sense include definite descriptions, like "the value taken by s", and its symbolic equivalent "δs". No expressions of this kind belong to **Q** or to **QN**. In these languages, there is no expression which is both complex (that is, contains other expressions as proper parts) and also capable of occupying the position occupied by name-letters. We have already seen (in chapter **4.10**, especially **(4.10.16)** and **(4.10.17)**) that sentences like those mentioned in **(10)** will not be classified in **Q** as identity sentences, but rather as existential generalizations, and the same applies to **QN**. If we try to formalize "$\delta s = 1$" in **QN**, the result would be something like

12) $\exists x(Fx\beta \;\&\; \forall y(Fy\beta \rightarrow y = x) \;\&\; x = \alpha)$

where "Fxy" corresponds to "x is a value taken by y", "β" corresponds to "s" (which here has to be thought of as the *name* of a sentence), and "α" corresponds to "1". **(12)** says that there is a unique thing, x, which is a value taken by s, and $x = 1$. This will not as it stands serve as a premise to the application of the substitutivity of identicals. If "name" is defined widely enough to include definite descriptions, then a language formalizable in **QN** is not extensional with respect to name-positions. Consider the pair:

13) Necessarily, the author of *Waverley* (if there is a unique author of *Waverley*) wrote *Waverley*

14) Necessarily, Scott (if there is a unique author of *Waverley*) wrote *Waverley*.

The first is true and the second false, despite the fact that the second results from the first by replacing an occurrence of a "name" ("the author of *Waverley*") by a co-extensive name ("Scott").[10]

[10] Strictly speaking, the notion of an extension has not been defined for definite descriptions. This is no accident. From the perspective of languages like **Q** and **QN**, definite descriptions do not have extensions: the interpretation rules do not assign entities to them. However, for dialectical purposes it is obvious how the notion of the extension of a description should be understood.

The **QN**-formalizations of (**13**) and (**14**) are, respectively:

15) $\Box(\exists x(Fx \& \forall y(Fy \rightarrow x = y)) \rightarrow \exists x(Fx \& \forall y(Fy \rightarrow x = y) \& Fx))$

16) $\Box(\exists x(Fx \& \forall y(Fy \rightarrow x = y)) \rightarrow F\alpha)$

with "*F*" corresponding to "wrote *Waverley*" and "α" to "Scott". (**15**) is true upon any interpretation, having almost the form "$\Box(p \rightarrow p)$". (**16**) is arguably not true upon an intended interpretation, since there are worlds in which someone other than Scott uniquely wrote Waverley, worlds *w* such that the antecedent of the conditional is true upon the interpretation at *w* and the consequent false upon the interpretation at *w*.

Ex. 5.19 It might be argued that authors of works are essential to those works being the very works they are. Someone might have written a book word-for-word like *Waverley*, but, on this view, it would not have been *Waverley*. Give an alternative pair of sentences which make the point (**6.13**) and (**6.14**) were intended to make, but which are not open to this objection. (It might be helpful to think back to the number of the planets.)

If we keep to the narrower conception of a name, according to which only an expression appropriately formalizable by a **QN**-name-letter is a name, then Frege's argument is invalid, since there is no justification for the claim that, regardless of the context, (ii) and (iii) have the same truth value. This is because "δp" and "δq" are not names, on this narrower conception.

The very different behaviour of names and descriptions in these contexts might well ground the view that they should not be treated as belonging to a common semantic category. We will not pursue this thought, but we will always think of a name in accordance with the narrower conception, which excludes complex expressions like definite descriptions.

The upshot of this discussion is that Frege's argument does not show that the name-extensionality of **QN**, and the associated expressions of de re modality, are inconsistent with **QN**'s non-truth functionality.

7 Trans-world identity

Evaluating de re sentences by the interpretation rules of **QN** requires attention to the *identity* of objects at various possible worlds. The rule for interpreting name-letters requires an interpretation to assign the same object to any name-letter with respect to each world at which the object exists. Hence this rule already incorporates the notion of trans-world identity: an interpretation must settle which object in, say, *w* is the object which it has assigned to "α" with respect to some other world.

The assumption of trans-world identity also emerges in connection with de re quantifications. Compare

1) $\Box\forall xFx$

2) $\forall x\Box Fx$.

In interpreting (**1**), a de dicto sentence, we do not have to consider how the entities at various worlds are related to the ones at the world of evaluation. For any interpretation, *i*, any world, *w*, any *n*-ary predicate, *F*, all we have to consider is whether all *n*-tuples formed from the members of D^w (whatever they may be) belong to $i_w(F)$. In interpreting (**2**), a de re sentence, with respect to *w*, we have to consider which objects from D^w also belong to the domains of other worlds. Suppose the question is whether (**2**) is true upon an interpretation, *i*, with respect to a world *w*. Then it is as if the quantifier of (2) ranged just over D^w, since the interpretation rule for "$\forall$" makes the truth of (2) turn on the truth of "$\Box F\alpha$" with respect to α-variants of *i* which assign something, say *o*, to α at *w*. Each of these variants has to determine, for each world, whether *o* belongs to what it assigns to "*F*" at that world. So each variant has to trace *o* through various worlds, and thus has to settle questions of "trans-world identity".

Imagine two types of atheistic views about God's existence. They are both "soft" atheisms, in that they agree that although there is no God, there could have been one. One view is formalizable:

3) $\Diamond\exists xFx$

with "*F*" corresponding to "is omnipotent, benevolent etc." (here insert a complete list of the attributes appropriate to God). The other view is formalizable:

4) $\exists x \Diamond Fx$

with the correspondence as before. The first view is de dicto: it is true upon an intended interpretation, *i*, with respect to *w*, iff for some world w', $i_{w}'(F)$ has at least one member. The second view is de re: it is true upon an intended interpretation, *i*, with respect to *w*, iff for some α-variant which assigns an object, say *o*, to α at *w* verifies "$\Diamond F\alpha$"; iff there is a world at which *o* belongs to $i(F)$. So (**4**) corresponds to the claim that someone exists who *could* have been God. The interpretation will settle whether someone who exists at the world of evaluation, *w*, is God at some possibly distinct world, w', so it involves trans-world identity.

Ex. 5.20 (a) Say which of the sentences (**7.3**), (**7.4**) entails the other. (b) Suppose that there is a single domain of actual and possible objects, common to each world. Show that in this case (**7.3**) and (**7.4**) would be equivalent.

We must agree with Quine that de re sentences, as **QN**-formalized, involve trans-world identity. If Quine is right, trans-world identity is unintelligible, and so **QN**, and all languages adequately formalizable therein, would likewise be unintelligible.

The issue is metaphysical, not epistemic. How, if at all, we *know*, concerning a possible world, that it does or does not contain a given individual is not at issue. If there is a fact of the matter which we do not know, then de re modality is intelligible, but permits the expression of propositions whose truth values we do not know.

It seems as if the intelligibility of trans-world identity is ensured by the fact that we can introduce a counterfactual situation by referring to a particular object. We might say: envisage a situation in which Nixon tells the truth. If this introduces a possible situation at all, there seems no room for doubt about whether it is a situation in which Nixon exists. We have specified the situation in terms of an object in it (cf. Kripke [1972]). On the face of it, Nixon belongs to at least two worlds, the actual world and a world where he is honest. How can there be some hidden difficulty here?

A difficulty arises if one holds the following two theses: (i) the only legitimate specification of a possible world is purely qualitative; and (ii) a purely qualitative specification is insufficient to determine which

actual individuals are present at a world. If the only facts which could determine trans-world identity, namely purely qualitative facts, fail to do so, then, of course, the trans-world identities are not determined. But theses (i) and (ii) are an unattractive combination. If you believe that all facts are determined by qualitative facts, then, fair enough, you will hold thesis (i); but, by the same token, you will reject thesis (ii). If you believe thesis (ii), then you have a reason to reject the view that all facts are determined by qualitative facts, and thus a reason not to agree that only qualitative specifications of worlds are legitimate.

That is the end of the discussion, unless we can find some arguments against the intelligibility of trans-world identity. I shall mention one, and briefly discuss another.

Lewis [1986a] has given a powerful argument, but it has two features which make it inappropriate for discussion here. First, it depends upon Lewis's preferred method of constructing non-actual worlds (as mereological sums of genuinely existing, though non-actual, spatio-temporally related things), and is therefore only available to one who is familiar with, and accepts, that construction. Secondly, it in no way impugns the intelligibility of de re modality, but simply justifies Lewis's preferred method of representing it.

The argument I shall briefly discuss is due to Quine. He compares trans-world identity with trans-moment identity:

> Our cross-moment identification of bodies turned on continuity of displacement, distortion and chemical change. These considerations cannot be extended across worlds, because you can change anything to anything by easy stages through some connecting series of possible worlds. ([1976], p. 861)

Quine is claiming that we can make sense of identity across moments, because this is determined by various kinds of continuity, but that we cannot make sense of identity across worlds, because it is not determined by anything. For example, this table on which I am writing, α_0, could have been made of slightly different parts. So there is a world, w_1, and a w_1-object, α_1, such that α_1 is identical with α_0 yet is made of slightly different parts. α_1 could also have been made of slightly different parts. So there is a world, w_2, and a w_2-object, α_2, such that α_2 is identical with α_1 but is made of slightly different parts. Continue this process through a hundred or a hundred million stages, making small variations to design as well as to parts, and you will end up with a submarine, or anything else you choose. So there are no limits on how something could be at another world. This is a *reductio ad absurdum* of the view that there are facts of the form: α at w is the same object as β at w' and is distinct from γ at w'.

Quine underplays the paradoxical nature of the argument upon which he relies. We do *not* happily accept that α_0 could have been made of the parts that some remote successor in the series is made of. So we are not happy to accept that "anything can be changed into anything by easy stages". The reasoning is on a par with the reasoning that seems to force us to accept that a heap of sand can never by destroyed by one-by-one removal of grains (for taking away one grain can never turn a heap into a non-heap). The reasoning is powerful, yet we all know that there must be something wrong either with it or with the premises, for we all know that the conclusion is unacceptable.

He also underplays the problem-free nature of identity through time. There are well-known puzzles. For example, we are drawn to a continuity account of identity through time, as is shown by the fact that we allow that a ship that has been endlessly repaired over many years, in gentle stages, is the very same ship, even if in its later years it is composed of none of its original parts. In addition, we are drawn to a compositional account of identity through time, as shown by the fact that if we imagine a ship's parts being successively replaced, but the old parts kept and finally reassembled into a ship, we have *some* inclination (of varying strength depending upon context) to hold that this is really the original ship, the one bearing a continuity relation to the original being merely a replica.[11]

Finally, Quine seems to assume something that these sorts of cases themselves give one reason to doubt: namely, that relations like "could have been made of slightly different parts" are transitive.

[11] Kaplan has suggested a context in which the inclination to take this view is strong: a museum has sent a philosopher to Greece to buy, crate and dispatch the ship of Theseus for reassembly in the museum. As the philosopher removes a plank, he replaces it with a brand-new one of exactly the same shape, so that when he has finished he has an assembled ship of new parts and a dismantled ship of old parts. He sends the latter to the museum. Should the museum director be seriously perturbed when he gets a phone call from the philosopher announcing that he has the *real* ship of Theseus, still in Greece?

Quine can indeed be justly read as presenting a believer in trans-world identity with a challenge to give a systematic account of it, but he cannot be said to have shown that it is any more incoherent than our talk of ships and heaps.

One way to take up the challenge has been proposed by Lewis. He does not, as I said, allow trans-world identity in the sense that Quine intends, though he does provide a substitute that is said to yield everything for which trans-world identity was needed, and is governed by fairly well-articulated principles. This theory is discussed in §**9**.

Kripke has suggested that some of the distrust of trans-world identity is fostered by a faulty picture: thinking of a possible world as like something viewed through the wrong end of a telescope. By contrast, he suggests, the question of whether Socrates could have been an alligator is not to be addressed by envisaging an alligator-infested world, and reviewing the individuals therein to "see" if one of them is Socrates. Rather, it is to be answered by connecting it with other questions, like: must an individual of a species be propagated by individuals of that species? Could anyone have had different parents (propagators) from the ones he actually had? Kripke seems to have in mind an epistemic version of the problem of trans-world identity. He is certainly right to say that possible worlds are not going to *supply* the answers to questions like the one about Socrates. Rather, worlds provide a way of expressing the answers, once found.

A version of the third thesis about de re modality mentioned in §**4** is correct as applied to **QN**: **QN**-formalizations expressing de re necessity do involve, in their **QN**-interpretation, questions of trans-world identity. (We will see in §**9** that there is an alternative approach.) We now very briefly consider the other two theses, (**4.1**) and (**4.2**), viz.:

> In ascribing de re necessity we attribute a property to a non-linguistic object; in ascribing de dicto necessity we attribute a property to a sentence.
>
> A de re necessary truth records how things are in the world; a de dicto necessary truth records only linguistic facts.

De re necessity does indeed involve the ascription of properties to non-linguistic objects, but we have seen no reason at all to suppose that de dicto expressions of necessity attribute it to a sentence. It is a property of *people* that they are, necessarily, unmarried if bachelors. Likewise, though our discussion sits happily with the thought that a de re necessary truth records how things are in the world, we have seen no reason to suppose that a de dicto necessary truth is made true by linguistic facts. There is a prima facie case against any such view. As far as we have seen, the de re/de dicto distinction is merely one of scope, so that the very same concept of necessity is involved in both cases. Hence one would expect it to be true in both cases or neither that they are "made true by linguistic facts".

8 "∀" for "□"

The semantics for **QN** have been given in terms of possible worlds: "□" is associated with universal quantification over worlds, "◊" with existential quantification. The connection suggests that we could have proceeded in a different way: instead of enriching **Q** with the non-truth functional sentence connectives, we could instead have enriched it with a further predicate constant, "W", stipulating that every interpretation assigns to "W" the set of all possible worlds. Let us call the result of enriching **Q** in this way **QW**. We might then hope to express a **QN**-sentence of the form "□ . . ." by a corresponding **QW**-sentence of the form "$\forall x(Wx\rightarrow \ldots x \ldots)$". Let us refer to any approach on these lines as a quantifier treatment of necessity.

Intuitively, the idea is to exploit the equivalence between it being necessarily true that A and it being true at every possible world that A. As soon as we look at the details of **QW**, however, some difficulties arise. What, in **QW**, should correspond to the **QN**-sentence "$\Box p$"? We cannot simply write

1) $\forall x(Wx \rightarrow p)$.

This severs the connection between "p" and the possible worlds: the idea was to say that "p" is *true at* each possible world. There are a number of different tacks one might take, of which I shall consider only two. In both cases, the base language (corresponding to our **Q**) is envisaged to have no sentence-letters, so let us imagine that modification of **Q** to have been made. Our original problem emerges in the same way if we ask: how should we fill the dots in "$\forall x(Wx\rightarrow \ldots x \ldots)$" when formalizing what in **QN** is formalized as "$\Box F\alpha$"?

The first suggestion I shall call the "extra argument place" treatment. Suppose we are to formalize

2) Socrates is necessarily human.

The proposal is that we use

3) $\forall x(Wx \rightarrow F'\alpha x)$

where "$F'xy$" corresponds, not to "human", but to "x is human at y". On an intended **QW**-interpretation, "F'" will be assigned the set of ordered pairs σ such that, for each world, w, and each object o in w, $\langle o,w\rangle$ belongs to σ just on condition that o is human at w.

The idea can be generalized. Every n-ary predicate in an English necessitation will be formalized by an $n+1$-ary predicate-letter of **QW**, the extra argument place being filled by a variable bound by a quantifier in the phrase "$\forall x(Wx\rightarrow \ldots x \ldots)$".

This approach provides an extensional treatment of modality. The extensionality of **QW** follows from the fact that its semantics are in essentials those of **Q**: the interpretation rules attend only to the extensions of expressions. Apparent evidence of non-extensionality disappears. Consider the invalidity of the argument:

4) Necessarily Socrates is human;
Socrates is human iff Socrates is snub-nosed.
Therefore necessarily, Socrates is snub-nosed.

(4) is evidence for non-extensionality only if we can construe the first premise as consisting in the application of a sentence connective to a sentence. On the extra argument place treatment, the premise is not construed in this way, but rather as having the logical form of (3) in which there is no sentential component. The invalidity of (4), as formalized in **QW** on the extra argument place approach, is quite consistent with **QW**'s extensionality.

Ex. 5.21 Provide a **QW**-formalization of (8.4), showing your correspondence scheme.

The approach provides a vivid example of how proposed logical forms may differ from the way a sentence would intuitively be supposed to be constructed. The divergence will hinder **QW** in formalizing as valid intuitively valid arguments. For example, intuitively the following argument is valid:

5) Necessarily Socrates is human. Therefore Socrates is human.

The present **QW** approach cannot, on the face of it, even discern a common constituent in premise and conclusion corresponding to "Socrates is human". "Human" in the premise has to be matched with a 2-place predicate letter, whereas in the conclusion it will, for all that has been said, be matched with a 1-place one.

Ex. 5.22 Provide a QW-formalization of (**8.5**), showing your correspondence scheme. Assuming that **QW** uses the standard notion of validity (paralleling **PN**- and **QN**-validity), is your formalization QW-valid?

To remedy this, we could use a 2-place predicate-letter in formalizing the conclusion, filling the additional argument place by a name for the actual world. Another remedy would be to use "′" to mark a special semantic relation between predicate letters, the interpretations being constrained to behave according to the following rule: if an interpretation assigns to "F" the set of all things meeting a condition, C, then it must assign to "F'" the set of all ordered pairs $<o,w>$ such that o meets condition C with respect to w.

Ex. 5.23 Evaluate the following argument against the suggestion that we should formalize apparently 1-place predicates, like "green", by 2-place ones:

> If we transform "x is green" to "x is green at w*" (where "w*" rigidly names the actual world) we will get the wrong results, for something green will be green-at-w* with respect to every world at which it exists, but will not be green with respect to every such world.

The extra argument place treatment takes all of a thing's properties as really relations between it and a world. You may think that this page

has the intrinsic property of being rectangular, but on the proposed treatment, we have to say that there is no such intrinsic property. To permit **QW**-formalization of, say,

6) This page is rectangular

we regard it as abbreviating something like

7) This page is rectangular-at-w

where the relevant world w is contextually determined (presumably as the actual world in a self-standing utterance of (**6**)). Lewis has criticized this implicit repudiation of intrinsic properties as unjustified. Being white or rectangular are properties which objects have in themselves, and are not relations the objects bear to other things.

Ex. 5.24 Evaluate the following response on behalf of the extra argument place account:

> On my view, to say that rectangularity is a property this page has "in itself" is simply to say that it is a *2-place* property of this page: a property the possession of which relates this page to only one other object, a world. A relational property of this page is one which is more than 2-place.

An alternative way to implement a quantifier approach to modality uses what I shall call "open-sentence formers" (cf. Lewis [1968, 1986a]). Lewis begins by suggesting that a phrase like "In Australia . . ." can be seen as a 1-place sentence operator which works by restricting the quantifiers in what follows. To evaluate

8) In Australia, all beer is good

one need attend only to a proper subset of all the beer there is, viz. that which is in Australia. A phrase like "At the actual world . . ." functions in a similar way, so that an affirmation of

9) At the actual world, all beer is good

is perfectly intelligible, and avoids commitment to the goodness of non-actual beer. (Compare: "As things are, all beer is good; but had the proposed Trans-National Brewery come into being, this would not have been so.")

Lewis's suggestion is that a phrase like "at v", where "v" is a variable ranging over possible worlds, can function in a similar way to "in Australia" and "at the actual world". The main difference is that such a phrase contains a variable of quantification that can fall in the scope of a quantifier, as in

10) $\forall x(\text{at } x(p))$.

Syntactically, "at x" forms from "p" an *open* sentence "at $x(p)$", and hence something from which a quantifier can form a closed sentence. I call any phrase "at v" (v any variable) an "open-sentence former": it takes a sentence, open or closed, to make an open sentence. Semantically, the basic idea is that "at $x(p)$" is true iff p is true with respect to x (a condition which will fail if x is not a world).

Let us take the syntax of **QW** to be enriched by the addition of the open-sentence former "at". How should the interpretation rules be modified? All **QW**-interpretations must be relativized to worlds: a sentence like "at $\alpha(p)$", which will be involved in interpreting a sentence which quantifies over worlds (assuming we leave the quantifier rules unchanged), should be true upon an interpretation iff "p" is true upon an interpretation *with respect to* whatever the interpretation assigns to "α".

To be true to Lewis's ideas, we cannot simply take over the **QW**-interpretation rules as they stand. Lewis wants quantifiers to range over absolutely everything there is, and he takes this to include non-actual objects as well as actual ones. Thus he takes there to be a reading of

11) There are talking donkeys

upon which it is true (with respect to the actual world): a reading which treats the quantifier as ranging over *everything*, actual and possible (cf. §**11** below). Context can implicitly restrict or derestrict quantifiers, and the same can be done explicitly by the "at v" operators. Lewis sees the effect of replacing "are" by "could be" in (**11**) as that of unambiguously ensuring that the quantifier will be completely unrestricted, and thus range over non-actual as well as actual objects.

One way to implement this idea within the present framework is to use two styles of variables and name-letters, one style for quantification restricted by some "at v" operator, and another for unrestricted quantification. A variable of quantification, say "x", occurring within the immediate scope[12] of an "at v" operator needs to be thought of as quantifying over objects in the world assigned to "v". Hence one style of name-letter replacing variables like "x" for the purposes of interpretation requires a corresponding relativity: the interpretation must assign it something in the world assigned to "v". Unrestricted quantifiers are thought of as quantifying over the totality of actual and non-actual objects (including the actual and all non-actual worlds). A corresponding distinct style of name-letter can be used, and to such a name-letter an interpretation is free to assign anything from any world. The quantifier rules will need to use name-letters of the unrestricted sort to mark positions occupied by variables whose quantifier does not fall in the scope of any "at v" operator, and name-letters of the restricted sort to mark positions occupied by variables whose quantifier does fall within the scope of an "at v" operator. The full details are not necessary for present purposes. The general idea is that "at n, X" is true upon an interpretation, i, with respect to a world, w, iff X is true upon i with respect to whatever i assigns to "n".

In **Q**, and therefore in **QW**, the order of quantifiers of the same sort is irrelevant. For example, the prefixes "$\exists x \exists y$" and "$\exists y \exists x$" are equivalent. But "$\exists x \Diamond$" and "$\Diamond \exists x$" are far from equivalent prefixes, the first expressing possibility de re, the second de dicto. (Compare (**7.3**) and (**7.4**).)

The difference is not brought out in **QW** by the following pair of formalizations:

12) $\exists x \exists y(Wy \;\&\; \text{at } y(\ldots x \ldots))$

13) $\exists y \exists x(Wy \;\&\; \text{at } y(\ldots x \ldots))$

for these are equivalent. Rather, Lewis will see the de re English phrase "there is something which could be . . ." as involving an implicit

12 What counts is the "at v" operator closest to the left of the first occurrence of the variable, say "x". The objects relevant to the quantification are those in the world which is assigned to "v", even if "x" subsequently occurs in the scope of some distinct operator "at v'".

restriction of the quantifier to the actual world. To bring this out in **QW**, we can add a name for the actual world, say "w★". (The ru will be that every interpretation, with respect to every world, w, assig the actual world to "w★". "At w★" thus forms closed sentences fro closed sentences.) The de re sentence could be formalized:

14) $\exists x(\text{at w}^\star(\exists y\ y = x)\ \&\ \exists y(\text{W}y\ \&\ (\text{at } y(\ldots x \ldots))))$.

(12) or **(13)** serve to formalize the de dicto sentence. The de re/de dicto distinction is not submerged.

QW (henceforth understood in its open-sentence former version and **QN** use pretty similar semantic resources, despite their syntactic differences. For example, in the semantics for **QN** there is quantification over worlds, including non-actual worlds, and their domains including domains containing non-actual objects. Is there any reason to prefer one of these languages to the other?

One might initially be tempted to suppose that there could be little to choose between the languages. However, Lewis has argued that **QW** has greater expressive resources than **QN**, revealed in the greater depth of the formalizations of English the former can provide.

As **QN** and **QW** stand, it is true that the expressive resources of the latter outstrip the former. One reason for this is connected with the addition made to **QW** of the name "w★". There is no corresponding device in **QN**. The difference can be brought out by considering the following sentence.

15) It is possible that everything that is actually red should also have been shiny.

This is formalizable in **QW** as

16) $\exists x(\text{W}x\ \&\ \forall y(\text{at w}^\star(Fy) \to \text{at } x(Gy)))$,

with "F" corresponding to "is red" and "G" to "is red and shiny". However, **(15)** cannot be adequately **QN**-formalized as either of

17) $\forall x(Fx \to \Diamond Gx)$

18) $\Diamond\forall x(Fx \to Gx)$

with correspondences as before.

Ex. 5.25 Show how the truth-upon-an-intended-interpretation conditions of **(8.17)** and **(8.18)** differ from the truth conditions of **(8.15)**. (See Davies [1981], pp. 220–1.)

A suitable supplementation of **QN** that keeps to the sentence connective approach to modality, and does for **QN** something like what "w★" does for **QW**, is a one-place sentence connective, say "$\boxed{A}$" and the interpretation rule:

19) For any interpretation, i, any world, w, $i_w(\boxed{A}X)$ is T iff X is true upon i with respect to the actual world.

One should then **QN**-formalize **(15)** as

20) $\Diamond\forall x(\boxed{A}Fx \to Gx)$.[13]

The claim that there are non-actual objects can be **QN**-formalized as

21) $\exists x(\Diamond\exists y\ y = x\ \&\ \neg\boxed{A}\exists y\ y = x)$,

or **QW**-formalized as

22) $\exists x\exists y(\text{W}y\ \&\ \text{at } y(\exists z\ z = x)\ \&\ \text{at w}^\star(\neg\exists z\ z = x))$.

If we build in to our definition of **QW**-interpretation that there are non-actual objects, then **(22)** will come out as **QW**-valid; otherwise not. Similarly, one has a choice whether or not so to engineer the interpretation rules of **QN**, in particular one's account of the worlds and their domains, as to make **(21)** valid, or even to make its negation valid. A common prejudice among logicians would be to prefer neutrality on this issue, which might well be seen to belong to "metaphysics".

Lewis gives various examples of natural claims formalizable in **QW** but allegedly not formalizable (to any reasonable depth) in **QN**. If even one allegation is correct, this would constitute a powerful reason for preferring **QW**. I shall consider three of the examples.

[13] Adding an "actually" operator may not the only way to express sentences like **(15)**: see Teichmann [1990].

23) It might happen in three different ways that a donkey talks.

In this example, we appear to talk of, indeed count, various different possibilities. These, says Lewis, can be represented by possible worlds but not by boxes and diamonds. (**23**) is **QW**-formalizable as

24) $\exists x\exists y\exists z(Wx \mathbin{\&} Wy \mathbin{\&} Wz \mathbin{\&} x \neq y \mathbin{\&} x \neq z \mathbin{\&} y \neq z \mathbin{\&} \text{at } x(\exists vFv) \mathbin{\&} \text{at } y(\exists vFv) \mathbin{\&} \text{at } z(\exists vFv))$

with "*F*" corresponding to "is a talking donkey". One might doubt if this is entirely adequate, since the differences between the worlds might be ones irrelevant to the way a donkey talked, whereas intuitively what is required for the truth of (**23**) is, for example, that a donkey might be made to talk by special training, or by the injection of a chemical, or by genetic engineering. However, it may seem that in **QN** one cannot get even as close as this, since there would seem to be no way in which numeral adjectives can be conjured out of boxes and diamonds.

In fact, one can use resources going at most a very little beyond **QN** to produce a formalization of (**23**) that rivals (**24**). The idea is to understand the English as saying that there are three different properties a donkey could have, and a possessor of any of these properties is a talker. I shall treat the quantification over properties in the way that simulates predicate quantification within a first order language, and so minimizes alterations to **QN** (compare the discussion surrounding (**4.19.4**)): I shall assume that properties are among the objects in some (or all) domains of interpretation. A **QN**-formalization is

25) $\exists x\exists y\exists z(Fx \mathbin{\&} Fy \mathbin{\&} Fz \mathbin{\&} x \neq y \mathbin{\&} x \neq z \mathbin{\&} y \neq z \mathbin{\&} \Box\forall v(Gvx \rightarrow Hv) \mathbin{\&} \Box\forall v(Gvy \rightarrow Hv) \mathbin{\&} \Box\forall v(Gvz \rightarrow Hv) \mathbin{\&} \Diamond\exists v(Jv \mathbin{\&} Gvx) \mathbin{\&} \Diamond\exists v(Jv \mathbin{\&} Gvy) \mathbin{\&} \Diamond\exists v(Jv \mathbin{\&} Gvz))$

with "*F*" corresponding to "is a property", "*Gxy*" to "*x* possesses (the property) *y*", "*H*" to "talks" and "*J*" to "is a donkey". (**25**) is better than (**24**), since it connects the different ways with different ways of *being a talker*. It may not be perfect, since the connections between having any one of the properties and being a talker are rather weak, but the fact remains that we have yet to find any expressive superiority of **QW** over **QN**.

This suggestion, like some later ones, could be criticized for its treating properties as among the individuals. This treatment would be wrong if there are no properties. Using this reply as part of a case for preferring **QW** to **QN** (as opposed to a case simply against **QN**) would depend on showing that, though properties do not exist, non-actual worlds do. This would form a controversial basis for a preference for **QW**. A separate point is that the main feature of the suggestion which led to (**25**) would be retained if properties were replaced by sets.

Another example which Lewis uses to ground a preference for **QW** is

26) A red thing could resemble an orange thing more closely than a red thing could resemble a blue thing.

A possible **QN**-formalization is

27) $\Diamond\exists x\exists y\exists z\exists v(Fx \mathbin{\&} Gy \mathbin{\&} Fv \mathbin{\&} Hz \mathbin{\&} Jxyvz)$

with "*F*", "*G*" and "*H*" corresponding to the three colour predicates "red", "orange" and "blue", and "*Jxyvz*" to "*x* resembles *y* more than *v* resembles *z*". Lewis objects that (**27**) wrongly requires there to be a single world containing all the objects, whereas it is enough for the truth of (**26**) that there be two worlds, one containing one pair, another the other. Lewis sums up his case by saying that English essentially involves cross-world comparisons of similarity. His account of (**26**) would suggest the formalization

28) $\exists x\exists y\exists z\exists v(Fx \mathbin{\&} Gy \mathbin{\&} Fv \mathbin{\&} Hz \mathbin{\&} Jxyvz)$,

the force of the English "could" being reflected simply in the fact that the quantifiers are quite unrestricted, each ranging over the totality of actual and non-actual objects.

There is certainly a formal difference between the claim that there could be things *x*, *y*, *z*, *v*, such that . . . and the claim that there could be things *x*, *y* and there could be things *z*, *v*, such that. . . . We could imagine worlds in which the very presence of a red and an orange thing ensures that there is no blue thing, thus preventing the "all in one world" comparison. This would not lead to a conflict with the intuitive truth of (**26**), which requires only one world in which all three colours are exemplified. If there is a world with a closely similar red-orange pair, and a world with a less closely similar red-blue pair, then surely there is *some* world where both pairs co-exist. (Indeed, this conclusion apparently follows from Lewis's own principles determining what worlds exist.) So it is still not clear that there are sentences of idiomatic English whose truth is differentially sensitive to the distinction Lewis makes between within-a-world and cross-world comparisons.

Ex. 5.26 Give the best formalization of the following in (a) **QN** and (b) **QW**:

My car could have been the same colour as yours actually is.

See Forbes [1985], p. 92.

A kind of example upon which Lewis places a good deal of weight are supervenience claims. These have the general form:

29) There could be no differences of one sort without differences of another sort.

A specific supervenience claim is that the mental supervenes upon the physical. In Lewis's words:

> The idea is that the mental supervenes on the physical . . . [i.e.] there could be no mental difference between two people without there being some physical difference, whether intrinsic or extrinsic. Reading the "could" as a diamond, the thesis becomes this: there is no world . . . wherein two people differ mentally without there being some physical difference, whether intrinsic or extrinsic, between them. That is not quite right. We have gratuitously limited our attention to physical differences between two people in the same world, and that means ignoring those extrinsic differences that only ever arise between people in different worlds. ([1986a], p. 16)

"Reading the 'could' as a diamond" goes over into my terminology as "supposing that the English is **QN**-formalizable".

There is no doubt that **QW** can express things which cannot be expressed in **QN**. In **QW** there is explicit quantification over worlds, and resources to count and differentiate distinct possibilities, and these are not explicitly available in **QN**. It may well be that for some philosophical discussions, about essence, origin and substance, for example, these resources are essential. However, the question I am raising here is whether they are required in the understanding of ordinary idiomatic English sentences and our intuitive judgements of whether these are true or false. The ordinary sentences are the neutral territory. The competing theorists are to formalize these in the languages **QW** or **QN**. If the truth-upon-an-intended-interpretation value of one of the theorist's formalizations does not match that of the English sentence (according to our intuitive judgements), that theorist loses a point. (He might regain it if he can convince his opponent that our intuitive judgements are faulty.)

The paragraph quoted from Lewis should identify some feature of the neutral territory not matched by a **QN**-formalization: a mismatch between the truth value of the English and the truth-upon-an-intended-interpretation value of the **QN**-sentence. What we in fact discover is some remarks couched in the non-neutral idioms of Lewis's preferred account, including reference to special features that he attributes to worlds. For example, on his construction of worlds, a person in a Riemannian spacetime cannot inhabit the same world as a person in a Lobachevskian spacetime. Such details, however, do not belong to the neutral territory. Only within the framework for which Lewis is arguing is there a gratuitous limitation of attention in the **QN** approach.

One might base a **QN**-formalization of the supervenience of the mental upon the physical upon the idea that the claim amounts to: necessarily, any things differing in mental properties necessarily differ also in physical ones. This suggests

30) $\Box\forall x\Box\forall y\Box(\exists z(Mz\ \&\ Fxz\ \&\ \neg Fyz) \rightarrow \exists z(Pz\ \&\ Fxz\ \&\ \neg Fyz))$,

where "*M*" corresponds to "is a mental property" "*Fxy*" to "*x* possesses (the property) *y*" and "*P*" to "is a physical property". This appears to be no weaker than the English.

Lewis allows that the alleged failure of **QN** with respect to this case makes little odds, but he suggests that it is more serious with respect to another, structurally similar, thesis, the supervenience of laws. The thesis is that two worlds could not differ in their laws without also differing in local qualitative character. Here is the supposed problem with **QN**-formalization:

> if we read the "could" as a diamond, the thesis in question turns into this: it is not the case that, possibly, two worlds differ in their laws without differing in their distribution of local qualitative character. That's trivial – there is no world wherein two worlds do anything. At any one world W, there is only the single world W. ([1986a], p. 16)

"World" here is playing two roles which should be kept distinct. One role is the special one which Lewis is developing and defending in his book: a world, in this *technical* sense, is the sort or thing quantification over which will be said to translate English modal idioms. As we might put it in our terminology: it is open to the defender of the expressive advantages of **QW** to add whatever further constraints he feels are necessary to his interpretation rule for "W". In the technical sense, it is certainly correct for Lewis to say: "At any one world w, there is only the single world w", for on his construction, for all worlds w, the only world which exists at w is w.

The other role for "world" is non-technical, and it cannot be taken for granted that any Lewisian thesis holds with respect to it. The non-technical sense is that used in stating the thesis that two *worlds* could not differ in their laws without also differing in local qualitative character. A suitable **QN**-formalization is:

31) $\Box\forall x\Box\forall y\Box((Gx \,\&\, Gy) \rightarrow (\exists z(Mz \,\&\, Fxz \,\&\, \neg Fyz) \rightarrow \exists z(Pz \,\&\, Hxz \,\&\, \neg Hyz)))$,

with "G" corresponding to "is a (non-technical) world", "M" to "is a law", "Fxy" to "x is governed by y", "P" to "is a local qualitative property" and "Hxy" to "x possesses (the property) y". No doubt a full appreciation of this thesis will require a more detailed understanding of what sort of entity will be assigned to "G" upon an intended interpretation. But (i) one must not assume that these entities are the possible worlds used in giving the semantics for **QN**; and (ii) a formalization can be entirely adequate even if it gives no analysis of the concepts employed. (One does not criticize the **Q**-formalization of "Socrates is human" as "$F\alpha$" on the grounds that it offers no analysis of the concept of humanity.) (**31**) is far from trivial: there are endless interpretations upon which it is false.

While we have not yet encountered an ordinary English sentence better formalizable in **QW** than in **QN**, it is certainly true that the expressive resources of **QW** outstrip those of **QN**. The existence of the predicate constant, "W", in **QW** but not **QN**, is enough to establish this. (For example, "There are worlds" is adequately formalizable as **QW**-valid, but not as **QN**-valid.) Further, a **QW**-sentence like

32) $\exists w Ww \,\&\, \forall w'(Ww' \rightarrow \forall z \text{ (at } w'(\exists y\ y = z) \rightarrow \text{at } w(\exists y\ y = z)))$

corresponds to an important claim about the structure of worlds (that some world contains all possible and actual individuals), yet has no obvious **QN**-correlate. (Cf. Hazen [1976]; Forbes [1985], pp. 90–1.)

33) $\Diamond\forall x(\Diamond\exists y\ y = x \rightarrow \exists y\ y = x)$

is not an adequate formalization, for it corresponds to the trivial claim that it is possible for all possible existents to exist.

One way to progress is to introduce a device which will enable "$\boxed{A}$" to refer back to what would have been actual if the possibility introduced by "$\Diamond$" were realized. We can achieve this effect in **QN** by indexing the operators. $\boxed{A}X$ will, by default, be true on an interpretation, i, with respect to a world, w, iff X is true upon i with respect to the actual world; but if it is indexed and occurs in the scope of an operator with the same index, $\boxed{A}_n X$ will be true upon an interpretation, i, with respect to any world, w, iff X is true upon i with respect to w', where w' is the world introduced, according to the semantics, by the previous n-indexed operator (cf. Forbes [1985], p. 91, n. 28 for a more accurate specification). Using this device, what corresponds to (**32**) is:

34) $\Diamond_1\forall x(\Diamond\exists y\ y = x \rightarrow \boxed{A}_1\exists y\ y = x)$.

The interpretation of "$\boxed{A}_1$" will pick up the variable attached to the existential quantifier introduced, by the interpretation rules, in the interpretation of "$\Diamond_1$".

These considerations suggest that the framework of **QN** can be extended to increase expressive power. Two questions remain: would there be any point in making these additions to **QN**, given the availability of **QW**? And do these additions keep to the spirit, as well as the letter, of the operator account?

The most promising basis for a positive answer to the first question is that there is a case for saying that **QN** but not **QW** can avoid committing itself to non-actual objects, for example, non-actual worlds. This view will be considered in §**11**.

In answer to the second question, a distinctive feature of quantification is the possibility of back reference that can be achieved by associating quantifiers with variables. If this feature is simply being mirrored by indexing, then it looks as if indexed operators are really quantifiers in all but name. What is certain is that the indexed operators are explicitly linked, by the envisaged rules of interpretation, to variables of quantification. It is natural to conclude that unless these rules mislead, the indices function as variables. If this conclusion is justified, then it would seem that we could also conclude that even the unindexed operators are "really" quantifiers over worlds. The proponent of the suggestion just mentioned, that **QN** can avoid commitment to non-actual entities, will need to speak to this question (see §**11**).

9 Counterpart theory

Lewis's quantifier treatment of modal operators is combined with another distinctive, but theoretically separable, view. He holds (for reasons we will not consider) that nothing exists at more than one world. If no changes were made in **QW**, this doctrine would imply that all de re ascriptions of possibility are false. If Socrates, for example, exists only in our world, there is no world in which he is foolish: he is not foolish at our world, and he does not exist at any other. If we formalized

1) Socrates might have been foolish

in the **QW** style of

2) $\exists w$ at $w(F\alpha)$

we would formalize a truth as a falsehood.

Lewis proposes that de re modalities address the question not of how actual objects are in other worlds, but how their *counterparts* are. A counterpart of an object, o, at a world is something than which nothing in the world resembles o more closely. To say that Socrates might have been foolish is to say that some counterpart of Socrates is foolish: something very like Socrates, or at any rate more like Socrates than anything else in its world, is foolish. "Something" quantifies over all actual and possible objects. Using "C" to express the counterpart relation, (**1**) is formalized:

3) $\exists x(Cx\alpha \mathbin{\&} Fx)$

with "α" corresponding to "Socrates" and "F" to "foolish". Likewise

4) Socrates is necessarily human

is formalized

5) $\forall x(Cx\alpha \rightarrow Fx)$.

Because each counterpart determines a world, we have no need to use the "at w" idiom in these formalizations. We do need to use it in other cases, for example in the formalization of

6) 9 is a necessary existent

which could become

7) $\exists y \forall w$ at w $y = \alpha$

with "α" corresponding to "9". Let us call **QC** the language which results from adding "C" to **QW**.

Ex. 5.27 Provide QC-formalizations of:
(i) Socrates is a contingent being.
(ii) All mathematicians must be mathematicians. (Bring out the possible ambiguity by providing two formalizations of this sentence, one de re and the other de dicto.)

It has been objected to counterpart theory that it fails to represent de re modality as about the right "rem" or thing (cf. Kripke [1972], p. 344n; Plantinga [1974], pp. 115–16; for a criticism, see Hazen [1979]). When we say that Socrates might have been foolish, we mean to speak of Socrates himself, and predicate possible folly of him rather than of someone else who is similar to him. Lewis [1986a] replies, entirely justly, that (3), for example, is about Socrates, and attributes possible folly to him by means of saying that he has a foolish counterpart. To the counterpart it attributes not possible folly, but folly.

Lewis explicitly allows an object in one world to have two counterparts at another. (If this were not allowed, one might wonder whether the counterpart relation differed from identity.) He is committed to this view by founding the counterpart relation upon overall similarity. A red circle and a red square are similar in one respect, and a red circle and a blue circle in another. There is no absolute fact about which other things the red circle is more similar to. This feature of similarity is not exclusive to counterpart theory: it is recorded in the consistency of such mundane beliefs as "x and y are very alike, and also very different".

Lewis can appeal to different respects of similarity to ward off another objection. Plantinga has claimed that, intuitively, Socrates could have been very much like Xenophon actually is, and Xenophon could have been very much like Socrates actually is. We can envisage a situation in which extreme versions of these possibilities both obtain: a world in which Socrates is just like Xenophon actually is and Xenophon is just like Socrates actually is. Suppose we make "F" correspond to some predicate which gives a reasonably comprehensive account of Socrates's features (those that Xenophon would have to possess to be "just like" Socrates), and "G" to some predicate which does the same for Xenophon's features. Then, according to the objection,

8) $\Diamond(G\alpha \,\&\, F\beta)$,

with "α" corresponding to "Socrates" and "β" to "Xenophon", is true whereas a **QC**-formalization is false: since counterparts are determined by similarity, the possessor of the features associated with "G" must be a counterpart of Xenophon and not of Socrates, and the possessor of the features associated with "F" must be a counterpart of Socrates and not of Xenophon.

Lewis replies that the features which Plantinga has in mind, corresponding to F and G, are only some among the similarities that can obtain between people in different worlds. Perhaps they are features of appearance, character and life history. But there are ways in which Socrates can resemble someone who is very different in appearance etc., for example by having similar genes and similar parentage. When we think of Plantinga's situation, we hold these similarities constant, and these are what can sustain a suitable counterpart relation.

Relational possibilities, like

9) Oxford might not have been north of London,

have a single formalization within **QN** ($\Diamond\neg G\alpha\beta$) but there are three possibilities within **QC**, depending on which name introduces a counterpart, or whether both do:

10) $\exists x(Cx\alpha \,\&\, \neg Gx\beta)$

11) $\exists x(Cx\beta \,\&\, \neg G\alpha x)$

12) $\exists x\exists y(Cx\alpha \,\&\, Cy\beta \,\&\, \neg Gxy)$.

Since the counterpart relation is reflexive, each of the first two entails the third. When the topic is a relation like "north of", which, according to Lewis, cannot hold between objects at different worlds, the distinctions between (**10**)–(**12**) correspond to no difference of substance. This may not be the case when the topic is interworld relations like non-identity: relations which can in principle hold between objects which belong to different worlds.

In **QN**, identity is a necessary relation, which implies that

13) $\alpha = \beta \,\&\, \Diamond\alpha \neq \beta$

is false upon every **QN**-interpretation, with respect to every world. One **QC**-formalization of (**13**), modelled on (**12**), is

14) $\alpha = \beta \,\&\, \exists x\exists y(Cx\alpha \,\&\, Cy\beta \,\&\, x \neq y)$.

Suppose that $\alpha = \beta$ and that, as Lewis allows is possible, α has more than one counterpart at some world: two objects both satisfy the condition that nothing at that world is more similar to α than they are. Then (**14**) is true. So there is a conflict between **QN** and **QC**.

Lewis ([1986a], ch. 4.5) holds that the conflict tells in favour of **QN**: identity is contingent, in the sense that there are truths of the form of (**13**) and (**14**). For example, consider a plastic utensils factory. The various utensils are manufactured by filling moulds with the precursors of plastics. The plastic itself is synthesized in the mould, so there is no gap between a certain lump of plastic coming into being and some plastic utensil coming into being. Suppose a bowl made in this way is incinerated a few days later, so that both it and the lump of plastic are destroyed. At every moment of time, both or neither the bowl and the lump of plastic exist, and when they exist, they do so in the same place, weigh the same, and so forth. This gives us reason to hold that the bowl *is* the lump of plastic, so the first conjunct of (**13**) is true. However, it is possible that the factory should have received a different order that morning. Suppose the precursors were already divided up into utensil-sized heaps, and that the heap which in fact became the bowl was instead made into a waste-basket. Suppose that the mould from which the bowl was made went unused that day, but was used on the next day to make a bowl (a bowl, say, fulfilling the special order which our actual bowl fulfilled, so we can properly speak of *the* bowl). The possibility we are describing appears to be one in which the original bowl and the original lump of plastic both exist, but are distinct, the lump of plastic being a different utensil, a waste-basket, and the bowl being a different lump of plastic. If so, this will make the second conjunct of (**13**) true.

In Lewis's scheme, what makes the waste-basket a counterpart of the original bowl is that it is made out of the same stuff; what makes the other bowl a counterpart of the original bowl is that both fulfilled the same order, were made in the same mould, and were the n^{th} bowl to be made by the factory. There are different dimensions of similarity, and this is one way in which there can be more than one counterpart.

This position might seem inconsistent with Leibniz's Law, which can be used to mount an argument for the necessity of identity, the claim that there are no truths of the form of (**13**).

15) (i) Leibniz's Law: identicals have all their properties in common. More formally, every instance of the following (obtained by replacing "**Π**" by any predicate) is true: $\forall x \forall y(x = y \rightarrow (\Pi x \rightarrow \Pi y))$.

(ii) $\forall x \Box x = x$ (assumption).

(iii) $\forall x \forall y(x = y \rightarrow (\Box x = x \rightarrow \Box x = y))$ (from (i), replacing "Π" by "$\Box x =$" (the predicate ascribing the property of being necessarily identical to x).

(iv) $\forall x \forall y(x = y \rightarrow \Box x = y)$ from (ii) and (iii).

This appears to establish quite generally that identicals are necessarily identical. (Cf. Barcan [1947]; also Wiggins [1980b], pp. 109–11 and 214–17.)

In material added to his [1968], Lewis objects to step (**15**iii): he denies that Leibniz's Law is correctly applied. He holds, in effect, that there is no unequivocal property *being necessarily identical to x* applicable now to x, now to y. The justification for this emerges from the counterpart-theoretic representation of (**15**iii). A necessitated relational expression, which would be represented in **QN** as

16) $\Box R\alpha\beta$

can be **QC**-formalized in more than one way (just as in the case of relational possibilities (**10**)–(**12**)); one possibility is

17) $\forall x \forall y((Cx\alpha \ \& \ Cy\beta) \rightarrow Rxy)$.

Applying this to (iii) yields:

18) $\forall x \forall y(x = y \rightarrow (\forall z((Czx \rightarrow z = z) \rightarrow \forall z'((Czx \ \& \ Cz'y) \rightarrow z = z')))$.

This does not result from (i) by substitution of a predicate (the same one on both occurrences) for "Π". On its first occurrence, "$\Pi\zeta$" (I use "ζ" to mark the gap in the predicate, the position to be filled by a name or variable) is replaced by

19) $\forall z(Cz\zeta \rightarrow z = z)$.

On its second occurrence it is replaced by

20) $\forall z'((Cz\zeta \ \& \ Cz'\gamma) \rightarrow z = z')$.

Since (**19**) and (**20**) are distinct, (iii) is not an instance of Leibniz's Law. This suggests that the argument of (**15**) cannot be used as a basis for rejecting counterpart theory. The defence needs some amplification: to accept the formalization of (**16**) as (**17**) we would need to be convinced that there are no cases in which objects have very few counterparts, so that (**17**) holds even when the objects are not necessarily *R*-related. Any qualms on this score can be set to rest by the formalization

21) $\forall w$ at w $\exists x \exists y(Cx\alpha \ \& \ Cy\beta \ \& \ Rxy)$.

This would not affect the point that the application of Leibniz's Law envisaged in (**15**) is counterpart-theoretically invalid.

Ex. 5.28 Discuss from the point of view of counterpart theory whether there is a pair of distinct objects which are necessarily distinct.

Counterpart theory offers rich expressive possibilities. It arises from a metaphysical position, the view that no object can exist at more than one world. From a strictly logical point of view, it cannot be criticized, for its syntax and semantics are essentially that of **Q**. Doubts about counterpart theory arise not from logic but from metaphysics: to lovers of desert landscapes, there is no appeal in an ontology which includes genuinely existing but non-actual worlds and inhabitants thereof.

10 Necessity and vagueness

Vagueness gives rise to borderline cases. Think, for example, of a colour spectrum. There are clear cases of red and clear cases of orange, and in between there are borderline cases: shades which we don't feel inclined to classify either as red or as orange. This feature of vagueness has its analogue in the case of the possible original constitution of artefacts. There are clear cases of possible differences in a thing's original parts; clear cases of impossible differences; and, in between, cases about which we don't know what to say.

It is natural to suppose that, if an area on the spectrum is red, then if you move a tiny distance in either direction, the area you get to must be red too; but if you move a large distance the area may not be red. In other words, a small difference does not matter to the correctness of applying "red" but a large difference does. The first fact amounts to what has been called a "tolerance principle": if two objects differ minutely in shade, then the predicate "red" applies to both or neither. This principle is in tension with the fact that large differences do make a difference to whether "red" is applicable, because you can create large differences out of a number of small differences.

Reasoning that makes this tension manifest, by delivering an explicit contradiction, is called "sorites reasoning", and the contradiction a "sorites paradox". In the case of red, we could construct a sorites argument by naming adjacent areas on a spectrum (imagine the spectrum spread out over a kilometre, and the area strips across it just 1 millimetre wide), $\alpha_0, \alpha_1, \ldots \alpha_n$, where α_0 is a clearly red area, and α_n clearly orange. Using "*F*" to correspond to "is red" we have:

1) $F\alpha_0$
$F\alpha_0 \rightarrow F\alpha_1$
$F\alpha_1 \rightarrow F\alpha_2$
...
$F\alpha_1$ (by modus ponens from the first two premises, followed by successive applications to yield)
$F\alpha_n$.

The conditional premises are licensed by the tolerance principle. Using the fact that what is orange is not red, we get our contradiction from the hypothesis that $\neg F\alpha_n$, which accords with the principle that a big difference does make a difference to the applicability of "red".

It seems that strictly analogous reasoning yields a similar paradox in the case of artefact concepts. The two principles that appear to be in tension are:

2) An artefact could have been constructed out of somewhat different parts from those actually used.

3) An artefact could not have been constructed out of totally different parts from those actually used.

(**3**) as it stands is inconsistent with Lewis's example of the plastic bowl which, he says, might have been made today out of one lot of plastic or tomorrow out of a completely different lot. So let us use "parts" to stand in for all the relevant features of an artefact's construction: not merely the components but also who (if anyone) ordered it, when it was made, with what implements, by whom, according to what design, and so on. It is obvious that it is not possible for there to be an arbitrarily large variation in these features, consistently with it being the very same artefact that gets made.

Consider an artifact α_0, and let "F_0" represent the relevant facts of its actual construction. Let "F_i" represent a property possessed by anything having one part different from anything possessing the property represented by "F_{i-1}". Since (**2**) entails that α_0 could have had one part different, we allow that, possibly, $F_1\alpha_0$. We also allow that had α_0 actually possessed F_1, it could have possessed F_2. This appears to allow us to infer that α_0 could have possessed F_2. This begins a slippery slope, which will have us in the end saying that α_0 could have had F_n, where a possessor of F_n has no part in common with a possessor of F_0. This is inconsistent with (**3**).

The most natural way to formalize the reasoning just informally presented is as follows:

4) $\Diamond F_1\alpha_0$
$F_1\alpha_0 \;\Box\!\!\rightarrow \Diamond F_2\alpha_0$
. . .
$\Diamond F_2\alpha_0$ (from the first two premises, and then further applications to yield)
$\Diamond F_n\alpha_0$.

(**4**) is invalid. Unlike (**1**), the principle used to detach the conclusion is not modus ponens. To show that it is not a truth-preserving principle, suppose we stipulate that a ship$\star$ is something which could have been made only out of a set of parts differing in at most one member from the set it was actually made of, and further stipulate that α is a ship$\star$. We will still hold to the first two premises. We will hold that α could have been F_1, and that if α had been F_1, then it could have been F_2. But the stipulation ensures that α could not have been F_2. (**4**) fails to represent the reasoning as sorites reasoning, for sorites reasoning is valid, according to classical principles of reasoning.[14]

Ex 5.29 Can you find an example in English, not involving the constitution of artefacts, to establish the falsehood of

$\Diamond A,\ A \;\Box\!\!\rightarrow \Diamond B \models \Diamond B$?

Forbes [1985] has suggested that the relevant reasoning can be represented in **QN** as follows:

5) $\Diamond F_1\alpha_0$
$\Diamond F_1\alpha_0 \rightarrow \Diamond F_2\alpha_0$
$\Diamond F_2\alpha_0 \rightarrow \Diamond F_3\alpha_0$
. . .
$\Diamond F_{n-1}\alpha_0 \rightarrow \Diamond F_n\alpha_0$
$\Diamond F_2\alpha_0$ (by modus ponens, followed by $n-2$ similar applications to yield)
. . .
$\Diamond F_n\alpha_0$.

This gives the reasoning precisely the form of (**1**) (replacing "F" by "$\Diamond F$" throughout). The conclusion formalizes "α_0 could have possessed F_n", that is, α_0 could have been made of entirely different parts. It is unclear whether this suggestion as it stands does justice to the original reasoning. (**5**) is, indeed, classically valid, but doubt must attend the truth-upon-an-intended interpretation of "$\Diamond F_1\alpha_0 \rightarrow \Diamond F_2\alpha_0$". At first sight, it seems that nothing in our intuitions commits us to saying that a ship could have differed in original construction by two parts, merely *because* we are committed to saying that it could have differed by one. If we believe a conditional like the second premise of (**5**), it will be because we already accept its consequent. Whereas sorites arguments appear to have true premises, there is room for doubt whether this feature is preserved by the formalization of (**5**).

For (**5**) to appear to be sound, we have to think of the conditionals as sustained by something like the following tolerance principle:

14 Some theorists treat iterated modus ponens as invalid, and so construe sorites reasoning as invalid: see Goguen [1969], or, for a discussion closer to present concerns, Forbes [1985], ch. 7, §3.

6) Small differences in membership of two sets make no difference to the applicability of "α could have been constituted out of . . ." to them.

If this is intuitively acceptable, then (**5**) genuinely constitutes a sorites paradox: it is valid, according to classical principles, has intuitively true premises, and an intuitively false conclusion. Sorites paradoxes are due to vagueness. The question I want to raise is whether in the present case the vagueness attaches to our modal concepts or rather to our artefact concepts. There is a prima facie case for the latter option.

Replace our actual artefact concepts by precise ones, and the paradox disappears. For example, "ship★" is a precise artefact concept: any given ship★ could have been made out of just one different part but not out of more than one different part. The tolerance principle (**6**) will obviously fail for ships★, and so there is no reason to think of an argument with the form of (**5**) as sound. There is now no whiff of a paradox. A complete explanation of the paradoxes has not alluded to special features of modal notions.

In the context of counterpart theory, the paradoxical reasoning does not even get off the ground, as I shall show by formalizing it in QC.

Rather than quantifying over counterparts, I shall assume they are named: α_1 is a counterpart at w_1 of α_0 (which exists at the actual world, w★), α_2 is a counterpart, at w_2, of α_1 (which exists at w_1), and so on. Since each α_i exists at only one world, reference to worlds can be suppressed. The argument is simply:

7) $F_0\alpha_0$
$F_1\alpha_1$
. . .
$F_n\alpha_n$.

The first premise specifies α_0's actual constitution. The second premise tells us that a counterpart of α_0, α_1, is constructed out of slightly different parts. To obtain anything paradoxical, we would need to assume that α_n is a counterpart of α_0. Then the conclusion would represent the claim that α_0 could have been constructed out of completely different parts. But, of course, we have every reason to suppose that α_n is not a counterpart of α_0, since it is completely unlike α_0 in its constitution. In short, we have every reason to think that the counterpart relation is not transitive, for it is based on the non-transitive relation of similarity.

Ex. 5.30 (a) Give an example, not involving modality, which shows that similarity is not a transitive relation.

(b) Evaluate the following argument:

If the counterpart relation were transitive, there would be a paradox. So we should see potential paradox as springing from our modal notions (perhaps incorrectly understood) rather than from our artefact notions.

If there is an F_i such that it is neither clearly the case that α_0 could have been F_i, nor clearly the case that α_0 could not have been F_i, then the counterpart relation needs to be vague. This presents no immediate philosophical problem, since the similarity relation is vague. If we are considering modal claims about the constitution of ships★, the counterpart relation will have to rule that there is a sharp step in the similarity curve: α_1 and α_2 are irrelevantly similar in that they differ by only one part, but relevantly dissimilar in that α_1 but not α_2 is a possible constitution for the ship★ α_0. Hence an appropriate standard of similarity will draw a sharp line between α_1 and α_2. Where there is vagueness in the counterpart relation, its source is our ambition to be faithful to the relevant concepts, like ship as opposed to ship★. It does not appear that modality in and of itself is the source of the vagueness.

The counterpart relation will mirror the vagueness or precision of the concepts, but this does not show that the *source* of the paradoxes of constitution lies with the expressly modal concepts. On the contrary, variations in the counterpart relation are owed to variations in the concepts invoked in the sentences it is used to interpret. This, in turn, is not to deny that, elsewhere, the concept of necessity imports vagueness. Non-indicative conditionals are an example.

11 Metaphysics

In the first four chapters, we managed mostly to steer clear of issues in metaphysics (the theory of being). We touched on the topic when, talking about formalizing empty names and descriptions, we pointed out that certain logical problems would not arise within a Meinongian ontology: one which includes non-existent things. We took no stand but showed how the same logical problems are standardly treated within a non-Meinongian framework. If this treatment is adequate, logic does not at this point determine metaphysics; if it is inadequate, logical matters may well be used to settle metaphysical ones.

In the present chapter, metaphysical problems have arisen at every turn, although we have done our best to avoid addressing them. The time has come to bring some of the most fundamental ones into the open. My aim is to display the connections between the (supposedly purely logical) problems we have been discussing, and metaphysical ones. I make no attempt to resolve the latter. If modality is a branch of logic at all, is not one which can be studied in isolation from the rest of philosophy.

First, realism. You are a realist about a certain subject matter if you think that there really are facts belonging to the subject matter that are not mere artefacts of our thought or language. We are all instinctively realists about rocks and rivers: there they are, and they are what they are whether we think about them or not and irrespective of how we think about them if we do. To the facts about rocks and rivers, the mind makes no contribution. We are all instinctively non-realists about fiction. We do not think that there is a fact, independent of our thought or language, concerning whether the Red Duchess's pepper was black or white (in *Alice Through The Looking-Glass*). If there is such a fact, it was created by Carroll's thoughts. If Carroll's thoughts never turned to the matter, then there is no fact either way.

Realism about modality is the view that there are modal facts, facts of the form "necessarily A" and "possibly A", which are what they are independently of our thought and language. Being a realist about modality involves no further commitment: in particular, it does not pronounce on whether these facts are properly formalized by **QN** or **QW** or **QC** or by some yet other language.

David Lewis uses the phrase "*modal realism*" to describe one particular brand of realism about modality. On his view, (i) modal facts are real and mind-independent, (ii) they are best represented in Lewis's preferred form of quantifier treatment, viz. counterpart theory, and (iii) there "really" are non-actual individuals and worlds. So it is a combination of what I have called realism about modality (i), what I will call "*quantifierism*" (ii), and "non-*actualism*" (iii).

Table 5.7

	Common sense	*Lewis*	*Ersatzism*
Realism about modality	✓	✓	Neutral
Quantifierism	×	✓	✓
Actualism	✓	×	✓
Realism about non-actuals	–	✓	–

Quantifierism is the doctrine that ordinary modal idioms are best represented as quantifications over worlds, and counterpart theory is a special form of quantifierism. Non-actualism is the view that there are things which are not actual. By contrast, an actualist holds that everything is actual. A non-actualist can hold that non-actual possible worlds are non-actual objects; an actualist who in some sense believes in non-actual worlds has to give a special account of what this can mean.

Non-actualists can differ in their opinion about the non-actual things they believe in. Realists about the non-actual hold that such things exist and are mind-independent, just like actual things like rocks and rivers. Non-realists about the non-actual hold that these things are in some way or other mind-dependent: rather than being "out there" they are creatures of our minds.

Lewis's "modal realism" combines realism about modality, quantifierism, non-actualism and realism about the non-actual. Some of the relevant logical space can be described by a diagram (see table 5.7). Common sense combines modal realism with rejection of quantifierism (for the natural expression of modality is by idioms which at least appear to be sentence connectives, and not quantifiers). In addition, common sense affirms that everything that exists actually exists, so it accepts actualism, and the question of realism or some other view with respect to non-actuals does not arise.

The label for the third view, *Ersatzism*, comes from Lewis [1986a]. This is the view, available alike to modal realists and modal anti-realists, that possible worlds and their occupants, even non-actual ones, are relatively familiar actual objects.

One could theoretically combine, without obvious inconsistency, being a non-realist about modality with being a non-actualist, but then one would be committed to being a non-realist about non-actuals. If one were a realist about non-actuals, then one would believe things like "there really is a non-actual possible world such that A, and this is so independently of our thought and talk", and that would commit one to believing that some sentences of the form "possibly A" are really true, independently of our thought and talk.

How could one fail to be a realist about modality? Kant said that experience can tell us how things are, but not that anything must be as it is (cf. Kant [1787], A1; for a criticism, see Kripke [1972], pp. 35 ff., and [1980], pp. 158–60). This line of thought could ground non-realism about modality. All there really is in the world is open to observation, directly or indirectly. If observation cannot tell us how things must be or could be, but only how they actually are, then any necessity or possibility we believe in must have its source in us, in our language, our sensibility or our thought. A traditional favourite, though not Kant's, identifies the necessary with the "*analytic*" – that which is true in virtue of meanings. Supposedly, an example would be "All bachelors are unmarried". That this is *necessary* is no fact about bachelors (so the doctrine goes) but about our language. This form of non-realism about modality has the disadvantage of disallowing the expression of de re modality, since it essentially involves seeing a modal expression as taking an entire sentence as its scope: this is the unit of language from which the necessity is said to derive.

Another form of non-realism about modality takes its cue from a species of non-realism about morality. Should one suppose that moral values are genuine components of the world, or are they artefacts of our responses to the world? If one takes the latter view, one might see the moral modalities – expressed by such sentences as "you *must* do this", "you *ought not* to do that" – as an attempt to objectify what is essentially a subjective response, and to warn that one's own response is something one is likely to act upon by approval, disapproval or whatever. One could extend this view to the modalities we have discussed in this chapter. The force of "$\Box A$", roughly speaking, is to indicate that one will not budge on the question of "A": it's non-negotiable.[15]

[15] There are at least two importantly different versions of the approach I am calling non-realist. One version does, the other does not, see an account of necessity, moral or logical, in terms of subjective responses as showing that our ordinary conceptions are in error. For the contrast, see Blackburn [1986b].

I mention these forms of non-realism about modality not to recommend them, but only to help clarify what is involved in being a realist about modality: rejecting these doctrines.

It might seem that the combination of realism about modality and actualism points one towards **QN** rather than **QW** or **QC** as providing the proper forms for representing modal idioms. One might argue: "**QW** and **QC** involve quantification over non-actual objects, and I cannot accept this because I am an actualist. But I can accept **QN**, because that does not involve quantification over non-actual things. Modality is represented by sentence operators, and not by quantifiers over anything." This argument is flawed in both its parts. First, there is a case for saying that one can combine quantifierism with actualism: this combination is ersatzism. Secondly, it is not obvious that **QN** is really free of non-actuals, because the interpretation rules for **QN** are couched in terms involving non-actual worlds and non-actual denizens of them.

The ersatzist regards "non-actual" worlds as relatively familiar actual objects. For example, he might say that "worlds" are set-theoretic abstractions with the capacity for representing things as they are and also as they are not. What we call "non-actual" worlds are actual set-theoretic abstractions which represent things as being other than as they in fact are. He might say: "I believe that a so-called non-actual world is an actual object, just as a picture is an actual object. It gets called 'non-actual' because the situation it purports to represent does not exist. But the non-actualist thinks that when we say, of a fanciful picture, that it depicts some state of affairs, there really must exist such a state of affairs, if not in this world, then in another. That is a mistake." It is a further question for the ersatzist whether he is a realist or not about (ersatz) worlds. That will depend on whether he thinks that sets, for example, are mind-independent, or the creation of set theorists. This question about realism is distinct both from realism about modality and from realism about non-actuals. It is distinct from the latter, since it arises within a position according to which there are no non-actuals. It is distinct from the former, since the question of realism or non-realism about sets (or whatever one takes worlds to be) is distinct from the question about the reality or otherwise of the facts that sets are used to represent: mind-independent sets could represent mind-dependent modal facts, and mind-dependent sets could represent mind-independent modal facts. If the inference from "there exist nonone would be committed to being a non-realist about non-actuals. If one were a realist about non-actuals, then one would believe things like "there really is a non-actual possible world such that A, and this is so independently of our thought and talk", and that would commit one to believing that some sentences of the form "possibly A" are really true, independently of our thought and talk.

How could one fail to be a realist about modality? Kant said that experience can tell us how things are, but not that anything must be as it is (cf. Kant [1787], A1; for a criticism, see Kripke [1972], pp. 35 ff., and [1980], pp. 158–60). This line of thought could ground non-realism about modality. All there really is in the world is open to observation, directly or indirectly. If observation cannot tell us how things must be or could be, but only how they actually are, then any necessity or possibility we believe in must have its source in us, in our language, our sensibility or our thought. A traditional favourite, though not Kant's, identifies the necessary with the "*analytic*" – that which is true in virtue of meanings. Supposedly, an example would be "All bachelors are unmarried". That this is *necessary* is no fact about bachelors (so the doctrine goes) but about our language. This form of non-realism about modality has the disadvantage of disallowing the expression of de re modality, since it essentially involves seeing a modal expression as taking an entire sentence as its scope: this is the unit of language from which the necessity is said to derive.

Another form of non-realism about modality takes its cue from a species of non-realism about morality. Should one suppose that moral values are genuine components of the world, or are they artefacts of our responses to the world? If one takes the latter view, one might see the moral modalities – expressed by such sentences as "you *must* do this", "you *ought not* to do that" – as an attempt to objectify what is essentially a subjective response, and to warn that one's own response is something one is likely to act upon by approval, disapproval or whatever. One could extend this view to the modalities we have discussed in this chapter. The force of "$\Box A$", roughly speaking, is to indicate that one will not budge on the question of "A": it's non-negotiable.[15]

[15] There are at least two importantly different versions of the approach I am calling non-realist. One version does, the other does not, see an account of necessity, moral or logical, in terms of subjective responses as showing that our ordinary conceptions are in error. For the contrast, see Blackburn [1986b].

I mention these forms of non-realism about modality not to recommend them, but only to help clarify what is involved in being a realist about modality: rejecting these doctrines.

It might seem that the combination of realism about modality and actualism points one towards **QN** rather than **QW** or **QC** as providing the proper forms for representing modal idioms. One might argue: "**QW** and **QC** involve quantification over non-actual objects, and I cannot accept this because I am an actualist. But I can accept **QN**, because that does not involve quantification over non-actual things. Modality is represented by sentence operators, and not by quantifiers over anything." This argument is flawed in both its parts. First, there is a case for saying that one can combine quantifierism with actualism: this combination is ersatzism. Secondly, it is not obvious that **QN** is really free of non-actuals, because the interpretation rules for **QN** are couched in terms involving non-actual worlds and non-actual denizens of them.

The ersatzist regards "non-actual" worlds as relatively familiar actual objects. For example, he might say that "worlds" are set-theoretic abstractions with the capacity for representing things as they are and also as they are not. What we call "non-actual" worlds are actual set-theoretic abstractions which represent things as being other than as they in fact are. He might say: "I believe that a so-called non-actual world is an actual object, just as a picture is an actual object. It gets called 'non-actual' because the situation it purports to represent does not exist. But the non-actualist thinks that when we say, of a fanciful picture, that it depicts some state of affairs, there really must exist such a state of affairs, if not in this world, then in another. That is a mistake." It is a further question for the ersatzist whether he is a realist or not about (ersatz) worlds. That will depend on whether he thinks that sets, for example, are mind-independent, or the creation of set theorists. This question about realism is distinct both from realism about modality and from realism about non-actuals. It is distinct from the latter, since it arises within a position according to which there are no non-actuals. It is distinct from the former, since the question of realism or non-realism about sets (or whatever one takes worlds to be) is distinct from the question about the reality or otherwise of the facts that sets are used to represent: mind-independent sets could represent mind-dependent modal facts, and mind-dependent sets could represent mind-independent modal facts. If the inference from "there exist non-actual worlds" to "there exist non-actual things" is blocked, the ersatzist can consistently accept all the joys in the worlds.

I want to look more closely at the question of whether a realist about modality who is also an actualist can simply dispense altogether with worlds, genuine or ersatz, as the common sense view tries to do. Such a theorist should reject any form of quantifierism, and nail his flag to **QN**. But the question arises: what is all this talk of non-actual worlds and non-actual individuals doing in the metalanguage in which the interpretation rules for **QN** are couched? How could one embrace the language, without embracing its semantics? If interpretation rules specify meanings, then the meaning of "□" according to **QN** is worlds-involving. I envisage two possible lines of reply. One is to take what one might call an "algebraic" view of the semantics for a language. **QN** was introduced for logical purposes, that is, for the purpose of formalizing arguments in order to enable them to be more readily assessed for validity. You might say that the so-called "semantics" for that language are a mere algebraic abstraction, designed to enable a calculation of whether or not an argument is valid. Worlds can be thought of as abstract objects, with no serious connection with thought, so one can take any view one likes about them, without any consequences for the realism one wishes to hold with respect to facts formalizable in **QN**. In short, as with ersatzism, you accept talk of non-actual worlds, but deny that this is talk of non-actual objects.

There is an analogy with ersatzism, but it is only partial. The ersatzist I have envisaged is a quantifierist. This combination would at least strongly suggest that the theorist thinks that what is actually going on when people use modal idioms is that they are quantifying over (ersatz) worlds: the sets (or whatever) are in some way before the minds of the users of the language. The common sense theorist who turns to an algebraic view of semantics to defend his position will deny this. The "worlds" that enter into the semantics are tools of the linguistic theorist, not objects of thought for the user of the language.

The algebraic view of semantics can be illustrated by the truth values in the semantics given in chapter 2 for the language **P**. One could say: the concept of *truth* does not enter into these semantics. The truth values are just a pair of arbitrary objects (many logicians choose the numbers 0 and 1, or even the numerals "0" and "1"). The calculation of validity simply requires the definition that an argument is valid iff

any interpretation assigning a designated one of these objects to all the premises also assigns that object to the conclusion. The nature of these objects is irrelevant to ordinary speech, and nor would any realist or non-realist thesis you might have about them have any special bearing on realism or non-realism with respect to the facts that are P-formalizable.

A quite different, non-algebraic, view of the role of semantics in logic has been adopted here. We have said that truth values do connect with the intuitive notion of truth: *having the truth value true* is simply *being true*. On this view, the semantics for **P** connect closely with other features of truth, for example: it is good to aim at believing something with truth value *true*, and it is good to reason **P**-validly because then you can never go from truth to falsehood.

The other route available to the common sense combination of non-quantifierism and actualism is to claim that the possible world quantifications in the metalanguage for **QN** owe their meaning to their sentence operator counterparts, and so are not "really" quantifiers at all. In justification of this view, one might claim, first, that there is an equivalence between sentence operator idioms like "□" and quantifications over possible worlds. An equivalence, however, is symmetric: it tells you either that the sentence operators are really quantifiers (the conclusion of quantifierism) or that the quantifiers are really sentence operators. Forbes urges us to take the latter course, on the ground that operators have a certain kind of primacy: they are closest to our mother tongue, our natural English modal idioms, and it is they that confer the meaning upon the quantifiers of the metalanguage, rather than the other way about.

A difficulty for this view is that quantifiers in general occur in all sorts of contexts: quantifications over books and rocks and numbers, as well as over worlds. The sense of a quantifier must be to some extent fixed by its occurrence in these other non-world contexts. This ensures that quantifiers do, quite generally, introduce objects, and thus do so when they quantify over worlds. To see the worlds-quantifiers of the metalanguage of **QN** as not really introducing entities would be to see them as quite separate idioms from our ordinary "all" and "some" (or "∀" and "∃"). It is open to the semantic theorist to stipulate that his metalanguage quantifiers, used in the semantics of **QN**, are just the ordinary, object-introducing ones. By contrast, it is of dubious intelligibility to claim, as Forbes does, that the metalanguage quantifiers are a special idiom, owing their sense to the operator counterparts in the object language.

Not every kind of semantics for a language with the operator syntax of **QN** involve quantifiers. Just as one could specify the contribution of "not" by a sentence like

1) A sentence "not-A" is true iff it is not the case that "A" is true

so one could, arguably, specify the contribution of "□" or "necessarily" by a sentence like

2) A sentence "□A" is true iff it is necessarily the case that "A" is true.

Whether such semantics will serve the purpose of defining validity is another question, and a disputed one.[16]

Finally, I turn to a second more or less metaphysical issue: analysis. Those looking for a reductive explanation, or "analysis", of what modal concepts mean will not have found even the beginnings of one in this chapter. A reductive explanation of a concept is one any part of which you can fully understand without yet understanding the concept to be explained. However, to be told that "necessarily A" is true iff "A" is true in all possible worlds is circular, relative to reductive aims, because the account makes free use of the notion of "possible", and "necessary" and "possible" are interdefinable: if we already understood "possible" we would not *need* the account, and if we did not understand it the account would be *useless*.

It was no part of our aim at any point to provide a reductive explanation. Exactly the same objection of circularity could be levelled at the account of "all" and "∀", if that were supposed to be reductive. We start with English "all"-sentences. We formalize them using "∀" and explain the meaning of "∀" by giving an interpretation rule which crucially involves "all". Were this an exercise in analysis, it would be unrewardingly circular. It was not such an exercise, but rather an attempt to fashion an artificial language in which the notion of validity would be more accessible to theory than it is in ordinary English. If that is the modest aim of introducing a language like **QN**, the alleged "circularity" is beside the point.

[16] One aspect of the dispute involves the relationship between truth theoretic semantics (as exemplified in (1) and (2)) and model theoretic semantics – the style of semantics in terms of interpretations provided throughout this book. Cf. Evans [1976].

One candidate for a reductive account of modality is that offered by David Lewis [1986a]. He says that one can say what a possible world is without using any modal notions: a possible world is a sum of (not necessarily actual) objects linked by space–time relations. The explanation essentially involves non-actualism. If one found the explanation attractive, it might help one overcome the commonsensical appeal of actualism.

Bibliographical notes

§1

Gamut [1991] (vol. 2, chs 1–3) provides good discussions of most of the topics of this chapter. For a comprehensive and fairly formal introduction to propositional modal logic, see Chellas [1980]; for a formal introduction to predicate modal logic, see Hughes and Cresswell [1996]. For a philosophical account of modality, which includes both formal semantics and arguments for many substantive essentialist claims, see Forbes [1985]; and, for critical discussions of this, Mackie [1987] and Edgington [1988]. Lewis's most famous use of "ways things could have been" as an introduction to possible worlds is in his [1973b], ch. 4.1, p. 84; for criticism, see McGinn [1981]. For Lewis's more recent view on possible worlds, see Lewis [1986a].

§2

For taxonomy (the distinction between indicatives and non-indicatives), see Dudman, esp. [1984a], Edgington [1991], Bennett [1995]. The earliest presentation of Lewis's theory is Lewis [1973a]; see also Lewis [1973b], [1979] (reprinted with a postscript in Lewis [1986b]). For Stalnaker's theory, see his [1975]; and for his response to Lewis's criticisms, see, for example, his [1984], ch. 7. For general discussion, see Edgington [1995].

§3

The historical source of the accessibility relation is Kripke [1963]. See also Chellas [1980] and Bull and Segerberg [1983].

§4

In the empiricist tradition, necessity was held to be de dicto, and coextensive with analyticity. For an expression of this view, see Quinton [1963]. For distinctions between necessary, analytic and apriori see Kripke [1972]. These lectures were responsible for a considerable revival of interest in de re necessity. The present formulation of the contrast between de re and de dicto derives from Forbes [1985], p. 48.

§5

For Quine's arguments, see his [1953c], [1953d] and [1960] §41. For criticism, see Plantinga [1974], App., pp. 222–51, and Linsky [1977], ch. 6; for detailed discussion see Neale [2000].

§6

For Quine's version, see his [1960], pp. 148–9. Davidson uses the argument in a number of places, for example [1967c], pp. 153. The account I follow most closely is Davies [1981], pp. 210–11. See also Neale [1995].

§7

For discussions of the problems of trans-world identity, see Plantinga [1974], pp. 88– Kaplan [1979], Forbes [1985], esp. chs 3 and 7, Van Inwagen [1985], Fine [1985] and Lewis [1986a], pp. 210–20. For a discussion of the linguistic theory of necessity, see Pap [1958], esp. part II, ch. 7, and Van Fraassen [1977].

§8

See Lewis [1968] and [1986a]. In the latter, see pp. 5–20 for the introduction of the open-sentence formers, and pp. 199–202 for the case against treating intrinsic properties as relations. (NB: although I use this to attack the extra argument place theory, in Lewis's book the argument has a different target.) Lewis, in both places, argues for counterpart theory (see §9). There are two separable questions: whether to adopt any quantifier treatment of modality, and whether to adopt the specific form of quantifier treatment embodied in counterpart theory. See also Hazen [1976], Davies [1981], ch. 9, esp. §1, Forbes [1985], pp. 89–95 and Melia [1992].

§9

The main texts for counterpart theory are, again, Lewis [1968] and [1986a]. See also Mondadori [1983], Forbes [1985], esp. chs 3.4 and 3.5, and App. 3, Ramachandran [1989]. For discussions of contingent identity, see Gibbard [1975], Wiggins [1980], and, especially, Forbes [1985].

§10

Lewis's remark about the non-transitivity of the counterpart relation shows that the paradox will not arise in his scheme, but it does not give any details of the semantic mechanisms. For such details, see Forbes [1985], ch. 7. For some current work on essentiality of origins, see Robertson [1998].

§11

For Lewis's views, see his [1986a]. See also Forbes [1985], esp. ch. 4, Kripke [1972] and esp. [1980] and Chikara [1998]. Non-realism about modality is endorsed by, for example, Wittgenstein [1921], Ayer [1936], and, in the form that necessary truths reflect merely human conventions, is famously opposed by Quine [1936]. A new twist has been given by Blackburn in various writings, e.g. [1986b]. See also: Wright [1986] and [1989], Craig [1985], Hale [1989]. Ersatzism is argued for in Plantinga [1974], ch. 8 and [1976] (though in this article Plantinga's target is the view that there are non-existent objects, rather than the view that there are non-actual objects). An argument against ersatzism by Lewis is discussed in Van Inwagen [1986]. For a criticism of Lewis's [1973b] argument for the reality of non-actuals, see McGinn [1981].

6

The project of formalization

1 Logical versus grammatical form

In chapter **1** I gave some preliminary motivations for studying validity through the medium of artificial languages. In subsequent chapters I presented some of these languages, indicating how they could be used to formalize arguments expressed in English, and in many cases illustrating detailed limitations. We now have to raise our heads from the trees, and try to discern the overall character of the wood.

The main subject of my discussion is the view that formalizing a sentence or argument of a natural language in one of the artificial languages we have discussed reveals something about the nature of the natural language, something that would otherwise be apt to remain hidden. It is this revelation which justifies the efforts we have expended in formalizing. **Q** has been especially favoured by proponents of this view, so it will occupy centre stage.

The revealed must look at least superficially different from the concealed, or there could be no revelation. Here are some apparent differences between natural language sentences and their formalizations.

1) Quantifiers: Some English universal (existential) quantifications not containing an occurrence of "if" ("and") are formalized by a **Q**-sentence containing "→" ("&").

2) Adjectives: English adjectival modification is formalized by **Q**-conjunction.

3) Descriptions: An English sentence of the form "The *F* is *G*" is subject-predicate but its **Q**-formalization is an existential quantification, containing constants like "→" and "=" whose correlates are not visible in the English.

4) "Exists" is a predicate in English but must often be matched by a quantifier in a **Q**-formalization.

5) Numeral adjectives: An English sentence like "Three men are at the door" contains neither an existential quantifier nor the identity sign, but its **Q**-formalization contains "∃" and "=".

6) Verbs of action: An English sentence like "John walks" is subject-predicate but its **Q**-formalization (on Davidson's proposal) is an existential quantification.

7) Adverbs: Adverbial modification is (on Davidson's proposal) formalized by **Q**-conjunction.

8) Propositional attitudes: An English sentence like "John believes that the earth is flat" is a single complex sentence, but its **Q**-formalization (on Davidson's proposal) is two separate sentences.

9) Necessity: "Socrates is necessarily human" is in English a necessitated atom, whereas its **QN**-formalization is the necessitation of a conditional, containing the extraneous concept of existence.

10) Counterpart theory: "Socrates is necessarily human" is in English a necessitated atom, whereas its **QC**-formalization is a universally quantified conditional, containing the extraneous concept of a counterpart.

The confident classifications of English sentences (e.g. the assertion that an English sentence of the form "The *F* is *G*" is subject-predicate) are intended to reflect, not a theoretically grounded view, but some sort of "intuition". The tradition I am describing has it that formalization shows many of our intuitive classifications of natural sentences to be incorrect: universal quantifications in English are shown, by formalization, to be "really" quantified conditionals, definite description sentences are shown to be "really" existential quantifications; and so on. These facts are concealed from the naive and untrained eye, which sees only "grammatical form", but they are revealed by formalization, and

these revelations are the project's main contribution to understanding natural language.

The thesis is sometimes expressed in the slogan: "Grammatical form misleads as to logical form" (cf. Strawson [1952], p. 51). The slogan may fail to capture the thesis. Using "logical form", as we are, to mean a sentence in some favoured artificial language, the truth of the slogan simply points up the existence of divergences like those noted in (1)–(10). The thesis, however, requires the further point that the divergences are only superficial, for *at bottom* the natural sentences have the features that are so readily visible in their logical forms. The slogan does more justice to the thesis upon a different interpretation of "logical form", according to which the phrase signifies not a formalization into some possibly alien language, but the intrinsic logical and semantic properties of the sentence. The thesis is that a natural sentence's logical form in my sense reveals its logical form in this other sense.

In the rest of this section, I shall elaborate some more precise versions of the traditional thesis about logical form. First, however, I want to emphasize how strange the thesis should initially appear. Nothing about our own language seems clearer than that "All men are happy" does not contain an expression for conditionality, or that "Shem kicked Shaun" does not contain an existential quantifier. Could one mitigate the strangeness of the thesis by likening logical theory to scientific theory? Perhaps logic, like physics, needs idealizations, and perhaps the discrepancies we are discussing are the differences between the ideal and the messier real.

Suppose we are concerned with the motion of real billiard balls on a real billiard table. We may find it convenient to make simplifying assumptions: the table is completely flat and frictionless, the balls are perfectly elastic, there is zero resistance from air and baize, and so on. We can with justice say that we are still studying the original concrete phenomena through the idealization. Laws statable in terms of the idealization can be applied to the real balls to yield fairly accurate (if not perfect) predictions of their motions. We could at any point achieve greater accuracy by removing some of the simplifying assumptions. No doubt we would do this until the cost of the additional time taken to solve the equations exceeded the benefit of extra accuracy, a balance that would depend on idiosyncratic needs and interests. There would be nothing sacrosanct about any one idealization, and there would be absolutely no temptation whatsoever, on finding, for example, that the assumption of perfect elasticity yielded adequately accurate predictions, to infer that the real balls are "really" perfectly elastic. The approach builds in the acknowledgement that the idealization differs from the phenomena.

The phenomena for logical theory are arguments in a natural language, say English, and the theory should pick out their validity-relevant features. An idealization can properly abstract from other features, for example from the actual mechanisms whereby specific truth conditions are expressed. Regarded as idealizations, formalizations might not capture all the features of English, but they should enable one to give reasonably accurate predictions of the validity of English arguments. A divergence, for example a case in which an intuitively valid argument in English is formalized as invalid in some artificial language, may simply reflect, what we knew already, that the idealization does not exactly correspond to the phenomena. Alternatively, it may lead us to reconsider our intuitive judgement, though a divergence alone could never be a good reason to abandon such a judgement.

While one could not reasonably quarrel with this approach to validity in natural language, it is not one which yields the distinctive theses of the tradition I have in mind. The approach would not license the attribution to English sentences of all the features of the idealization. For example, to say that "All men are happy" is "really" a quantified conditional would be as ludicrous as saying that billiard balls are "really" perfectly elastic.

To understand the traditional conception of logical form, we need to distinguish variants. Let us start by making a very rough distinction, to be refined in §4, between *logical* features of a sentence or argument and *semantic* features. A prominent logical feature of a sentence would be its logical constants, and the pattern of occurrence of the non-logical expressions. Logical features would be those relevant to validity, or at least formal validity. Semantic features would include logical features, but would also include any other features pertaining to the meaning of the sentence and the words which compose it. We can distinguish two groups of theses about logical form: those which speak only to logic, and those which speak also to semantics. I shall begin by discussing a series of theses ((**11**), (**14**) and (**15**)) which fall into the latter category.

(11) The logical form (i.e. adequate and deep formalization) of a sentence of a natural language gives a *representation of its truth conditions*.

We imposed the following condition of adequacy upon formalization: a formalization is adequate iff the recovered sentence or argument agrees with the original in truth conditions; iff the truth-upon-an-intended-interpretation conditions of the formalization agree with the truth conditions of the original. This ensures the truth of (**11**). Some of the most famous claims about the truth conditions of English sentences, like Russell's claim about the truth conditions of sentences containing definite descriptions, have arisen within the context of the project of formalization.

(11) has no exotic consequences for the real logical nature of the natural language sentences. In general, two sentences may agree in truth conditions, but in other respects be quite unlike, for example the pair:

(12) Either snow is white or it is not

(13) If snow is white it is white.

The identity of the truth conditions gives us no reason to say that (**12**) is "really" a conditional, or that (**13**) is "really" a disjunction. The fact that a natural sentence's logical form matches the sentence in point of truth (-upon-an-interpretation) conditions does not in itself show that there is any other interesting relation between the two. Sameness of truth conditions is symmetric. It accordingly gives as much reason for thinking that the arrow in a Q-formalization of an English quantification is not "really" there as for thinking that the English "really" contains an (invisible) "if".[1]

The following strengthening of (**11**) would introduce an asymmetry:

(14) The logical form of a sentence of a natural language gives a *perspicuous* representation of its truth conditions.

[1] I use "truth conditions" in the way explained in chapter **1.9**: the truth conditions of a sentence are the actual or possible circumstances in which it is or would be true. There is, however, another usage, according to which the truth condition (singular) of a sentence is its meaning.

Perspicuity is normally defined as the co-incidence of syntax and semantics. To attain the definition, we need to define syntactic and semantic categories. Let us say that two expressions belong to the same *syntactic category* iff for every sentence containing one, the result of replacing it by the other is a meaningful sentence. In **Q**, the syntactic categories include the following:

unary sentence connectives: ¬
binary sentence connectives: &, ∨, → and ↔
name-letters
1-place predicate-letters
2-place predicates (=) and predicate-letters
3-place etc.
. . .
quantifiers: ∀, ∃
sentences (including sentence-letters)

Let us say that a *semantic category* is determined by the following test: expressions e_1 and e_2 belong to the same semantic category iff either they are assigned the same kind of entity by the interpretation rules, or else they are treated by the same kind of interpretation rule. In virtue of the first disjunct, all and only name-letters belong to one single semantic category, since they, and only they, are unrestrictedly assigned members of the domain; likewise all and only sentences belong to one single semantic category, since they, and only they, are invariably assigned truth values; likewise predicate-letters of any given degree, say, 3, belong to a single semantic category since they, and only they, are all assigned a set of triples of members of the domain. The second disjunct is intended to be read in such a way that all binary sentence connectives count as belonging to a single category, since they, and only they, are given rules of interpretation which fix a truth value for the resultant sentence on the basis of the truth values of two components. (We could have managed with just the first, more specific, disjunct in the definition had we required that an interpretation assign n-ary truth functions to each n-ary connective, and also made a suitable assignment of kinds of functional entity to the quantifiers.)

A language is *perspicuous* iff its semantic and syntactic categories coincide; that is, iff for any syntactic category there is a semantic one containing just the same things; and vice versa. **Q** comes out as perspicuous. This is not surprising: **Q** was devised with perspicuity in mind; and my account of the semantics and syntax of **Q** was guided by the wish that it should count as perspicuous. Otherwise, I might have counted a semantic category for each interpretation rule, corresponding to the thought that members of a semantic category have "the same kind of meaning", but this would have placed "&" and "∨" in different semantic categories, even though they are in the same syntactic category. I could also have defined syntactic category differently, stipulating one syntactic category for each clause in the specification of a **Q**-sentence (**4.1.6**). This would have placed the sentence connectives in different syntactic categories, but would have eliminated the connection with the idea of a syntactic category as one fixed by substitution with preservation of meaningfulness. This, in turn, would have made it unclear whether we could apply the notion of syntactic category to English; or rather, would have made it unclear that we could apply it to English in the rather casual way characteristic of the logical form tradition, and in advance of having, for English, the analogue of the recursive specification of a **Q**-sentence.

An alleged example of the non-perspicuity of English is that names and quantifier phrases belong to the same syntactic category but different semantic categories. The semantic difference is fairly uncontroversial: the semantics for a name like "Carter" require it to refer to a particular man, but this is not so for a quantifier phrase like "no one". The sameness of syntactic category is harder to establish, as it requires looking at every possible context. True, both "Clinton is a bachelor" and "No one is a bachelor" (**1.12.10** and **11**) are meaningful, and the one results from the other by replacing a name by a quantifier phrase. But to establish that "Clinton" and "no one" belong to the same syntactic category, it would be necessary also to consider sentences like "No one ever complains" (cf. Ex. **1.30**), "Clinton never complains", "No one with back ache complains" and many more (cf. Oliver [1999], pp. 253–4). These show decisively that names and quantifier phrases do not belong to the same syntactic category by the criterion under consideration.

Ex. 6.1 Refute by example the following claim:
"No one" is an exception. Other quantifier phrases do belong to the same syntactic category as names.

For a more promising example of sameness of syntactic category with difference of semantic category we might turn to names like "Carter" and definite descriptions like "The current US President". The semantic difference is, in the first instance, that "Carter" is semantically simple, whereas definite descriptions are semantically complex. If a view considered in chapter **4.12** is correct, the semantic differences are considerably more far-reaching, surfacing (in the apparatus of this book) as the suitability of names, but the unsuitability of descriptions, for formalization by a name-letter. Sameness of syntactic category may initially look more promising than with names and quantifier phrases, but in the end it, too, proves impossible to establish: "No Carter ever told a lie", but not "No the current US President ever told a lie"; "The remarkable and talented Carter addressed the Senate", but not "The remarkable and talented the current US President addressed the Senate"; and many others (cf. Oliver [1999], pp. 253–4). One is forced to conclude that, at least on this account of syntactic category, the logical form tradition has been too quick to brand English as non-perspicuous.

The motivation for the branding is that perspicuity is connected with two features prized by logicians. One can fairly easily devise *rules of proof* for a language only if it is perspicuous; these rules, stated in terms of the physical make-up of sentences, would specify patterns of derivations of sentences from others in a way which mirrors validity. The other prized feature is that one can fairly easily devise *rules of interpretation* for a language only if it is perspicuous; these rules would apply to sentences specified in terms of their physical make-up, delivering for each a truth (-upon-an-interpretation) condition. For languages like Q we possess rules of both kinds; we do not possess comprehensive rules of either kind for English. Although it clearly does not follow that English is not perspicuous, the difficulty of devising such rules for English has, I suspect, encouraged this conclusion. In order to continue our exploration of the logical form tradition, we will need at least provisionally to accept the non-perspicuity of natural languages.

Ex. 6.2 Show why the non-perspicuity of English does not follow from the claims about the ease of devising rules. (You might find it helpful to formalize the premises, using "→".)

Given this provisional acceptance, thesis (**14**) specifies a relation which, unlike sameness of truth conditions, is not symmetric: a sentence of an artificial language like **Q** is a sentence of a perspicuous language, whereas its equivalent in a natural language is not. However, there is no inference from this to some more exotic thesis, expressible by such remarks as that English universal quantifications are really universally quantified conditionals. Given that English is not perspicuous, there must be some vital difference between the way in which an English sentence and its formalization present their truth conditions. Indeed, there would be something paradoxical in the assertion that those features of the artificial languages which make the formulation of rules of proof and interpretation relatively easy are really present in the English sentences, for which the formulation of such rules is hard or impossible.

We get closer to what many people have had in mind by the idea of logical form in the following strengthening of (**14**):

(15) The logical form of a sentence of a natural language gives a perspicuous and *systematic* representation of its truth conditions.

The notion of a systematic representation, which will be refined in §6 below, is linked to the notion of *compositional semantics*. The rough idea is as follows. The meaning of a sentence is determined by the meanings of the words of which it is composed, and by their manner of composition. Compositional semantics for a language will specify word meanings, and the semantic import of modes of composition, in such a way that, from these specifications, the meaning of any sentence in the language can be derived. As with meaning, so with truth conditions. Setting aside ambiguity, a word makes the same contribution to the truth conditions of any sentence in which it occurs, and a compositional semantics will specify this contribution. A systematic representation of truth conditions, alluded to in (**15**), is a representation within the perspective of a compositional semantics. On this view, a sentence's logical form will contribute to an understanding of how the words in the sentence contribute systematically to the truth conditions of the whole.

Thesis (**15**) offers an important aspiration. However, it is at first sight hard to see how its achievement would be consistent with some of the specific classical claims about logical form. For example, on the face of it, one who aspires to give a systematic representation of the truth conditions of universal quantifications in English should not pretend that such quantifications standardly contain an expression for conditionality.

I return to thesis (**15**) in §6. (**15**), in common with its weaker predecessors, (**11**) and (**14**), concerned semantic features. I now want to introduce two theses relating specifically to logical features:

16) If an argument in natural language is adequately formalizable as valid, then it is *formally* valid.

17) The formalization of a natural sentence renders the proposition it expresses accessible to formal deductive manipulation.

(**16**), which I discuss more fully in §**2**, does not entail that the nature of the features responsible for formal validity is intrinsically characterized by the logical form. For example, the **Q**-validity of the formalization of, say, "All men are mortal, Socrates is a man, therefore Socrates is mortal" assures us, according to (**16**), that the argument is formally valid, but gives us no assurance that it is valid in virtue of having a quantified conditional as its first premise. The English will have a formal feature corresponding to being a universally quantified conditional, but (**16**) makes no commitment to the intrinsic nature of that feature. Such a position requires that one have a standard of formal validity independent of formalization. This question is discussed in §5. The notion of deductive manipulation, required by (**17**), is discussed in §**3**.

In the next section, I look briefly at a very ambitious version of the thesis of the revealingness of logical forms.

2 Analysis and the Tractarian vision

Russell toyed with the idea, and Wittgenstein embraced it in the *Tractatus*, that all validity is formal validity: indeed, is validity in virtue of **Q**-logical form. Contrary appearances are to be explained by the fact that our language is not fully analysed. Analysis would reveal all validity as formal. If we find a valid argument that apparently does not formalize as **Q**-valid, this shows that there is some defect in our formalization: we have not uncovered enough logical structure, we have not carried analysis far enough.

First, a historical correction. The language of logical forms that Russell had in mind was not **Q**, but the richer language of *Principia Mathematica*. The language of logical forms that Wittgenstein had in mind was also not **Q**, but a language that at least superficially seems less rich. The differences between them, and between them and **Q**, had no impact either on the basis of the vision or on its subsequent rejection by both philosophers.[2]

Let us see how this *Tractarian vision* would view a very simple case. A standard example of a valid argument which is not formally valid is:

1) Tom is a bachelor, so Tom is unmarried.

One might suppose that the deepest **Q**-formalization is:

2) $F\alpha; \neg G\alpha$

with "α" corresponding to "Tom", "F" to "is a bachelor" and "G" to "is married". However, the Tractarian view has it that we have overlooked some hidden structure in the English. We can reveal it, and at the same time show (**1**) to be formally valid, by the formalization

3) $F\alpha \,\&\, \neg G\alpha; \neg G\alpha$

with correspondences as before except that "F" now corresponds to "is a man". The formalization depends upon the analysis of "is a bachelor" as "is a man and is not married".

(**3**) counts as an adequate formalization of (**1**) by the standards previously given, and nothing we have so far said gives grounds for objecting to it. These standards are consistent with intuitively more unacceptable proposals. In general, if an argument A; C is valid, our current standard of adequacy counts as adequate a formalization "$p \,\&\, q$; q". For if A entails C, then "A" and "$A \,\&\, C$" have the same truth conditions. In this sense it is trivial that all validity can be represented as formal validity, and the maxim of formalizing in such a way as to

2 The Tractarian dream was dreamed before the notion of completeness for logical systems was available. No doubt the incompleteness of second order logic would have affected Russell's version of the dream. (A consequence of incompleteness is that some valid sentences would be formalized by unprovable ones.)

maximize the amount of validity that can be represented might lead to no more than this trivial result.

Russell always had doubts about the possibility of representing all validity as formal, and although Wittgenstein explicitly affirmed this position in the *Tractatus*, he explicitly retracted it later. For both of them, the decisive factor was that the vision was unrealizable because there are valid arguments which no amount of analysis can represent as formally valid. They were both influenced by the case of colour incompatibilities. Thus the valid argument

4) This is red (all over), so it is not blue (all over)

cannot be represented as formally valid, however one analyses. One might try analysing "red" as "not yellow and not blue and not etc." But first, it is doubtful whether the analysis can be brought to a conclusion, and secondly it would involve treating "yellow" etc. as primitive (not to be analysed) and a similar valid argument could be stated using just these primitives.[3]

Contemporary opinion would reject the Tractarian vision not just for this kind of reason, but also on the grounds that it fails to distinguish between a claim about logical form and a claim about analysis. Davidson, among others, has emphasized the distinction ([1967a], pp. 31 ff.; [1970a]). If it can be made good, it puts a curb on the Tractarian vision. Something like (**3**) would be said to be unsatisfactory as a logical form of (**1**) even if satisfactory as an analysis.

The distinction between analysis and form is quite intuitive. For example, a traditional "analysis of knowledge" might start off on lines such as these:

5) *S* knows that *A* iff
 (i) *A*
 (ii) *S* believes that *A*
 (iii) . . . .

The rule of the game is that you do not re-use the word "knows" in the numbered clauses after the "iff". The clauses thus in some sense

[3] It is presupposed that an expression has its analysis once and for all. Otherwise, the second point could be evaded by proposing different sets of primitives for different cases.

explain the meaning of "knows". This does not even begin to look like a claim about logical form. By contrast, Davidson's account of propositional attitudes rules that the logical form of "*S* knows that *A*" is:

6) $F\alpha\beta, p$

(with "*Fxy*" corresponding to "*x* knows *y*", "α" to "*S*", "β" to "that" and "*p*" to "*A*"). This does not even begin to look like an analysis of knowledge. There is no attempt, in the logical form, to avoid the word "knows". Davidson's thought is that we must first get the logical form straight, leaving analysis as a separate issue.

Russell and Wittgenstein would not have been moved by this distinction. In the relevant period, both philosophers were primarily interested in thought or judgement rather than in language. For them, the structure revealed by logical form is the structure of thought, and this is what is so often hidden by natural language. Revealing the thought expressed by the premise of (**1**) as a conjunction is part and parcel of the very same enterprise as revealing the thought expressed by "All men are happy" as a quantified conditional. Once you become hardened to, or even rejoice in, the marked differences between the logical structure of thought and the grammatical structure of natural language, one sort of difference is likely to seem much like another. Language condenses complex structures in various ways: hiding the conjunction really present in the thought expressed by (**1**) is no different in kind from hiding the conditionality really present in the thought expressed by "All men are happy". In both cases, it is the job of "philosophical logic" to reveal what is hidden.

Davidson's own basis for the distinction between analysis and logical form depends upon considerations rather foreign to the discussion so far, considerations which will be introduced in §**6** below. In the remainder of this section, I want to consider whether a well-grounded curb on the Tractarian vision can emerge from the kind of perspective we have adopted. In particular, the fact that (**3**) counts as adequate by our present standards may seem to constitute a criticism of these standards. The view for which I shall argue is that curbing the Tractarian vision on the basis of ideas we have already to hand is possible only by disqualifying, as inadequate, some standardly accepted Q-formalizations of English.

One way to motivate the distinction between logical form and analysis is to say that the logical form of a sentence should specify its logical constants, and the way in which they relate to the pattern of occurrence of the non-logical expressions. Analysis, by contrast, should concern the contribution to meaning of some expression, ideally providing a more complex equivalent expression built up out of unanalysable primitives. This simple-minded idea could be implemented by the following constraint upon formalizations:

> **7)** A formalization is adequate only if each of its logical constants is matched by a single English expression making the same contribution to truth conditions.

On this view, formalization effects only notational changes, so far as the logical constants are concerned.

(**7**) corresponds to a deeply unambitious conception of logical form. The logical form of a sentence is given by how its logical constants occur, and the pattern of occurrence of the non-logical expressions. An artificial language would be just a convenient way of schematizing the non-logical expressions, rather as we did in chapter **1**, and providing a usefully standardized notation for the logical constants. The idea of an artificial language would not essentially enter into the account of the nature of logical form.

Although (**7**) would rule that formalizations like (**3**) of (**1**) are inadequate, since they introduce constants (in this case "&") having no corresponding expression in the English, it would play havoc with traditional practice in formalization. English universal quantifications could not be formalized by universally quantified conditionals since that would involve importing a constant ("→") to which there corresponded no English expression; similarly for English existential quantifications. Adjectival and adverbial modification could not be formalized by the conjunctive method. Necessitated atoms could not be formalized in **QN**. This narrow and unambitious view appears to be required, if one is to hold to the thesis (**1.16**):

> If an argument in natural language is adequately formalizable as valid, then it is *formally* valid.

If such a thesis is not to be trivial, it requires a conception of formal validity that applies directly to English without any detour through

formalizations. The only available conception is that already offered in chapter **1.10**: an argument is formally valid iff its validity turns only upon the meanings of the logical constants it contains and upon the pattern of occurrence of the non-logical expressions. On this definition,

> **8)** This is a red house, so this is a house

is not formally valid. Hence, by (**1.16**), it should not be adequately formalizable as valid, and the only way to secure this result would appear to be the adoption of (**7**).

(**7**) is one way in which one could give a criterion for the difference between logical form and analysis, and thus limit the Tractarian dream. The motivation is coherent, but is far from doing justice to ambitious theses about logical form. Is there an alternative criterion?

We might prefer to think along the following lines. Logic, and thus formalization, must have no truck with non-logical expressions. These must be recognized as contributing only through the pattern of their occurrence, and not through their specific meaning. Hence a formalization should not, as (**3**) does, introduce a predicate-letter (in this case "*F*") which, according to the correspondence scheme, corresponds to an expression ("is a man") which does not occur in the sentence to be formalized. Let us call a generalization of this condition the "*correspondence requirement*". It rules that if the correspondence scheme associated with a formalization has it that, say, "*F*" corresponds to "...", then that actual expression, "...", must occur in the sentence of natural language which is formalized. This requirement is clearly not met by (**3**) relative to (**1**): the correspondence scheme mentions "is a man", which does not occur in (**1**). Yet the requirement would seem more liberal than (**7**), since it allows modes of composition in English to be represented as the application of **Q**-constants.

The correspondence requirement must be understood with some generosity if all its liberality is to be reaped. For example, the spirit of the requirement should allow the conjunctive formalization of adjectival modification. Yet in a standard **Q**-formalization of (**8**) as

> **9)** $F\alpha$ & $G\alpha$; $G\alpha$

we find ourselves saying that "*F*" corresponds to "is red", whereas that expression does not literally occur in the premise of (**8**). While the difference is important for a detailed understanding of the workings of natural language, let us agree to interpret the correspondence requirement in such a way that such differences will not count.

With this rather vague liberalization, the correspondence requirement comes closer than (**7**) to the traditional view, in that it allows the standard formalizations of sentences containing, for example, adjectival modification and quantifiers. It rules out a few of the Q-formalizations proposed in this book, but there is in addition at least one major category of formalizations which it brings under suspicion: Davidson's treatment of verbs of action and some of their adverbs. Consider, for example, the formalization of

10) Shem kicked Shaum

as

11) $\exists x(Fx \mathbin{\&} G\alpha x \mathbin{\&} Hx\beta)$

with "F" corresponding to "is a kick", "Gxy" to "x kicked y", "Hxy" to "x was received by y", "α" to "Shem" and "β" to "Shaum". Perhaps our vague liberalization of the correspondence requirement allows that (**10**) contains something close enough to "is a kick" (viz. "kicked") and also contains something close enough to "x kicked y" where the position marked by y is supposed to be occupied by an expression standing not for a person or a rock but for a kick. But the liberalization cannot allow that (**10**) contains an expression close enough to "was received by". Either the tradition or else the correspondence requirement is at fault.

The correspondence requirement is clearly violated by **QW** and **QC**.[4] On our current understanding of what is to count as a logical constant, furnished by the list in chapter **1.11**, the predicates "W" and "C" of these languages are not logical constants. Hence they are non-logical expressions, yet the English expressions to which they correspond ("is a world", "is a counterpart of . . . in . . .") do not occur in the natural language sentences.

[4] Even one who accepted the correspondence requirement need not be worried by this fact. Those who have done so much fruitful work on languages like **QW** and **QC** would have no cause for alarm if some of the results were to get classified as contributing to analysis rather than logical form.

Ex 6.3 Do **QN**-formalizations meet the correspondence requirement?

With the ideas currently to hand, I can find no way of curbing the Tractarian vision by providing a criterion for the distinction between logical form and analysis, except by stipulations which classify as incorrect some traditional logical form proposals. I believe that this shows that the ideas to hand are insufficient. To justify the kinds of logical form proposal one finds in the founders of the logical form tradition, Russell and Frege, one needs ideas, notably that of compositional semantics, that only came fully into view in the second part of the twentieth century.

3 Proof

A comprehensive grasp of the activities of the logician requires understanding the notion of proof. In addition to its intrinsic importance, and its connection with certain traditional aspirations in logic, it has a role to play in giving an account of what it is to be a logical constant.

I mentioned in chapter **1** that a traditional logical aspiration has been the mechanization of inference. Where there is disagreement about what follows from what, it would be good to feed premises and conclusion into a calculating engine, and wait for the computation of a verdict. Such an engine has to be fed sentences. It cannot be fed propositions, for these are too abstract. It is important that one express the argument in a language in which the validity-relevant features correspond conveniently to the physical make-up of the sentences, for it is to this physical make-up that, in the first instance, we can expect the machine to be sensitive. The mechanical aspiration thus provides a motivation for creating perspicuous artificial languages.

We saw in chapter **4.14** that this aspiration cannot be satisfied in full with respect to a language as rich as **Q**, for there is demonstrably no decision procedure for **Q**. This fact has added to the importance of the notion of proof. For it is demonstrably the case for **Q** that there are systems of proof in which, if an argument is valid, a proof will be found for it in a finite number of steps. This does not add up to a decision procedure because if following the system has not resulted in a proof of a certain conclusion after a million or ten million steps, you do not know whether this is because a few or a million more steps are needed or because the argument under test is invalid.

There are various different kinds of rules of proof, but I shall give a sketch of a system of natural deduction. A reader wanting to become proficient in using natural deduction must look elsewhere (for example to Lemmon [1965]), and one already proficient could skip to §4.

Confining our attention just to propositional logic, a standard system of natural deduction associates with each sentence connective two kinds of rules of proof: *introduction rules*, which specify from what premises one can derive a **P**-sentence dominated by the connective, and *elimination rules*, which specify what conclusions one can derive from premises containing a **P**-sentence dominated by the connective. The intuitive idea behind the introduction rule for "&" is that from a pair of sentences as premises you can infer their conjunction. For the elimination rule, the idea is that from a conjunction you can infer either conjunct. In setting out a full system, it is easiest to express these ideas in a slightly more complicated way. Using "Γ" and "Δ" to stand for (possibly empty) sets of sentences, and a horizontal line to express that from what is above one can derive what is below, we write the rules as follows:

$$\&\text{E:}\quad \text{if}\quad \frac{\Gamma}{X \,\&\, Y}\quad \text{then}\quad \frac{\Gamma}{X}\quad \text{and}\quad \frac{\Gamma}{Y}$$

and

$$\&\text{I:}\quad \text{if}\quad \frac{\Gamma}{X}\quad \text{and}\quad \frac{\Delta}{Y}\quad \text{then}\quad \frac{\Gamma \cup \Delta}{X \,\&\, Y}.$$

(Here "∪" represents set-theoretic union: $\Gamma \cup \Delta$ is the set of all the sentences belonging to Γ or to Δ.) The elimination rule, &E, tells you that if X & Y is derivable from Γ, then each conjunct is also derivable from Γ. The introduction rule, &I, tells you that if X is derivable from Γ and Y from Δ, then the conjunction can be derived from the premises obtained by adding Γ to Δ. "Derivable" in these contexts is not supposed to have any independent meaning. Rather, its meaning is to be fixed by the specification of the rules of proof for all the connectives.

For "¬" the elimination rule is straightforward:

$$\neg\text{E:}\quad \text{if}\quad \frac{\Gamma}{\neg\neg X}\quad \text{then}\quad \frac{\Gamma}{X}.$$

The introduction rule is based on *reductio ad absurdum* style argument, for example:

$$\neg\text{I:}\quad \text{if}\quad \frac{\Gamma, X}{Y \,\&\, \neg Y}\quad \text{then}\quad \frac{\Gamma}{\neg X}.$$

This rule tells you that if you can derive a contradiction from a sentence X together with zero or more other premises Γ, then you can derive $\neg X$ from Γ.

In addition to the rules for each connective, the system also requires general structural rules. All the rules mentioned so far are hypothetical. They tell you that if such-and-such is derivable, then so is something else. But unless there is at least one categorical rule, there will be no categorical facts of derivability. The categorical rule is

$$\text{A:}\quad \frac{X}{X}$$

sometimes called, perhaps misleadingly, the rule of assumptions. The only way in which derivations can get going is by one or more applications of A.

To illustrate an application of these rules, let us adopt a standard convention. We will write sentences on numbered lines, the numbers enclosed in parentheses. To the left of the line number, we will indicate, by numbers, not enclosed in parentheses, the line number of "assumptions": sentences introduced into the proof by A and used as premises at some point, possibly remote, in the derivation of the sentence on the line. To the right we cite the rule to which appeal is made in writing the line:

1	(1)	$p \,\&\, \neg p$	A
	(2)	$\neg(p \,\&\, \neg p)$	¬I.

In the application of ¬I on line (2), Γ has no members, and the instance of Y & $\neg Y$ is p & $\neg$ p. No number occurs to the left of "(2)", since any number appearing there should refer to a member of Γ.

The system is completed by associating each connective with suitable introduction and elimination rules.

Ex. 6.4 State suitable introduction and elimination rules for ∨ and ↔.

The motivation for the rules is to reflect valid patterns of reasoning, but in the statement of the rules there is no mention of validity, truth or any other "semantic" notion. Whether or not something is a correct application of a rule can be determined just by inspecting the shapes of the sentences. Mechanical testing can determine whether or not a rule has been correctly applied. The rules fix the extension of a relation of derivability, which can be written "$\vdash_{\mathbf{P}}$". Thus our little proof above established that

1) $\vdash_{\mathbf{P}} \neg(p \mathbin{\&} \neg p)$,

in other words, that $\neg(p \mathbin{\&} \neg p)$ is derivable by the rules relating to **P** on the basis of no assumptions.

How is "$\models_{\mathbf{P}}$" related to "$\vdash_{\mathbf{P}}$"? We shall say that for a language, L, any system of rules of proof, π, for L, and any semantics, σ, for L, in terms of which validity is defined, π is *complete* with respect to σ iff

2) if $\Gamma \models_{\sigma} X$ then $\Gamma \vdash_{\pi} X$;

and that π is *sound* with respect to σ iff

3) if $\Gamma \vdash_{\pi} X$ then $\Gamma \models_{\sigma} X$.

There are standard systems of natural deduction which are both sound and complete with respect to the semantics we gave for **P**, so, using "$\vdash_{\mathbf{P}}$" to relate to such a system,

4) $\Gamma \vdash_{\mathbf{P}} X$ iff $\Gamma \models_{\mathbf{P}} X$.

In short, "$\vdash_{\mathbf{P}}$" and "$\models_{\mathbf{P}}$" are equivalent, and there are rules for **Q** relative to which "$\vdash_{\mathbf{Q}}$" and "$\models_{\mathbf{Q}}$" are also equivalent. For this result to have significance, it is obviously important that the two relations be defined in different ways. In the standard terminology, "$\vdash_{\mathbf{P}}$" is defined purely *syntactically* whereas "$\models_{\mathbf{P}}$" is defined *semantically*.[5]

A hypothesis is that there is no way of devising rules of proof for English that would come anywhere near being both sound and complete relative to our intuitive semantics. So if you want to be able to mechanize inference along the lines of rules of proof, it helps to transform English into some language for which there are rules of proof: **Q** would be an example of such a language. This hypothesis is, in effect, **(1.17)**:

> The formalization of a natural sentence renders the proposition it expresses accessible to formal deductive manipulations.

This claim can be understood in a way upon which it is true. However, it does nothing to establish an ambitious thesis according to which the semantic mechanisms of a formalization are the very ones at work in the natural sentence. By using a quantified conditional to express the proposition expressed in English by a quantification, you may be able to use rules of proof to prove some conclusion; but this does not begin to show that the English sentence is really itself a quantified conditional. It shows at most that it has the same truth (upon-an-intended-interpretation) conditions as such a conditional, but these truth conditions may be expressed by quite different mechanisms.

4 Formal and structural validity

A traditional view is that the logical form of a sentence shows how its primitive expressions are organized so as to engender its overall

[5] The contrast is not as clear as it might appear. Using the word "truth" in a definition does not automatically render it semantic, and the "semantic" features ascribed by interpretation rules are identified on the basis of purely syntactic features, the shapes of the sentences. The algebraic view of semantics mentioned in chapter **5.11** detaches semantics from the ordinary notions of truth etc., so could be regarded as belonging to syntax, yet it mentions no rules of proof. Equally, if in stating rules of proof you read the horizontal line as showing that what is below *follows from* what is above, you are viewing the rules in a semantic light.

meaning: logical form displays semantic structure. One way in which this thought arises is as follows: formal validity is validity purely in virtue of structure, abstracted from content. However, Gareth Evans has argued, I think entirely convincingly, that it is possible to distinguish a coherent and important notion of validity in virtue of semantic structure (structural validity) which contrasts sharply with the traditional notion of formal validity. The main point of contrast is that formal validity turns on the specific contribution – the "content" – of the logical constants, whereas the conception Evans articulates is one in which validity depends upon no such specific contribution: the validity is *purely* structural.

Evans's starting point is the notion of a categorial semantics: a semantic theory having a number of semantic categories, each primitive expression being assigned to just one category. Members of the same category are assigned the same kind of entity by the semantics. Thus an n-place predicate might be assigned a set of n-tuples; an n-place sentence connective an n-ary truth function (i.e. a function from n-tuples of truth values to a truth value); and so on. Roughly speaking, a structurally valid inference is one whose validity is independent of the particular assignments that are made within the semantic categories, but which is wholly dependent on the pattern of the categories in the argument. For example, it is arguable that

1) John is a large man, so John is man

is structurally valid. Imagine a categorial semantics which treats "John" as belonging to the category of names, marked by the fact that to each member of the category the semantics assigns a member of the domain; which treats "man" as belonging to the category of 1-place predicates, marked by the fact that to each member of the category the semantics assigns a set of members of the domain; and which treats "large" as belonging to the category of extensional adjectives, marked by the fact that to each member of the category the semantics assigns a function from a set (the set associated with the one-place predicate to which the adjective applies) to a subset of that set. Relative to this semantics, (**1**) is structurally valid, since any argument whose premise consists of a name followed by an extensional adjective followed by a one-place predicate, and whose conclusion consists of that name followed by that one-place predicate, is valid. The validity of (1) is thus

due to its semantic structure: due only to the semantic categories to which its elements belong, and not to the special contribution the elements make within their categories.

If there were no restrictions on the categorial semantics, then any inference whatsoever would count as structurally valid (relative to some semantics or other). For example, we might subdivide the category of one-place predicates into three as follows: (a) those to each of which the set of all and only bachelors is assigned; (b) those to each of which the set of all and only unmarried persons is assigned; (c) those to which any other set of objects is assigned. In the light of this categorization

2) Tom is a bachelor, so Tom is unmarried

is structurally valid, since any argument whose premise consists of a name followed by an (a)-type one-place predicate and whose conclusion consists of that name followed by a (b)-type one-place predicate is valid.

Evans places the followed constraint upon categorial semantics:

> We will construct a new category out of an older and more comprehensive category only when we can make an assignment to members of the new category which provides a *different* explanation for the behaviour which members of the new category had in common with the old, the provision of which explanation would show that the apparent unity in the behaviour of members of the old category was deceptive, concealing deep differences in functioning. ([1976], p. 64)

A semantics which makes the envisaged threefold division among one-place predicates will not meet this constraint. By contrast, here is a putative case in which a division would be justified. Imagine a language containing both names and definite descriptions of the form "the so-and-so", and a semantic theory which lumps them into the single category of "singular terms". The theory assigns members of the domain to some (perhaps all) names and some (perhaps all) definite descriptions. Some inferential behaviour is common to singular terms. For example, there is a valid pattern of inference from a premise applying to a singular term a predicate, F, to a conclusion which says that there is at least one *F*. Now imagine a new semantic theory which proposes a division of the category of terms, separating names from descriptions, treating names as before, and associating with "the" a function from pairs of sets to truth values: the function takes the value *true* iff the first set in the pair has a unique member and that member belongs to the second. The new semantic theory gives a new explanation of the inferential behaviour common to singular terms. In the case mentioned, there are two explanations, one for the case in which the premise contains a name, patterned on the explanation provided by the old theory, and another for the case in which the premise contains a definite description. In the second case, the explanation alludes to two assignments, that to "the" and that to "*F*", whereas the old theory alluded to a single assignment to "the *F*". This surely satisfies Evans's condition that the new theory's division of the category of singular terms shows "that the apparent unity in the behaviour of members of the old category was deceptive, concealing deep differences in functioning".

In our earlier listing of the semantic categories of **Q**, the 2-place connectives were bundled into a single category, and this categorization is justified by Evans's test. The only inferential behaviour common to all of them is the substitution of equivalents, which can be represented in the following scheme:

3) For any 2-place connective, ¢, any interpretation, *i*, if $i(Y) = i(Z)$ then $i(X ¢ Y) = i(X ¢ Z)$ and $i(Y ¢ X) = i(Z ¢ X)$.

Specification of the different truth functions expressed by the three 2-place sentence connectives of **Q** gives no new explanation of this common behaviour. The explanation still resides simply in the fact that the connectives express some truth function or other. No further subdivision of the semantic category of 2-place connectives is justified. Hence an argument of the form

4) *p* & *q*; *p*,

while it may count as *formally* valid, does not count as *structurally* valid. The relevant structure is:

5) [sentence$_1$] [sentence connective] [sentence$_2$] ; [sentence$_1$]

But not all instances of this structure, permuting just within the specified categories, are valid; for example:

6) $p \vee q$; *p*.

The notion of a structurally valid argument corresponds to the idea that there are some arguments which are valid, not in virtue of the specific meanings of their expressions, but rather in virtue of the way in which the sentences are constructed. Though formal validity is sometimes thought to answer to this intuitive idea, it does not, for formal validity is validity partly in virtue of the specific meanings of certain favoured expressions, the logical constants, and if this idea is to have importance there has to be a deep reason for selecting some expressions rather than others as the logical constants.

On Evans's view, a sentence's semantic structure will be represented by a sequence of categories. A **Q**-formalization will not constitute such a representation. If you looked to **Q**-logical form to give answers about semantic structure you would, arguably, get wrong answers. You would infer that English quantifiers are unary, whereas the evidence is that they are binary. You would infer that English universal generalizations contain an expression belonging to the category of two-place sentence connectives, whereas they do not. You would infer that adjectives belong to the same category as predicates, being associated with a set of sequences, whereas the category they really belong to is of expressions assigned functions from sets to subsets.

Instances of adjectival detachment (for example (**2.8**): This is a red house, so this is a house) and instances of explicit conjunction elimination (on the pattern of (**4**)) both suggest *patterns* of validity. Evans's insight is that the pattern of adjectival detachment abstracts from the meaning of any specific expression, whereas the pattern of conjunction elimination essentially involves a sign for conjunction.

If the draconian standard of (**2.7**) is adhered to, whereby each **Q**-constant must be matched by an English expression, there is a good sense in which a formalization reveals the logical structure of the natural sentence. It shows how the English sentence is structured around the logical constants it contains: what those constants are, and how the non-logical expressions contribute to truth conditions. The same is not true of the more relaxed standard provided by the correspondence requirement. (For example, the requirement permits the formalization of adjectival modification by conjunction, whereas there may be no expression for conjunction in such an English idiom.) None the less one can see how people might be led into thinking that formalizations meeting this requirement speak to the question of logical structure. Given, finally, that logical structure might well be confused with semantic structure one can see how the view that logical form reveals semantic structure found space within the traditional conception of logical form. Crudely put, the progression goes as follows: formalization singles out the logical constants of English and the way in which the non-logical material is organized by the constants; thus formalization identifies logical structure, the mechanisms whereby expressions contribute to truth conditions; thus logical form identifies semantic structure. The starting point is not secured by traditional practice in formalization, and the steps are unjustified.

5 Logical constancy

One goal we attributed to the logician is the characterization not merely of validity, but also of formal validity. A formally valid argument is one valid just in virtue of the meanings of the logical constants it contains and the pattern of occurrence of the other expressions. An account of the goal of logic thus essentially involves an account of what it is to be a logical constant. Once we have that, we will be able to specify the border between logic and other disciplines. A language fit for logic will have no constants other than logical constants. A sentence will be logically true iff it is true in virtue of the meanings of the logical constants it contains together with the pattern of occurrence of non-logical expressions. Logical necessity is necessity owed to the meanings of the logical constants, and to the meanings of no other expressions.

Within a formal language like **Q**, it is simply stipulated which are the logical constants: they are the expressions which receive a constant interpretation. However, there is no limit to what expressions could be accorded a constant interpretation. For example, one could introduce a **Q**-like language with the additional stipulation that some symbol was, upon every interpretation, to be assigned the set of cats. The present section asks what principles underlie our sense that the symbol thus treated in that language as a logical constant is not really one.

In chapter **1** we gave a list of the logical constants, mentioning that this does not even hint at a rationale for grouping the listed expressions together. I quite cavalierly ignored the original list in chapter **5**, in order to investigate the logic of modality.

We will consider four accounts of what makes an expression a logical constant. We will find that there is a large measure of convergence between them. The first, briefly mentioned in chapter **1.11**, is that the logical constants are *topic-neutral*: they can be distinguished by the fact that they introduce no special subject matter. Thus "if" and "some" are intuitively not "about" anything at all, whereas a name like "Ronald Reagan" is about Reagan, and "happy" is about happiness.

The view is encouraged by the thought that logic concerns reasoning in general. It should therefore give special attention just to those expressions which can be used in reasoning on any subject matter whatsoever, and this may suggest that the logical constants should not introduce any specific subject matter.

If standard views about what expressions are logical constants are correct, the topic-neutral account as so far formulated does not give a sufficient condition for an expression to be a logical constant. Take a world like "therefore", which can certainly be used in connection with reasoning on any subject whatsoever, and is as topic-neutral as any expression one can think of. Yet orthodoxy does not include it among the logical constants. "Therefore" is normally used not as part of an argument, but rather to indicate what the conclusion of the argument is. Since it is not part of the argument, it cannot contribute to its validity. If we think of an argument in an abstract way, as a collection of sentences one of which is the conclusion, its validity is independent of whether anyone has ever "indicated" which sentence is the conclusion. Since logical constants should certainly contribute to the validity of arguments, "therefore" should not count as a logical constant, despite its topic-neutrality.

Ex. 6.5 Discuss the following suggestion:

"⊢" is a logical symbol and represents "therefore", so the latter is, in effect, counted as among the logical constants.

This suggests how the topic-neutral theory should be reformulated. A logical constant is an expression whose meaning contributes to the validity of arguments and which introduces no special subject matter. With this amendment, the main problem is the vagueness of the notion of topic-neutrality, which leaves unsettled, concerning some expressions, whether they are logical constants or not. One example is "but", not normally accounted a logical constant. My earlier suggestion was that it makes the same contribution to validity as "and", since it expresses the same truth function, while "*A* but *B*" suggests that in the context of *A* it is surprising, poignant, or worthy of special note or emphasis, that *B* is true. Does this additional component in the meaning of "but" constitute the introduction of a specific subject matter (surprise, poignancy or special notice)? The topic-neutral theory delivers no answer, whereas traditionally, and intuitively, "but" does not count as a logical constant.

The vagueness of the topic-neutral theory is conspicuous in the case of a traditional logic constant, "=". On the one hand, this perhaps ought not to count as topic-neutral, for it might seem that identity is a specific topic, not involved in all reasoning. On the other hand, perhaps it ought to count as topic-neutral, for no specific objects are introduced when we introduce "=". The second idea seems to point in a helpful direction, but we cannot rest content with it as it stands, for it could as well be said that identity introduces *every* object, since every object is related by it to something.

A possible approach is to refine the notion of "introducing objects" in terms of understanding conditions. Perhaps there are no objects, and no kind of object, knowledge of which is required to understand "=". Arguably, there are objects, or a kind of object, knowledge of which is required to understand a non-logical constant like "cat".

This criterion would arguably count "necessarily" as a logical constant, for even if a philosophical account of the notion introduces worlds, it would appear that knowing about these is not essential to a grasp of the adverb. By contrast, the criterion would exclude the constants "W", "w*" and "C" of **QW** and **QC** respectively, since there are objects, worlds and counterparts, which one must know about to understand these expressions.

The criterion would exclude the set-theoretic expression "∈" (for "is a member of"), since in order to understand this expression one has to know about sets. However, it would remain a matter for dispute whether or not the criterion counted second order logic as logic; whether, that is, it counted second order quantifiers as logical constants. On the one hand, arguably it does not, since these quantifiers have a

special subject matter, properties or sets, and one needs to know about these to understand the quantifiers. On the other hand, the criterion must obviously be so understood as to count first order quantifiers as topic-neutral, which means that the fact that their range is confined *to objects in the domain* must not count against their neutrality. However this is achieved, it might have the result that the fact that the range of a second order quantifier is (for example) confined to *subsets from the domain* should also not be a mark of non-neutrality.

Although topic-neutrality corresponds to an intuitive strand in traditional thinking about the logical constants, it is hard to make the idea precise. We will therefore turn to other criteria.

Ex. 6.6 Consider a sentence operator: "It was the case that . . .". Does this count as a logical constant by the topic-neutral criterion? Are there any other reasons for judging the issue either way?

The second criterion to be considered, the *rules criterion*, focuses in a different way upon the role of the logical constants in reasoning. The claim is that a logical constant is an expression whose meaning can be wholly specified by rules of proof, rules stating what you can validly infer from a sentence dominated by it, and from what you can validly infer a sentence dominated by it.

Various constraints can be appropriately imposed on the rules exploited by the rules criterion.[6] Those for a target expression must not mention any other expressions in the language. Without this restriction, the door would be open to counting a large number of expressions as logical constants which intuitively are not. For example, I might otherwise claim that "ruish" has a meaning wholly fixed by the following introduction rule and pair of elimination rules:

$$\text{RUISH I:}\quad \text{if } \frac{\Gamma}{x \text{ is Jewish and } x \text{ is Russian}} \text{ then } \frac{\Gamma}{x \text{ is ruish}}.$$

$$\text{RUISH E (i)}\quad \text{if } \frac{\Gamma}{x \text{ is ruish}} \text{ then } \frac{\Gamma}{x \text{ is Russian}}$$

$$\text{RUISH E (ii)}\quad \text{if } \frac{\Gamma}{x \text{ is ruish}} \text{ then } \frac{\Gamma}{x \text{ is Jewish}}.$$

6 For requirements not mentioned here, notably that of "harmony" between elimination and introduction rules, see Dummett [1991], esp. pp. 215 ff. See also Harman [1986], esp. pp. 117 ff.

The justification for excluding such rules is that the target expression is supposed to sustain the validity of the reasoning all by itself. If other expressions were mentioned in the rules, there would be a suspicion that these were responsible for some of the validity-relevant features.

A further appropriate restriction is that introduction rules should not contain any occurrences of the target expression in the conditions for the introduction. This would exclude the introduction rule for "¬" given earlier:

$$\text{if} \quad \frac{\Gamma, X}{Y \,\&\, \neg Y} \quad \text{then} \quad \frac{\Gamma}{\neg X}.$$

The justification is that it should not be assumed that there is some way justifiably to get the expression into the language other than through the introduction rule; for if there were, that other way could be relevant to the meaning of the target expression. This means that in order to count "¬" as a logical constant by the rules criterion, one needs to specify some arbitrary falsehood, say "2 + 2 = 5", perhaps abbreviated "⊥", and give the introduction rule as:

$$\text{if} \quad \frac{\Gamma, X}{\bot} \quad \text{then} \quad \frac{\Gamma}{\neg X}.$$

On plausible assumptions, the rules criterion determines that the English truth functional sentence connectives are logical constants. The assumptions are that rules for "and" just like those given for "&" enable the derivation only of valid arguments, and that the meaning of "and" is fixed by the fact that a conjunction is true iff both conjuncts are. The validity of the elimination rule ensures that the truth of a conjunction is sufficient for the truth of each conjunct, and the validity of the introduction rule ensures that the truth of each conjunct is sufficient for the truth of their conjunction.

Although exactly the same rules would be valid for the binary connective ". . . and 2 + 2 = 4 and . . .", we need not fear that the latter will wrongly be counted a logical constant, for it is not plausible to say that its *meaning* is fixed by the rule that "*A* and 2 + 2 = 4 and *B* is true iff *A* is true and *B* is true": a component of the meaning is left undetermined.

This sort of example poses a different problem for rules theory. Consider the connective "¢", where "*A* ¢ *B*" expresses the denial of the disjunction of "*not-A*" and "*not-B*" (cf. Peacocke [1987]). The rules of "and"-introduction and elimination are valid for ¢, yet it is plausible to hold that although "¢" expresses the same truth function as "and", the two expressions differ in meaning. Someone could sincerely accept "*A and B*" and sincerely deny "*A* ¢ *B*"; this is some evidence that they believe that *A and B* and do not believe that *A* ¢ *B*, which in turn is some evidence that the sentences differ in meaning. If so, one would have to confess that the rules criterion does not rule "and" a logical constant, since its meaning is not fixed by the validity of the rules, since the validity of the rules does not distinguish between the meaning of "and" and the meaning of "¢".

If we replace "meaning" by "truth conditions" in the rules criterion, we would be returned to the difficulty that ". . . and 2 + 2 = 4 and . . ." would counterintuitively be accounted a logical constant.

Intuitively the meaning of "¢" is more detailed or specific than the meaning of "and". This suggests amending the rules criterion as follows: an expression is a logical constant iff its meaning is the least specific meaning which validates the introduction and elimination rules. The problem now is that "¢" will not be a logical constant, either relative to the rules for "and" or to any others, since to assign it the less specific meaning of "and" will validate any rules which it would validate, yet will not assign it the correct meaning. This problem can be overcome by distinguishing between primitive and defined constants. The category of primitive logical constants is as determined by the amended rules criterion. The category of logical constants is the closure of this category under definition.

Ex. 6.7 (a) How does the rules criterion treat "but"?
(b) Could there be a logical constant that was primitive in the sense of being typically learned directly, and not via a definition, and yet which had the meaning of "¢"?

We have tried applying the rules criterion to English expressions. Now let us check whether it delivers the right results for our formal languages. The **P**-validity of the rules for "&" entails that every interpretation upon which a conjunction is true is one upon which each conjunct is true; and every interpretation upon which each conjunct is true is one upon which their conjunction is true. To recover the actual interpretation rules for "&", we need to add the assumption that every sentence not true upon an interpretation is false upon that interpretation.

Similar considerations, upon similar assumptions, substantiate the claims of the other truth functional sentence connectives, in English and in **P**, to be logical constants.

Quantifier elimination and introduction rules for **Q** would take the form:

$$\forall \mathrm{E}: \quad \frac{\forall \nu X}{X^{n}_{\nu}},$$

where X^{n}_{ν} results from X by replacing one or more occurrences of ν by n.

$$\forall \mathrm{I}: \quad \text{if } \frac{\Gamma}{X^{n}_{\nu}} \quad \text{then} \quad \frac{\Gamma}{\forall \nu X} \quad \text{provided that } n \text{ does not occur in } \Gamma.$$

The validity of ∀E ensures that any interpretation upon which "$\forall \nu X$" is true is one upon which "X^{n}_{ν}" is true, regardless of what the interpretation assigns to "n". Since each interpretation assigns just one thing to each name-letter, this enables us to infer the stronger: any interpretation upon which "$\forall \nu X$" is true one such that "X^{n}_{ν}" is true upon it and upon any n-variant interpretation. The correctness of ∀I ensures that if "X^{n}_{ν}" is true on some bunch of interpretations which include every possible assignment to "n" then "$\forall \nu X$" is true upon any member of the bunch. Thus the interpretation rules for ∀ are recoverable from the validity of the rules of proof.

It would not be easy to apply similar reasoning to English. For one thing, although one can see roughly what sort of rules should correspond to ∀E and ∀I, it is unclear how they should be couched in detail. It would also be unclear whether the least specific meaning verifying these rules would be the meaning of "all" or "every". For example, an important type of reasoning involving "all"-sentences, inductive reasoning from instances of a generalization to the generalization itself, might be argued to be partially constitutive of the meaning of "all", and it is not obvious whether this part would be captured by rules of "all" elimination and introduction. It might be better

to argue indirectly: if "all" matches ∀ in meaning, and the latter is a logical constant, then so is the former.

Ex. 6.8 Propose suitable introduction and elimination rules for "=". (See Tennant [1978], p. 77.)

What does the rules criterion exclude from the category of the constants? Name-letters and predicate-letters are not even candidates for constancy, since they have no fixed interpretation. Likewise, names in English could not be constants, since there appear to be no valid rules special to any name.

The case of predicates is more delicate, since the classical view is that the English predicates "is the same as" and "exists" should count as logical constants, in view of their formalizability by the **Q**-constants "=" and "∃". On anti-Meinongian assumptions, the rule

$$\text{EXISTS I:} \quad \text{if } \frac{\Gamma}{\ldots n \ldots} \quad \text{then} \quad \frac{\Gamma}{n \text{ exists}}$$

is valid for extensional parts of English, and it would be plausible to argue that the least specific meaning assignable to "exists" which would validate it is its actual meaning. The significance of this is diminished by the fact that EXISTS I is not in general valid in English. There is a valid rule for **Q** corresponding to EXISTS I, so that one would expect an existential predicate to be a (non-primitive) logical constant of **Q**, as indeed it is: "$\exists x(x = \ldots)$". Rather similar remarks apply with respect to identity. The non-extensionality of English means that there is no generally valid elimination rule, whereas the extensionality of **Q** means that there is such a rule.

Ex. 6.9 Give a counterexample to EXISTS I as a rule for English.

This means that the rules criterion turns out to have an implicit relativity to a language. We must say either that "is the same as" and "=" do not mean the same, or else that of two expressions with the same meaning, it may be that one is a logical constant (in its language) while the other is not (in *its* language).

Though the topic-neutral criterion counts "□" among the logical constants, this ruling has been disputed, so it would be useful if it could be confirmed by an alternative criterion. The elimination rule appropriate to **PN** is obvious:

$$\Box\text{E: if } \frac{\Gamma}{\Box X} \text{ then } \frac{\Gamma}{X}.$$

Validating this rule (by the standards of **PN**-validity) requires that for all worlds w and interpretations i, if $\Box X$ is true at w upon i, then X is true at w upon i. By itself, this does not entail any significant feature of the interpretation rule for $\Box$. For example, it is consistent with $\Box$ having a much less specific meaning than it actually has, for example being a "null" operator, with the interpretation rule:

$\Box X$ is true at w upon i iff X is true at w upon i.

If the rules criterion is going to count "$\Box$" as a logical constant, everything must turn upon the introduction rules.

One possible such rule is:

$\Box$I(i) if $\frac{\Gamma}{X}$, and every member of Γ has "$\Box$" as its main connective, then $\frac{\Gamma}{\Box X}$.

If this connection holds for **PN**-validity, then if for all interpretations i and all worlds w such that all the Γ are true at w upon i, X is true at w upon i, then for all interpretations i and all worlds w such that all the Γ are true at w upon i, $\Box X$ is true at w upon i. The standard interpretation rule will verify this conditional (given the assumption about the composition of Γ) but it is hard to see whether this is the least specific meaning that will do so. One cannot derive "$\Diamond X \rightarrow \Box \Diamond X$" from $\Box$I(i), so this rule does not determine the meaning "$\Box$" possesses in **PN**.[7]

The following rather similar rule, but with a less demanding restriction upon Γ,

[7] Cf. Prawitz [1965]. $\Box$I(i) determines the logic known as S4. (In this system of classification, deriving from Lewis and Langford [1932], **PN** is known as S5.) Arguably, the English word "necessarily" is ambiguous, and one can see some modal logics, including S4, as specifying various disambiguations of it.

$\Box$I(ii) if $\frac{\Gamma}{X}$, and every sentence-letter in every member of Γ is within the scope of some occurrence of "$\Box$", then $\frac{\Gamma}{\Box X}$,

permits the derivation of just the arguments that are **PN**-valid. This is some evidence that it, together with the elimination rule, fixes the interpretation rules for "$\Box$"; but it is not decisive, since these interpretation rules may attribute more meaning than is needed for the validities.

Though not conclusive, the rules criterion adds some support to the case for the constanthood of "$\Box$". However, it would appear to give no encouragement to the constanthood of "W", "at", "w*" or "C". So far as I know, no one has attempted to formulate elimination and introduction rules for these expressions, and I have no idea how such an attempt would begin. This is an odd result, given that what is expressible in **QN** is expressible in **QW** and **QC**. The criterion has some other counterintuitive features.

First, as it stands, the theory does not provide a sufficient condition for constancy. Let us stipulate that the meaning of the two-place sentence connective, "$\star$", is to be fixed by the two rules:

$$\star\text{I} \quad \frac{X}{X \star Y}$$

$$\star\text{E} \quad \frac{X \star Y}{Y}.$$

The rules together ensure that we can correctly infer any sentence from any sentence (by first using the introduction rule for "$\star$", then the elimination rule), and no expression can have a meaning which legitimates that. Hence the fact that one can state introduction and elimination rules for an expression does not even ensure that the expression has a coherent meaning, let alone that it is a logical constant.

Ex. 6.10 Show why no expression can have a meaning such that $\star$I and $\star$E are both valid, making explicit any assumptions upon which you need to rely. Cf. Prior [1960].

The argument shows that one who would introduce his constants by rules should place further restrictions on the nature of the rules. A commonly suggested restriction, which would suffice to rule out the combination of ★E and ★I, is this (cf. Belnap [1962]). Suppose we have a set, σ, of rules, one pair of which relates to a constant, ¢. Now imagine removing the ¢-rules from σ, and call the diminished result σ′. The proposed restriction is that σ′ should permit all the derivations not involving ¢ that σ permits. That a set containing just ★ does not satisfy this restriction is shown by the fact that the derivation of Y from X (neither containing ★), which is available in the presence of ★, is unavailable in its absence. With this restriction, "★" gives no reason to think that an expression could meet the rules criterion without being, intuitively, a logical constant. The doubt is not whether the account rules in too much, but whether it rules in enough.

It looks as if the meaning of "few" could not be fixed by elimination and introduction rules; and, depending upon the other expressive resources of the language, the same might go for "most". But "few" and "most" would appear, intuitively, to have as much right to count as logical constants as "all" and "some", and this intuition is supported by the topic-neutral criterion. To exclude them is to show that one is willing to count as a logical constant only what yields the sorts of results in which one is interested, and this suggests that there is no objective borderline between the constants and other expressions.

The third criterion for constancy derives from Davidson [1973] and from Dummett [1981]. It is motivated by the thought that the logical constants are the expressions which act as "cement" in the construction of longer sentences out of the bricks of the symbols of the language; I shall call it the "*cement criterion*". Davidson has implemented this idea by saying that a logical constant is an expression which, in a semantic theory for a language containing it, receives a recursive rather than a basis clause. In the kind of semantic framework we have used, a basis clause is the specification of what an interpretation may or must assign to an expression (an object from the domain, a set of n-tuples, or whatever). The recursive clauses are the "interpretation rules". Their effect is to transform the question of the truth-upon-an-interpretation conditions for a sentence into questions about the truth-upon-an-interpretation conditions for its parts. A part may contain a further occurrence of the expression treated by the rule of interpretation, so that that rule, and also perhaps others, may have to be reapplied before the question of the truth-upon-an-interpretation conditions is resolved. The final resolution is purely in terms of the basis clauses. Looking at it from the bottom up, rather than the top down, once the assignments of the basis clauses are fixed, so is the truth upon the interpretation of every sentence in the language.

The criterion excludes "=", since this receives a basis clause. This is not a decisive refutation of the cement theory, since we cannot assume that our original list of logical constants is well motivated. More seriously, the cement theory will include as constants expressions which clearly are not, for example "large" (cf. Evans [1976], p. 69). In terms of our semantic framework, we can find a rule which, rather than assigning something outright to "large", as a basis clause does, makes what is assigned depend systematically upon the assignments effected by the basis clauses. For example, one could introduce "large" into **Q** by the syntactic rule that attaching it to a 1-place predicate (or predicate-letter) forms a new 1-place predicate, and the semantic rule that an interpretation which assigns a given set, σ, to a 1-place predicate, ϕ, must assign the subset of large members of σ to "large ϕ". Just as assignments to the sentence-letters of **P** thereby determine the truth values of all **P**-sentences in virtue of the recursive rules for the sentence connectives, so the assignments to the "large"-free predicates and predicate-letters thereby determine the assignments to all the predicates and predicate-letters in virtue of the recursive rule for "large".

Ex. 6.11 Some logics have a constant to express falsehood, or some arbitrary absurdity: "⊥" is often used. Do you think "⊥" would receive a basis or a recursive clause? What does your answer show about the appropriateness of defining logical constants as just those which receive recursive clauses?

Though the cement theory must be rejected, one can see how holding it makes it hard to find use for a distinction between logical structure and semantic structure. Semantic structure might be thought of as the way a sentence is built up out of its parts. On the cement theory, this structure is supposed to be given by the logical constants.

The fourth criterion is due to Peacocke [1973]. I shall call it the "apriori" criterion. Consider this property of truth functional sentence connectives: if you know what truth values an interpretation has assigned to the components of a sentence dominated by a truth functional sentence connective, and you know the interpretation rule for the connective, then you can work out apriori what truth value the interpretation accords to the resultant. Perhaps some generalization of this property will constitute a distinctive feature of the logical constants.

In the semantic framework we have adopted, an interpretation assigns a truth value to every sentence of the language (relative to a domain, and perhaps also relative to a world – but we will omit these qualifications). It does so in the following way: it makes outright assignments of entities to the sentence-letters, name-letters and predicate-letters. The operators, that is, the sentence connectives and quantifiers, are not themselves assigned anything, but each is associated with a clause specifying a condition upon which a sentence dominated by the operator would be assigned the value true. The outright assignments are the basis clauses, the assignments upon a condition the recursive clauses. We can say that an outright assignment "treats" the expression to which the assignment is made, and that a recursive clause "treats" the expression which dominates the type of sentence which the clause addresses.

The version I shall consider of Peacocke's formulation of the generalization (adapted to our semantics and terminology) is:

1) α is a logical constant iff α is non-complex and, for any expressions $\beta_1, \ldots \beta_n$ on which α operates to form the expression $\alpha(\beta_1, \ldots \beta_n)$, knowledge of what each interpretation assigns to each β_i, together with knowledge of how α is treated, enables one to infer apriori what each interpretation assigns to $\alpha(\beta_1, \ldots \beta_n)$.

This excludes the case in which some β_i or $\alpha(\beta_1, \ldots \beta_n)$ is an operator (and so not assigned anything upon an interpretation). This has no practical importance, and could be avoided without damaging the spirit of the proposal.

(1) clearly rules that truth-functional sentence connectives are constants. Also it excludes, for example, predicate-letters. We can apply the test to "$F\gamma$", where we think of this as corresponding to "$\alpha(\beta)$" in (1) (so that "F" takes the place of "α" and "γ" of "β_1"). Suppose you know what each interpretation assigns to "γ", and so you know in particular that, for some specific interpretation, i, $i(\gamma)$ is Ronald Reagan. Suppose you also know how the unary predicate-letter "F" is treated (each interpretation assigns it a set), and you know in particular that $i(F)$ is the set of honest men. You cannot infer apriori what truth value i assigns to "$F\gamma$", because nothing in what you know about i's assignment to "γ" and treatment of "F" contains any information about Reagan's honesty. So a predicate-letter fails the test for being a logical constant, which is as it should be.

(1) does not count **Q**-quantifiers among the constants. One reason is the trivial one that, given the semantics we have adopted, the truth-upon-an-interpretation condition of a quantification depends not upon how the interpretation treats any component, but rather upon how certain interpretations treat a sentence related in a certain way to the contained open sentence. I shall ignore this problem: let us assume, to assist the discussion, that we are dealing with a semantics in which interpretations assign objects to variables. Then it might seem as if (1) does count the quantifiers as constants. For suppose you know what each interpretation assigns to "Fx" (truth or falsehood, as the case may be), and you know the interpretation rule for "$\forall$", then it may seem that you can come to know apriori what each interpretation assigns "$\forall xFx$". For you know that an interpretation, i, assigns truth to this sentence just on condition that all interpretations agreeing with i on their assignment to "F" assign truth to "Fx".

However, knowledge of what i assigns to "$\forall xFx$" cannot be extracted apriori from the knowledge we have supposed. One might not know that all the interpretations whose assignments to "Fx" one knew about are all the interpretations there are. Moreover, there is an unclarity about what it is to *know what* something, for example a predicate-letter, is assigned. Suppose that i_1 assigns the set of featherless bipeds to "F" and i_2 assigns to this letter the set of men. It is not an apriori matter whether they agree on their assignment to "F". You might know that "Fx" is true upon both i_1 and i_2, but because you do not know that these interpretations agree on their assignment to "F", you might not realize that i_2 is one of the interpretations relevant to the truth upon i_1 of the universal quantification. To deal with this problem, Peacocke ([1973], pp. 226–7) in effect stipulates that included in the knowledge which is to be the basis for the apriori inference is this: for any letter, L, any pair of interpretations, if the interpretations agree on what they assign to L, you know that they do.

As Peacocke stresses, this does not entail that, for any pair of letters, if an interpretation assigns the same thing to each, you know that it does. For this reason, identity does not, as the criterion stands, count as a logical constant. Suppose you know that an interpretation assigns Hesperus to "α" and Phosphorus to "β", and you also know that, like every interpretation, it assigns to "=" the set of ordered pairs whose first member is the same as the second. You cannot work out apriori whether or not "$\alpha = \beta$" is true upon the interpretation. You need in addition the non-apriori information whether or not Hesperus is identical to Phosphorus.

This limitation, like the previous one relating to quantification, could be stipulated away, but then one would begin to wonder whether such stipulations have any rationale other than that of endorsing a predetermined list of constants, a list drawn up on the basis of some criterion other than (**1**). Why not add "is the same height as" to the list, by stipulating that the knower has access to the relevant information? It would seem that the way to argue for the rejection of such a suggestion is not from within the criterion (1), but from the independent intuition that information pertaining to heights must be irrelevant to logic. Again, "□" can be included among the constants, *if* we are prepared to add the following to the available information: with respect to which worlds the interpretations make the sentences of the language true, and, for each world, what each interpretation assigns to each letter with respect to that world. But the question arises whether or not we should be willing to suppose this information to be available.

It would, of course, be pleasing to have a way of giving our intuitions about constancy a precise technical embodiment, and this may well be provided by Peacocke's account, but it seems that it will not serve to underwrite these intuitions. Doubts about whether. "□" or even "=" is a constant would not be resolved by the criterion.

To summarize this part of the discussion, it is not clear whether it is possible to make a marked improvement on the first of the criteria – the topic-neutral one. It does not give definite rulings on certain cases, but perhaps this is simply because the concept of logical constancy is vague. The cement criterion seemed to be on the wrong lines, but the other three criteria showed some encouraging convergence. The indeterminacy in the rules theory related to the question whether the rules-determined meaning was really the least specific that would sustain the validities; and there were problems concerning whether it introduced a language-relative notion of constancy. The apriori criterion did not conflict with the topic-neutral criterion, but, rather, appeared to rely upon it when it came to crucial decisions about the formulation of the criterion.

The boundary to the class of logical constants determines the boundary to the class of logical truths: truths true solely in virtue of the meaning of the logical constants, together with the pattern of occurrence of non-logical expressions. This in turn provides a boundary to the domain of logic: it could be identified with the class of logical truths.

Let us use "C" to refer to the claim that a truth is a logical truth iff it is true just in virtue of the meanings of the logical constants and the pattern of occurrence of other expressions. C has come under attack by Quine [1963]. In Quine's work one finds attacks on at least the following five claims which might be held to be consequences of C:

(**2**) Logical truths are true by convention.

(**3**) Logical truths are true independently of how the world is.

(**4**) "The truths of logic have no content over and above the meanings they confer on the logical vocabulary" (ibid., p. 109).

(**5**) A truth of logic "is a sentence which, given the language, automatically becomes true" (ibid., p. 108).

(**6**) Logical truths are "true by virtue purely of the intended meanings . . . of the logical words" (ibid., p. 110).

The assertion of (**2**) is not part of the intention of the adherent of C, and I will ignore the question of whether it is a consequence of C. (It may seem to be, on the assumption that meanings are conventional.) C is certainly not intended to entail (**3**), and (**4**) is, on the face of it, false. As Quine remarks, the truth of "$\forall x\ x = x$" depends not just on words but on the world, on the fact that everything is self-identical. So we should understand C as saying that to the extent that meaning contributes to the truth of a logical truth it is only the meaning of the logical constants and the pattern of occurrence of the non-logical expressions that are relevant. (**4**) is foreign to the approach of this book, in which no assumptions are made about how the logical vocabulary of English gets its meaning, and according to which, if anything confers meaning on the **Q**-logical vocabulary, it is the interpretation rules of the metalanguage and not the logical truths of the object language. (5) and (6), as recently qualified so as not to exclude a contribution to truth "from the world", are indeed consequences of criterion C.

Quine's argument against these last two theses, the only ones to which C is clearly committed, stems from a quite general scepticism about meaning. Certainly, if there is no such coherent concept as meaning there is no coherent doctrine of truth in virtue of meaning. It would take us too far afield to discuss this general scepticism, so I will simply turn aside from it.

On one interpretation, to say that a sentence is logically necessary is just to say that it is logically true. This requires amplification. If truths like

7) Hesperus is Phosphorus

are *necessary*, then there is a kind of necessity that it is not logical necessity, in the sense just defined, while also not being merely physical, epistemic or moral necessity, or necessity which is a restriction of logical necessity (in the sense of logical truth). (**7**) exemplifies the kind of necessity that Plantinga refers to as "broadly logical" necessity and which Kripke, for example, has called "metaphysical necessity". By contrast, the logical necessity that equates with logical truth is termed by Plantinga "narrowly logical" necessity. Not all broadly logical necessities, like (**7**), are knowable apriori. So if logic is to fulfil its promise of being an apriori subject, it should study and use only a narrower notion. This means that our original definition of validity in chapter **1.3** should be understood as employing a narrower notion.

6 Language, form and structure

The sentences of the artificial languages I have discussed have a strange feature upon which I have not explicitly remarked: for the most part they do not say anything, and cannot be characterized as true or false. The reason is that the letters, sentence-letters, name-letters and predicate-letters, have no fixed interpretation. They are empty vessels waiting to be filled; the fillings are interpretations. Thus a **Q**-sentence like "$\exists xFx$" does not say anything, nor is it true or false. Rather, it awaits an interpretation. Relative to an interpretation which, say, assigns the set of dogs to "F", the sentence says that there are dogs, and is true. In **Q** there is no truth (or saying) *simpliciter*, but only truth (or saying) upon an interpretation.

We could have proceeded differently, introducing **Q**-symbols like "F" and "α" as abbreviations of specified expressions in a natural language, for example "is a dog" and "Fido". Let us call such a language **Q***. It will have the same definitions of interpretation and validity as **Q**, but its non-logical symbols will be thought of as (non-logical) constants for which just one interpretation is correct, and this interpretation, call it $i^\star$, will be specified. For **Q***-sentences, truth equates with truth upon $i^\star$.

This alternative procedure makes **Q*** a real language, one whose sentences all say something, and are all true or false. For the purposes so far discussed, it makes little difference whether one formalizes in **Q** or in **Q***. For this section we need to focus on **Q***. The suggestion I want to discuss is that we can use semantics for **Q*** to give a compositional semantics for natural language. In more detail, the suggestion is that in devising a semantics for English, we proceed in two stages:

1) First, English sentences will be translated into **Q***.
2) Secondly, the semantic resources discussed in connection with **Q** will be applied to the sentences of **Q***; the English expression "true" will be regarded as abbreviating "true upon $i^\star$".

The second step assigns intuitively correct truth (that is, truth-upon-$i^\star$) conditions to the **Q***-sentences, on the basis of the contributions systematically made by their parts. The first step ensures that none but correct truth conditions get assigned to English sentences which have **Q***-translations. Let us call a two-tiered theory of this kind a *proxy semantics*. In the version given above, English semantics are given by proxy of **Q***, which can be called the *proxy language*.

Something like this idea may well have been influential for some time, for example in Russell's early work, but it is only rather recently that a version of it has been explicitly formulated. Donald Davidson has proposed that English should be given proxy semantics. He uses

Q★ as an example of a suitable proxy language, while explicitly denying any commitment to the view that this is the only possible proxy language. He does not favour the type of semantic theory we have introduced, in which symbols are assigned entities of various kinds, preferring instead a "truth theory". I will not attempt to describe what this is, but will note that a truth theory recognizes exactly the same semantic categories as the semantics presented here, and, like our semantics, it provides truth conditions for every sentence on the basis of the elements from which it is constructed, and it is because of the perspicuity of Q★ that it is quite easy to produce a truth theory for it.

The essentially Davidsonian project of proxy semantics specified by (**1**) and (**2**) provides a definition of logical form in the spirit of (**1.15**):

3) The logical form of a sentence is the sentence of the proxy language into which it needs to be translated in order to provide its semantics.

(Cf. Davidson [1977], esp. p. 203.) This gives the project of formalization an importance of a new kind. Formalization would not merely help to characterize validity, or formal validity, for a natural language, but, more widely, would help to give a general characterization of meaning, or at least truth conditions, for natural language. (**3**) would vindicate the view that logical forms provide insight into the semantic mechanisms of natural language.

On Davidson's view, logical form is essentially relative to a proxy language. Relative to a propositional language, the logical form of an English universal quantification will just be "*p*". This tells us very little about the semantics of the English sentence. For example, it would be absurd to infer that the English quantification is "really" unstructured. Some constraints must be placed upon the choice of proxy language.

Intuitively, we want to say that an appropriate proxy language will mirror the structure of the natural language. It would be a pity to make this into a criterion of adequacy for a proxy language, for we wish to allow logical form proposals to provide a conduit for discoveries about the semantic structure of natural languages. For example, Davidson has proposed that a sentence like "Shem kicked Shaum" should be matched with an existential quantification in the proxy language, and it would defeat such a purpose to protest, on the basis of some pretheoretical "intuition" to the effect that the English sentence is atomic and

not a generalization, that this proposal does violence to semantic structure.

The intuitive idea which, according to Davidson, should guide us is that semantics, including proxy semantics, should provide an answer to the question: "What are these familiar words doing here?" If we ask this of "Shem kicked Shaum", the answer given by Davidson's proposed **Q**★-translation would be that "Shem" and "Shaum" serve to refer to objects, that "kicked" introduces kicks, and that the sentence as a whole introduces existential quantification. (I follow Davidson in ignoring tense.)

One more precise test of the adequacy of proxy semantics upon which Davidson insists is that if an English argument is, intuitively, formally valid, its translation in the proxy language should be formally valid. For example, an important component in his argument for his logical form proposal for action sentences is that the **Q**★-translation (or **Q**-formalization) of "Shem kicked Shaum in the face, so Shem kicked Shaum" is **Q**-valid. This places significant limitations upon what could be an adequate proxy language for English. For example, it rules out the adequacy of a propositional language.

Ex. **6.12** Could the same effect be achieved by requiring that the translation procedure envisaged in (**6.1**) be statable in a finite number of rules? Could this requirement rule out the claim that an English quantification has a sentence-letter as its logical form?

A further Davidsonian adequacy condition upon semantic theory is that the theory have only finitely many axioms. This forces a semantic theory to recognize some structure in the object language. Languages typically have infinitely many sentences, so a finitely axiomatized semantic theory can deliver a correct assignment of truth conditions to every sentence only by seeing sentences as made up of parts drawn from a finite total stock, each part making a distinctive contribution to the truth conditions of whatever sentence may contain it. Applied to proxy semantics, this means that there must be a finite manual for translating English into the proxy language, and a finitely axiomatized semantic theory for the latter. Let us call this the *finite axiomatization constraint*.

The idea of proxy semantics enables one to make good sense of some extreme-sounding claims about logical form, for example, the claim that "Shem kicked Shaum" is "really" an existential quantification. The claim amounts to no more than that such a sentence will be translated into an existential quantification for the purposes of proxy semantics. Likewise the claim that "All men are happy" is "really" a quantified conditional amounts to no more than that such a sentence will be translated into a quantified conditional for the purposes of proxy semantics. This is more ambitious than the claim that, say, an English quantification and its $\mathbf{Q}^\star$-rendering are alike in truth conditions, for it makes the further claim that their likeness in this respect is of a special kind, engendered by a single scheme of translation; yet it is suitably less ambitious than the absurd claim that an English quantification contains an invisible occurrence of "if" or "→".

Some puzzlement may remain. Consider two logical form proposals for English quantifications: one made by standard $\mathbf{Q}^\star$-translations, the other by translations into a language containing binary quantifiers. Both proposals might meet all of Davidson's criteria, yet we are intuitively inclined to believe that they cannot both be right, cannot both be true to the semantic structure of English. Intuition clearly sides with treating English quantifiers as binary rather than unary. It seems perfectly possible that a natural language sentence should have a Davidsonian logical form, and yet exploit different mechanisms to achieve the same truth conditions from the same non-logical primitives. Hence it seems possible that a proxy semantics meeting Davidson's constraints should give an incorrect account of semantic structure.

The root of the puzzlement is that whereas Davidson's notion of logical form is relative to a proxy language, we tend to suppose that there is an absolute fact about the semantic mechanisms a sentence exploits.

We saw that the choice of a suitable proxy language is crucial to the value of Davidson's conception of a proxy semantics. Intuitively, we want to say that the proxy language must mirror English in point of semantic structure. We should not make this a condition of correctness, or else the project of proxy semantics will be unable to contribute to our understanding of what the semantic structure of English is. Davidson's finite axiomatization constraint is an attempt to require proxy semantics to recognize semantic structure, without explicit allusion to this notion. However, it fails in this aim: even if it forces the recognition of some structure, it does not guarantee recognition of the *right* structure. For example, it does not prevent a proxy semantics, adequate by Davidson's standards, from regarding "Socrates is wise" as a conjunction, since the requirement of finite axiomatization would not be flouted by a finite translation manual with the rule that an English atom of the form "x is wise" be translated into $\mathbf{Q}^\star$ by something of the form "Wise x & $(2 + 2 = 4)$". Moreover, it would not give a basis for choosing between proxies with binary quantifiers and proxies with unary quantifiers.

When we speak of the "right" semantic structure, or of the "intuition" that English quantifiers are binary, to what standard are we appealing? An extreme Davidsonian might argue that there is no intelligible standard beyond what emerges from the project of proxy semantics. To object that, for example, it is "counterintuitive" to regard "kicked" as a 3-place predicate is to invite the response that the intuition is worthless. Semantic classifications have no role other than to assist in the specification of truth conditions of all the sentences of a language on the basis of the contributions of subsentential expressions. If this is provided by a theory which regards "kicked" as 3-place, there is no room for a standard relative to which this is the wrong semantic classification.

Davidson at one point suggested ([1967a], p. 25) that semantic structure mirrors the structure of our ability to speak and understand a language. This suggests that an appropriate standard is *psychological*: semantic structure should reveal the features which one who understands a sentence exploits in coming to understand it. One cannot tell apriori what these features are, though certain "intuitions", for example, that English quantifications are not all quantified conditionals, constitute our unsubstantiated hunches. These need to be tested empirically, since we know that introspection is not always a reliable guide to our mental processes.

Empirical data for such views could be obtained in a number of ways, of which the following is an extreme example. Suppose we found that severing a certain nerve in the brain had the effect of making a person incapable of understanding explicit conditionals, sentences like "if John is a man, then he is happy". Suppose that this same person continued to understand quantifications like "All men are happy". This would be evidence that the semantic structure of English quantifications is not that suggested by their Q-formalizations.

The criterion of psychological correctness provides a basis for preferring one rather than another of two logical form proposals which meet all the other criteria. It sets a standard for semantic structure that is not relative to a proxy language, and harmonizes well with the tradition. For example, Russell on more than one occasion said, concerning a logical form proposal, that it gave a better picture of what was going on in the mind of a person using or understanding a sentence (e.g. [1912], p. 29). This would be assured, if logical form mirrors actual processing mechanisms. This makes logical form the form of thought as well as the form of language.

The various constraints on logical form, regarded as a proxy in **Q***, designed to serve the purposes of proxy semantics, are that: (i) the logical form should coincide in truth conditions with the English sentence; (ii) semantics and translation manual should be finitely axiomatized; (iii) English arguments we intuitively count as formally valid should correspond to **Q**-valid proxy arguments; and (iv) the semantics for **Q*** should in some way reflect the processing mechanisms speakers of English actually employ. One ought to be cautious in supposing that, by these standards, English sentences have unique logical forms.

For some English sentences, there is room for doubt whether there is even one logical form meeting all these conditions. We have already registered the thought that one who understands an English quantification like "All men are happy" may not need to bring to bear an understanding of the conditional, so that the **Q***-formalization may not meet all the envisaged constraints. Various other constructions raise the same worry, for example:

4) Tom is married.

This needs to be **Q***-formalized as an existential quantification, so as to represent the argument "Tom is married to Susan, so Tom is married" as formally valid. Yet it is very much an open question whether speakers invoke their understanding of the existential quantifier in understanding (**4**), so the fourth constraint, (iv) of the preceding paragraph, may not be met. One reason this problem arises is that **Q**-predicates are of fixed degrees (unary, binary or whatever) whereas arguably there are English predicates of variable degree, like "are compatriots", which seems to make a sentence however many names (separated by "and") we put in front of it. This kind of problem may show not a problem with the general methodology, but an inadequacy of

Q*. Future work may deliver languages as perspicuous as **Q*** but with predicates of variable degree.

One should be cautious in supposing that **Q*** is uniquely well suited to be the proxy language. Russell's theory of descriptions seems much less likely to meet criterion (iv) (mirroring processing mechanisms) than one which treats definite descriptions by proxy languages closer in form to English: either those which treat "the" as a binary quantifier (e.g. Neale [1990]) or those which treat it as a singular term (e.g. Burge [1974]).

The proxy language approach to logical form takes for granted that it must be possible to provide compositional semantics for English, but this assumption can be disputed. It is based on the idea that we come to understand sentences in virtue of our understanding of the meanings of their parts. While this may be true in some sense, there is room for doubt whether the meanings of the parts *entail* the meaning of the whole. Perhaps we often understand sentences by guesswork and probabilistic reasoning, so that there is no clear demarcation between our specifically linguistic competence and our general cognitive abilities. For example, it would appear that no correct assignment of semantic properties to "carpet", "vacuum", "sweeper" and "cleaner" *entails* that a carpet sweeper is something to sweep carpets with, whereas a vacuum cleaner is not for cleaning vacuums. Yet speakers of English have no trouble picking up the use of these compounds once they understand their constituents. Another example is the use of demonstratives, with respect to which general knowledge about what objects are likely to be conspicuous for speakers plays a crucial role in interpretation. Context plays a role in such cases, but perhaps it is one which can be brought under the heel of theory.

A more radical challenge is that context can exert an influence that could not be systematized in a way consistent with the traditional conception of compositional semantics. On this conception, meaning determines truth conditions; so, relative to the same state of affairs, a sentence with a single meaning can have only one truth value. Suppose some brown leaves have been painted green and consider an utterance, concerning those leaves, of "Those leaves are green". Suppose we are part of a commission inspecting Vietnam to determine whether Pentagon denials that it has used exfoliants are true. Brown leaves are signs of the early stages of the action of exfoliants. Charles Travis has suggested that, relative to these concerns, the truth about the painted leaves is that they are not green. The utterance has a meaning upon which the facts just described make it false. Now suppose that we are trying to select camouflage material. Only green things will do, and more or less anything green will do. Relative to these concerns, Travis suggests the truth about the leaves is that they are green. The utterance has a meaning upon which the same leaf-related facts make it false. Yet the sentence uttered is not ambiguous, so it ought to have only one meaning, determined by the meaning of its parts; so relative to a single state of affairs, it ought to have only one truth value, yet it seems to have two. This gives rise to doubt whether there is, in every case, anything both compositionally determined and determinative of truth conditions.

The conception of logical form guided by proxy semantics carries a further debatable commitment. Whereas we incline to think of our language as public, in the sense of being invariant among a community of speakers of (what we perhaps fecklessly call) a single language, there is no inference from this to the conclusion that the common language is implemented in the same way in the psychology of the individual speakers. Proper names are a much discussed case in which it seems quite likely that there is individual variation (and Frege, at least tentatively ([1892b], first footnote), and Russell quite explicitly ([1912] p. 29) said that there was). It may well be that my ability to use a name like "London" is bound up with various things I know about the city, and that your ability is bound up with a number of different things you know about the city. Even though we mean the same thing by the word (judged by ordinary standards), the underlying psychological processes may be different. If semantic structure has to mirror this underlying psychology, it may lose touch with the shared language. Russell thought that this was just how things are, which is why he firmly rejected the classification of his theories of names and descriptions as contributions to the philosophy of language as opposed to the philosophy of thought or judgement. What we think of as a single common language is no such thing, but is an artefact of a certain isomorphism among the way individual thinkers use words. For Russell, the fundamental subject matter was the individual thinker.

In its early days, the project of formalization seemed to give centre stage to notions of validity and proof; but more ambitious logical form proposals cannot be defended with just these materials. A conception of logical form as a proxy in a formalized language which will serve the purposes of semantics is better able to justify such claims. It faces the difficulty that logical form may not be unique, as different proxies may serve the semantic task equally well. If at this point we try to narrow the possibilities by requiring proxies to be psychologically realistic, to mirror somehow the semantic processing speakers actually engage in, we reach a position which, however controversial, is in some ways quite close in spirit to that which originally animated the project. In particular, it helps explain why the founding fathers of the project saw thought or judgement as more fundamental than language.

7 Conclusion

The original motivation for the introduction of artificial languages into logical studies was the lack of perspicuity in natural languages. If the only ambition is to use the perspicuity of an artificial language to attain some generalizations about validity in natural language, the connection between the artificial and the natural is unproblematic. The adequacy of a formalization ensures that if the formalization is valid, so is what it formalizes; and the notion that the formalization is an idealization diminishes any anxieties about validities in the natural language that are not reflected in adequate formalizations.

Once the aim is to segregate out *formal* validity in the natural language, a tension arises. This is because there are independent, even if vague, tests for whether or not a valid argument in natural language is formally valid. It turns out that arguments in natural language which, though valid, are not formally valid (for example, arguments involving adjectival detachment), can be adequately and validly formalized. It would seem a pity to pass up this opportunity; but seizing it is inconsistent with concentrating upon only formal validity in the natural language.

The relationship between the artificial and the natural seems most strained in connection with exotic logical form proposals, for example the view that English universal quantifications are quantifications of conditionals. Making sense of such claims requires Davidson's conception of logical form, but this does not necessarily substantiate the claims. This is because resistant intuitions (for example, the intuition that English quantifiers are binary) are some evidence, albeit inconclusive, that the logical form proposals, even if they get the truth conditions right within the project of proxy semantics, do not satisfy the psychological constraint.

The semantics that can be provided for languages like **Q*** offer a model of rigour and precision, and have provided an essential stimulus to research. What is overambitious is to suppose that these semantics are already, by proxy, what we seek for our natural language, since there is little evidence that they satisfy the psychological constraint. Indeed, it is overambitious to *assume* that compositional semantics of any form are possible for natural languages. But if they are not, that will be a most important fact, and one only statable against the background of what compositional semantics for artificial languages are like.

Davidson's conception offers the most ambitious prospect for the project of formalization, but let us not underestimate its other achievements. First, the very thought that specifying truth conditions is a useful contribution to specifying meaning (and perhaps sometimes exhausts what is required) grew up in the context of the project. Looking in his (somewhat distorting) rear-view mirror, Russell said that in his early years he thought that the logician's task in connection with "the" was to identify some weird abstract object to which it referred (Russell [1959], p. 150). Specification of the way in which the truth of "the"-sentences depends upon the truth of their components is a vastly improved goal. Secondly, it is surprising how much room for disagreement there is about the truth conditions of the sentences of the language we actually speak. We have encountered cases in connection with apparently very simple sentences: universal and existential quantifications, and sentences containing definite descriptions. Such disagreements could in practice only be discovered and expressed in connection with attempts at formalization in artificial languages for which truth conditions can be precisely stated. Thirdly, it is hard to overestimate the importance of formalization in the resolution of ambiguity. For example, once you have been introduced to the quantifier shift fallacy in the context of formalization, your ability to spot instances of it in English is greatly enhanced.

My concern with formal logic has been guided by the limited aim of describing and understanding the project of formalization. Formal logic is a discipline – a branch of mathematics – which can be pursued for its own sake and with no eye to this project. Many of the results obtained in this discipline have – or have been claimed to have – a philosophical importance of a kind not envisaged within the project of formalization.[8]

Bibliographical notes

The traditional view of logical form derives from Frege and Russell. Neither of them is very explicit. Russell comes closest, in the quotation with which this book opened: "Some kind of knowledge of logical forms, though with most people it is not explicit, is involved in all understanding of discourse" [1914], p. 53.

For the Tractarian vision in its purest form, see Wittgenstein [1921]. Russell's version [1918] already contains doubts. For its repudiation, see Wittgenstein [1932].

For an introduction to the techniques of natural deduction see Lemmon [1965] or (for more advanced texts) Prawitz [1965] and Tennant [1978].

For discussions of the logical constants, see Peacocke [1973] and [1987], Harman [1986], Prawitz [1977] and [1979] and Warmbrōd [1999]. For logical truth, see Quine [1960], esp. §13, Peacocke [1987]. Quine's attacks on meaning begin with [1951] and are continued in many later works, e.g. [1960]. Attacks on this position include those by Chomsky [1975], Evans [1975] and Boghossian [1996].

For a criticism of certain uses of schematic letters (like the name- and predicate-letters of **Q**), see Smiley [1982].

For compositional semantics see Davidson [1965], [1967a], [1970b], [1973], Davies, [1981], Schiffer [1987] and Travis [1996]. For the relation between semantics and the psychology of speakers, see Davies [1987], Evans [1981] and Wright [1981].

For Davidson's conception of logical form, see especially his [1967b], [1970a] and [1973]. For an excellent presentation and criticism, see Foster [1976]. For a gentle shift of perspective, see Wiggins [1985]. For a quite different interpretation from mine of Davidson's view, see Lycan [1984], pp. 31–2.

Further reading in philosophical logic could well take the form of following up works mentioned in connection with specific discussions. For something both wide-ranging and introductory I recommend Haack [1978]. Gabbay and Guenthner [1983], [1984], [1986] and [1989] contain survey pieces at an advanced level. For shorter introductions see Read [1994], Sainsbury [1996].

8 To choose two examples from dozens: John Lucas [1961] claimed that Gödel's incompleteness theorem shows that men are not mechanisms; and Hilary Putnam [1980] claimed that the Löwenheim–Skolem theorem has farreaching consequences for metaphysics.

Glossary

accessibility: Relation between worlds used to restrict which worlds are relevant to truth upon an interpretation.

actualism: The view that everything is actual.

adequate: A formalization is adequate iff the recovered argument (sentence) says the same as the original English.

ampersand: Name for "&".

analytic: True in virtue of meanings.

antecedent: First component of a conditional.

apriori: A proposition can be known apriori just on condition that it can be known without recourse to experience or experiment.

argument: A collection of a propositions, one of which is designated as the *conclusion*, the remainder being the *premises*.

argument-claim: An argument-claim refers to an argument, and says of it either that it is valid (a positive argument-claim), or that it is invalid (a negative argument-claim).

argument-form: Pattern from which an argument results by replacing letters or variables by expressions from a natural language.

assertible: An utterance's degree of assertibility varies inversely with the extent to which it is open to criticism, on any grounds at all. Being misleading or uninformative will lower an utterance's degree of assertibility.

atom of Q: a predicate or predicate-letter of degree n combined with n names.

binary quantifier: One which takes two open sentences to form a sentence.

binary sentence connective: One which takes two sentences to make a sentence.

box arrow: Name of "$\Box\!\!\rightarrow$", where $p \Box\!\!\rightarrow q$ corresponds to "if it had been the case that p it would have been the case the q".

box: Name of "$\Box$" (corresponding to "necessarily").

C: A sentence is a logical truth iff its truth is determined just by the meanings of the logical constants it contains.

cancellable: If A implicates but does not entail B, then "A but not B" is consistent. Grice called this the cancellability of the implicature.

cardinality: The cardinality of a set is the number which specifies how many members the set has. The cardinality of the set of the apostles is 12; of the set of flying horses, 0.

cement criterion: Logical constants are expressions which serve to combine the primitive expressions of the language into more complex ones.

complete: π (a system of proofs) is complete with respect to σ (a semantics) iff: if $\Gamma \models_\sigma X$, then $\Gamma \vdash_\pi X$.

component: Sentence which a sentence connective takes to make a complex sentence.

compositional semantic theory: Theory assigning properties to words and modes of composition, on the basis of which the meanings or truth conditions of all the sentences of the language can be derived.

conditional probability: The probability of A, given B, written: $Pr(A \mid B)$.

conditional proof: If $[A_1, \ldots A_n, B \models C]$ then $[A_1, \ldots A_n \models$ if B then $C]$.

conditionals: Sentences of the form "if . . . then . . .".

consequent: Second component of a conditional.

constructional history: Metaphorically, the steps taken in building up a sentence from the symbols of the language. Literally, the sequence of sentences that have to be cited in using the definition of *sentence* to establish that a given expression is a sentence.

contingent: A proposition is contingent (or contingently true) iff it is true but is not necessarily true.

contradictories: Two propositions are contradictories iff it is logically impossible for both to be true and logically impossible for both to be false.

contraposition: For "if", this obtains only if: [if A then not-B] ⊨ [if B then not-A].

correspondence requirement: If the correspondence schema associated with a formalization has it that, say, "F" corresponds to ". . .", then that actual expression, ". . .", must occur in the sentence of natural language which is formalized.

correspondence scheme (for a formalization): Shows which English expressions are replaced by which letters.

counterexample to the validity of an argument: An interpretation upon which the premise(s) are true and the conclusion false.

de dicto necessity: Necessity which is not de re.

de re necessity: QN-formalizable by a sentence in which either there is a name within the scope of some modal operator or a modal operator within the scope of a quantifier.

decision procedure: A mechanical method for determining, with respect to an arbitrary sentence, and in a finite number of steps, whether or not it is valid. Truth table methods serve as a decision procedure for **P**. There is no decision procedure for **Q**.

declarative sentence: One that could be used, and typically is used, to make an assertion, to affirm that something is or is not the case.

deductive logic: The study of what it is for conclusions to *follow from* premises.

deep: One adequate formalization is deeper than another iff it reveals more of the structure of what it formalizes.

definite description: Expression of the form "The so-and-so".

degree: The degree of a predicate is the number of names it takes to form a sentence.

diamond: Name of "◊" (corresponding to "possibly").

domain: The domain of an interpretation is a set of entities ("individuals"). The interpretation may assign to symbols only entities in its domain, or sets constructed from these.

dominant operator: Expression taking the whole of the sentence in question as its scope.

dominates: An occurrence of a sentence connective dominates the sentence which is its scope.

elimination rule: Specifies what conclusions one can derive from premises containing a sentence dominated by a given expression.

empty domain: Domain containing no entities.

entails: A entails B iff it is logically impossible for A to be true without B being true.

equivalent: There are many different standards of equivalence. One is: sentences are equivalent iff every **P**-interpretation assigns them the same truth value.

ersatzism: Quantifierism plus actualism.

exclusive disjunction: The truth function which associates a pair of different truth values with true and a pair of the same truth values with false.

existential (universal) quantification: Sentence dominated by an existential (universal) quantifier.

existential quantifier: In English, "there is a" or "some" or equivalent expressions. In **Q**, "∃".

extension: Of a sentence, its truth value; of a name, its bearer; of a predicate of degree n, the set of n-tuples of which it is true.

extensional: A sentence is extensional iff its truth or falsity turns on nothing except what objects the names it contains refer to, of what things the predicates it contains are true, and the truth values of any unstructured sentences it contains. A sentence is *extensional with respect to a position for an expression (sentence, name or predicate)* iff replacing an expression in that position with any other expression having the same extension leaves the truth value of the whole sentence unchanged. A sentence is *extensional, tout court,* iff it is extensional with respect to all its positions for sentences, names and predicates.

finite axiomatization constraint: Applied to proxy semantics, the constraint is that there be a finite manual for translating English into the proxy language, and a finitely axiomatized semantic theory for the latter.

formal validity: An argument is formally valid iff it is valid in virtue of its form or pattern.

iff: If and only if.

implicature: What is conveyed but not said.

inclusive disjunction: The truth function expressed by "$\vee$".

inconsistent: A collection of propositions is *inconsistent* iff it is logically impossible for all of them to be true.

indicative conditionals: exemplified by (**2.4.34**): "If Oswald didn't shoot Kennedy, someone else did."

indicative sentence: Contrasted with *subjunctive sentence,* and introduced by examples.

intended interpretation: An interpretation of a formalization which relative to an English sentence to be formalized, assigns entities in keeping with the meaning of the corresponding English expressions. (For example, relative to some English sentence, an intended interpretation of a **P**-sentence is one which assigns to all the **P**-letters of the sentence the truth value the corresponding English sentences actually possess.)

intension: The set of all ordered pairs whose first member is a possible world and whose second member is the extension of the expression with respect to that world.

interpretation: Assignment of entities to the non-logical symbols of a language.

intersection: The intersection of two sets, $\Gamma \cap \Gamma'$, is the set of things belonging to both Γ and Γ'.

introduction rule: Specifies from what premises one can derive a sentence dominated by a given expression.

invalid: Not valid.

language of propositional calculus, of predicate calculus: These are the languages **P** and **Q** introduced in chapters **2** and **4** respectively.

law of identity: The validity of every instance of "$\alpha = \alpha$".

lexical ambiguity: Ambiguity in a sentence due to some "lexical item", i.e., roughly, word, which it contains.

logical constant: See list at (**1.11.1**).

modal collapse: The property (possessed by S5 modal languages) that given a non-modal sentence, X, there are only two non-equivalent fresh sentences, $\Box X$ and $\Diamond X$, that can be formed from it just by adding modal operators.

modal idioms: Words which express modal notions, for example "possibly", "necessarily", "must", "may", "might".

modal notions: Notions expressed by modal idioms, for example the notions of necessity and possibility.

modal realism: The view that modal facts are real and mind-independent.

modus ponens: The principle licensing inference from a conditional, together with its antecedent, to its consequent. It can be expressed by the assertion: $[A, \text{if } A \text{ then } B] \models [B]$.

monotonic: Deductive validity is monotonic in that, if you start with a deductively valid argument, then, no matter what you *add* to the premises, you will end up with a deductively valid argument. Cf. (**1.6.5**).

naive syntactic test: Two expressions belong in the same syntactic category, by the naive syntactic test, iff you can replace one by the other, wherever it occurs, without turning sense into nonsense.

name quantifier: Quantifier whose variables occupy a position in a sentence appropriate to a name (name-position).

name-letters: $\alpha, \beta, \gamma, \alpha', \ldots$ etc. These are used to correspond to ordinary English names like "John".

negation: The negation of a proposition is what results from inserting "not" (or some equivalent expression) into it in such a way as to form a contradictory of it.

n-tuple: A sequence of n objects. For example, a 3-tuple (or triple) is a sequence of three objects. Sequences differ from sets in that sets introduce no order into their members whereas sequences do.

objectual quantifier: An objectual quantifier (for example, the **Q**-quantifiers) is one which is interpreted in terms of how things are with some or all objects.

one-one mapping: a one-one mapping between the sets Γ and Γ' is a relation, R, such that each member of Γ bears R to exactly one member of Γ' and each member of Γ' bears R to exactly one member of Γ. Since we can think of a relation as a set of n-tuples, we can think of a one-one mapping between Γ and Γ' as a set Σ of ordered pairs meeting the following conditions: for any x in Γ, there is exactly one y in Γ' such that $<x, y>$ is in Σ; and for any x in Γ', there is exactly one y in Γ such that $<y, x>$ is in Σ.

opaque context with respect to names: One in which there is no guarantee that two co-referring names can be substituted for one another without affecting the truth value of the whole.

open sentence: Expression resulting from a **Q**-sentence by replacing some name-letter by a variable not already contained in the sentence.

operators: The sentence-connectives of **P**, together with "$\forall$" (the universal quantifier, corresponding to "all") and "$\exists$" (the existential quantifier, corresponding to "some" or "a").

ordered pair: Sequence of two objects.

perspicuous: Language whose syntactic and semantic categories coincide.

persuasive: An argument is persuasive for a person only if he is willing to accept each of the premises but, before the argument is propounded to him, is unwilling to accept the conclusion.

PN-validity: $X_1, \ldots X_n \models_{PN} Y$ iff for all **PN**-interpretations i and all worlds w such that all of $X_1, \ldots X_n$ are true at w upon i, Y is true at w upon i.

Pr(A) = 0.5: The probability of A is a half.

Pr(B|A): the probability of B, given A; equivalently, the conditional probability of B upon A. Equivalent to: $Pr(B \& A) \div Pr(A)$, if $Pr(A) > 0$.

predicate: An expression which takes one or more names to form a sentence.

predicate position: The kind of position that a predicate may properly occupy in a sentence.

predicate quantifier: Quantifier whose variables occupy a position in a sentence appropriate to a predicate (predicate position).

predicate-letters: F, G, H, F', . . . etc. These are used to correspond to English verbs, like "runs", some adjectives, like "is hungry", and some nouns, like "is a man".

predicative adjective: The following condition is a rough test for the predicativity of any adjective, A: where n is any name, C any noun which A qualifies, n is a (or an) $AC \models n$ is A and n is a (or an) C.

probabilistically valid: $[A_1, \ldots, A_n \vdash C]$ iff necessarily the uncertainty of the conclusion does not exceed the sum of the uncertainties of the premises.

proper subset: Γ is a proper subset of Γ' ($\Gamma \supset \Gamma'$) iff every member of Γ is a member of Γ' and some member of Γ' is not a member of Γ.

proposition: Something which can be asserted, believed, denied, supposed, etc., and which can be expressed, in a given context, by a meaningful, declarative, indicative sentence.

proxy language: Language given semantics directly, and thus giving proxy semantics for any language translatable into it.

proxy semantics: The semantics for a language given by translating it into another, and giving semantics for the latter.

P-validity: An argument in **P**, $X_1, \ldots X_n; Y$, is **P**-*valid* iff every interpretation upon which all the premises are true is one upon which the conclusion is true. Abbreviation: $X_1, \ldots X_n \models_P Y$.

QN-validity: $X_1, \ldots X_n \models_{QN} Y$ iff for all **QN**-interpretations i and all worlds w such that all of $X_1, \ldots X_n$ are true at w upon i, Y is true at w upon i.

quantifier shift fallacy: Argument from premise of the form $\forall\exists$ to conclusion of the form $\exists\forall$.

quantifierism: The view that ordinary modal idioms are best represented as quantifications over possible worlds.

Q-validity: An argument is **Q**-*valid* iff every interpretation upon which all the premises are true is one upon which the conclusion is true.

recovered argument: The result of replacing the letters in a formalization by the corresponding English expressions (as determined by the associated correspondence scheme), and then replacing the connectives by the corresponding English connectives.

reductio ad absurdum: Reasoning which uses the validity of an argument, together with the falsehood of the conclusion, to infer the falsehood of one of the premises.

reflexivity: To say that validity is reflexive is to say that: If C is among the $A_1, \ldots A_n$, then $[A_1, \ldots A_n \models C]$.

resultant: Sentence formed by a sentence connective from one or more components.

robust with respect to the antecedent: Evidence for the antecedent will not undermine evidence for the conditional.

rules criterion: Logical constants are expressions whose meaning is exhaustively determined by introduction and elimination rules.

rules of interpretation: Rules which determine for each sentence of a language, relative to the entities assigned by an interpretation to the symbols of the language, a condition under which the sentence is true (false) on the interpretation.

rules of proof: Rules stated in terms of the physical make-up of sentences, specifying derivations of sentences from others.

scope (of an occurrence of a sentence connective): The shortest sentence in which it occurs.

semantic category: e_1 and e_2 belong to the same semantic category iff either they are assigned the same kind of entity by the interpretation rules, or else they are treated by the same kind of interpretation rule.

semantics: Systematic specifications of meanings for all the sentences of a language. For example, the interpretations and interpretation rules for a language.

sentence connective: Expression which forms a sentence out of one or more sentences.

sequence: A series of objects in a fixed order. Thus one could write the series consisting just of Reagan and the number seven as <Reagan, 7>. The importance of the order is brought out by the fact that <Reagan, 7> ≠ <7, Reagan>. An n-tuple is a sequence with n members.

singleton: A one-membered sequence, or 1-tuple.

sound: An argument is sound iff it is valid and has true premises. A system of proof, π is sound, with respect to a semantics, σ, iff: if $\Gamma \vdash_\pi X$ then $\Gamma \models_\sigma X$.

strong necessity: $\Box X$ is true on an interpretation i iff X is true on i at every world.

stronger than: A is stronger than B iff A entails B but B does not entail A.

structural ambiguity: Non-lexical ambiguity in a sentence.

structural features: Facts about the recurrence of non-logical elements in a sentence.

subjunctive conditionals: Exemplified by **(2.4.35)** "If Oswald hadn't shot Kennedy, someone else would have."

substitution test: Replacing a component by one with the same truth value invariably leaves the truth value of the resultant unchanged. The test is for the truth functionality of the connective dominating the component to which the test is applied.

substitutional quantifier: A quantifier interpreted in terms of the truth or otherwise of sentences resulting by replacing the quantified variables by constants.

substitutivity of identicals: An interpretation upon which "$\alpha = \beta$" is true is one upon which: ". . . α . . ." is true iff ". . . β . . ." is true.

syntactic category: e_1 and e_2 belong to the same syntactic category iff, for every sentence s containing e_1, the result of replacing e_1 by e_2 is a meaningful sentence.

syntactic irregularity: Failure of match between logical properties and syntactic properties as determined by the naive syntactic test.

syntax or grammar: A set of rules which determines how sentences are constructed from the language's vocabulary.

the psychological constraint: is met by a logical form proposal only if the logical form reflects the way in which the corresponding sentence of natural language is processed in an act of understanding.

tilde: Name for "$\neg$".

topic neutral criterion: Logical constants are expressions which introduce no special subject matter, or understanding which does not require learning about any actual or possible objects.

Tractarian vision: View, found in Wittgenstein's *Tractatus*, that all validity is formal validity.

transitivity: Of "$\rightarrow$": $[X \rightarrow Y, Y \rightarrow Z] \vdash_P [X \rightarrow Z]$. Of "if": [if A then B, if B then C] $\models$ [if A then C]. Of validity: If $[A_1, \ldots A_n \models C]$, and $[B_1, \ldots B_k, C \models D]$, then $[A_1, \ldots A_n, B_1, \ldots B_k \models D]$.

truth conditions: Circumstances under which a sentence is or would be true.

truth functional: A sentence connective is truth functional iff whether or not any resultant sentence it forms is true or false is determined completely by whether its components are true or false.

truth tables: Tabular representation of the way in which truth values of complex sentences depend upon the truth values of their compounds.

truth value: A sentence possesses the truth value *true* just on condition that it is true; the truth value *false* just on condition that it is false.

turnstile: Name for $\models$ (double turnstile) and $\vdash$ (single turnstile).

unary predicate: One which takes one name to form a sentence.

unary quantifier: One which takes one open sentence to form a sentence.

unary sentence connective: One which takes one sentence to form a sentence.

universal quantifier: In English "all" or "every" or equivalent expressions. In **Q**, "$\forall$".

valid in virtue of P-logical form: Has an adequate and **P**-valid **P**-formalization.

valid: An argument is valid iff it is logically impossible for all the premises to be true yet the conclusion false. Compare also **P**-valid and **Q**-valid.

variables: $x, y, z, x', \ldots$ etc.

vel: Name for "$\vee$".

verbs of propositional attitude: Verbs expressing the *attitude* (e.g. belief, saying, wondering) adopted by a subject to a proposition.

weak necessity: $\Box X$ is true on an interpretation i iff X is true on i at every world at which all the objects which i assigns to any name-letters in X exist.

weaker than: A is weaker than B iff B entails A but A does not entail B.

wide scope: An occurrence of a sentence connective has wide scope relative to another occurrence of a connective iff the scope of the latter is a proper part of the scope of the former.

$\mathbf{X}\frac{n}{v}$**:** The result of replacing every occurrence of the variable v in X by the first name-letter n not occurring in X.

List of symbols

Chapter 1

A, B, C	etc. are used as variables ranging over English sentences or propositions. (p. 23)
$A_1, \ldots A_n; C$	gives the general form of an argument (in English) with n premises, $A_1, \ldots A_n$, and conclusion C. (p. 23)
$\models$	called "double turnstile": "$A_1, \ldots A_n \models C$" abbreviates: "$A_1, \ldots A_n; C$" is valid. As a special case, "$(A_1, \ldots A_n) \models$" abbreviates: "$(A_1, \ldots A_n)$" is inconsistent. As a further special case, "$\models A$" abbreviates: "it is logically impossible for A to be false". (p. 24)
$\not\models$	"$A_1 \ldots A_n \not\models C$" abbreviates: "$A_1, \ldots A_n; C$" is not valid. (p. 24)
F, G, H	etc. are used in chapter **1** to mark places which can be occupied by predicates, and in chapter **4** as predicate-letters (more or less the same role). (p. 39)
α, β, γ	etc. are used in chapter **1** to mark places which can be occupied by names, and in chapter **4** as name-letters (more or less the same role). (p. 39)

Chapter 2

P	the propositional language. (p. 54)
$p, q, r, \ldots$	sentence-letters. (p. 54)
$\neg, \&, \vee, \rightarrow, \leftrightarrow$	symbols of **P**. (pp. 54–5)
X, Y, Z	etc. are used as variables ranging over **P**-sentences (later, over sentences in whatever artificial language is under discussion). (pp. 55, 58)
$\models_P$	abbreviates "**P**-valid". (p. 58)

Chapter 3

$\mid$	$\Pr(A \mid B)$ abbreviates: the probability of A, given B. (p. 127)
$\not\vdash$	$A_1 \ldots, A_n \not\vdash C$ iff necessarily the uncertainty of the conclusion does not exceed the sum of the uncertainties of the premises. (p. 131)

Chapter 4

Q	the quantificational language (sometimes called the language of first order (predicate) logic, or the language of quantification theory). (p. 153)
$\exists$, $\forall$	quantifiers of **Q**. (p. 154)
x, y, z, x'	etc. variables of **Q**. (p. 154)
<. . .>	denotes a sequence. The names of its members fill the dots, distinct names separated by commas. (pp. 157, 400)
$\vDash_{\mathbf{Q}}$	abbreviates "**Q**-valid" (pp. 160, 182).
$\alpha = \beta$, $\alpha \neq \beta$	abbreviate, respectively, "$= \alpha\beta$" and "$\neg = \alpha\beta$". (p. 182)
ι	iota-operator used by Russell to formalize "the" as a quantifier. (p. 194)
T, W	unary quantifiers supposed to correspond to "most" and "few". (p. 225)
Q+	**Q** augmented by "T" and "W". (p. 225)
μ, ϕ	binary quantifiers corresponding to "most" and "few". (p. 227)
QB	**Q** augmented by "μ" and "ϕ". (p. 227)
λ	binary quantifer corresponding to "every". (p. 229)
$\lvert\sigma\rvert$	the cardiality of the set σ. (p. 230)
A, E	substitutional universal and existential quantifiers. (p. 231)
QS	**Q** augmented by names and "A" and "E". (p. 232)
$\triangle$, ∇	unary universal and existential predicate quantifiers. (p. 236)
$f, g \ldots$	variables for "$\triangle$" and "∇". (p. 236)
$\supseteq$	inclusion: $\Gamma \supseteq \Gamma'$ iff every member of Γ is a member of Γ'. (p. 230)
QF	a language with the syntax of **Q** but with interpretation rules not requiring every interpretation to assign every name-letter an object. (p. 241)

Chapter 5

$\Box$	box, corresponding to "necessarily". (p. 258)
$\Diamond$	diamond, corresponding to "possibly". (p. 259)
QN	**Q** augmented by "$\Box$" and "$\Diamond$". (p. 257)
PN	**P** augmented by "$\Box$" and "$\Diamond$". (p. 258)
$\Box\!\!\rightarrow$	box arrow: $p \Box\!\!\rightarrow q$ corresponds to "if it had been the case that p it would have been the case that q". (p. 272)
PNS	**PN** augmented by "$\Box\!\!\rightarrow$". (p. 272)
$\delta\sigma$	expression denoting 1 or 0, depending as σ is true or false. (p. 296)
W	Predicate constant of **QW** assigned, on every interpretation, the set of all possible worlds. (p. 304)
QW	**Q** augmented by "W", and, later, by "at", and, later, by "w*". (p. 304)
at v	such operators, for some variable v, form open sentences from closed sentences. "At v, p" says, roughly, that p is true at world v. (p. 308)
w*	name in **QW** of the actual world. (p. 310)
Ⓐ	corresponds to "actually". (p. 311)
QC	**QW** augmented by "C". (p. 319)
C	Cxy corresponds, in **QC**, to "x is a counterpart of y". (p. 319)

Chapter 6

$\cup$	set theoretic union: something belongs to $\Gamma \cup \Gamma'$ iff either it belongs to Γ or it belongs to Γ'. (pp. 230, 356)
$\vdash_{\mathbf{P}}$	called single turnstile: the derivability relation in **P**. (p. 358)

Bibliography

Adams, E. W. [1970] "Subjunctive and indicative conditionals", *Foundations of Language* 6, pp. 89–94.

——[1975] *The Logic of Conditionals*. Reidel, Dordrecht.

Anderson, A. R. and Belnap, E. [1975] *Entailment*. Princeton University Press, Princeton.

Anscombe, G. Elizabeth M. [1959] *An Introduction to Wittgenstein's Tractatus*, Hutchinson, London.

Appiah, Anthony [1985] *Assertion and Conditionals*. Cambridge University Press, Cambridge.

Armstrong, David [1978] *Nominalism and Realism: Universals and Scientific Realism*, vol. 1. Cambridge University Press, Cambridge.

Ayer, Alfred, J. [1936] *Language, Truth and Logic*. Gollancz, London; 2nd edn 1946.

Bach, E. [1974] *Syntactic Theory*. Holt, Reinhard and Winston, London.

Bach, Kent [1987] *Thought and Reference*. Clarendon Press, Oxford.

Barcan, Marcus [1947] "The identity of individuals in a strict functional calculus of second order", *Journal of Symbolic Logic* 12, pp. 12–15.

Barwise, J. and Cooper, R. [1981] "Generalized quantifiers and natural language", *Linguistics and Philosophy* 4, pp. 159–219.

Belnap, N. [1962] "Tonk, plonk and plink", *Analysis* 22, pp. 130–4, reprinted in P. Strawson (ed.), *Philosophical Logic*, Oxford University Press, Oxford, pp. 132–7.

Bencivenga, Ermanno [1986] "Free logics", in Gabbay and Guenthner [1986], pp. 373–426.

Bennett, Jonathan [1974] "Counterfactuals and possible worlds", *Canadian Journal of Philosophy* 4, pp. 381–402.

——[1995] "Classifying conditionals: the traditional way was right." *Mind* 104, pp. 331–54.

Blackburn, Simon [1986a] "How can we tell whether a commitment has a truth condition?" in Charles Travis (ed.), *Meaning and Interpretation*, Blackwell, Oxford, pp. 201–32.

——[1986b] "Morals and Modals", in Graham Macdonald and Crispin Wright (eds), *Fact, Science and Morality*. Blackwell, Oxford, pp. 119–41.

Boghossian, Paul [1996] "Analyticity reconsidered", *Noûs* 30, pp. 360–91.

Boolos, George [1975] "On second-order logic", *Journal of Philosophy* 72, pp. 509–27; reprinted in his *Logic, Logic and Logic*. Harvard University Press, Cambridge, Mass., 1998, pp. 37–53.

Boolos, George and Jeffrey, Richard C. [1974] *Computability and Logic* (3rd edn 1989), Cambridge University Press, Cambridge.

Bostock, David [1997] *Intermediate Logic*. Clarendon Press, Oxford.

Bull, Robert, A. and Segerberg, Krister [1983] "Basic modal logic", in Gabbay and Guenthner [1983], pp. 1–88.

Burge, Tyler [1974] "Truth and singular terms." *Noûs* 8, 309–25; reprinted in Lambert [1991a], pp. 189–204.

——[1986] "On Davidson's 'Saying that'", in LePore [1986], pp. 190–208.

Carroll, Lewis [1872] *Through the Looking Glass and What Alice Found There*. Macmillan, London.

Chellas, B. F. [1980] *Modal Logic*. Cambridge University Press, Cambridge.

Chikara, Charles S. [1998] *The Worlds of Possibility: Modal Realism and the Semantics of Modal Logic*. Clarendon Press, Oxford.

Chomsky, Noam [1975] *Reflections on Language*. Pantheon, New York.

Clark, Romane [1970] "Concerning the logic of predicate modifiers", *Noûs* 4, pp. 311–35.

Cohen, L. Jonathan [1962] *The Diversity of Meaning*. Methuen, London.

Craig, Edward [1985] "Arithmetic and fact", in P. Hacking (ed.), *Exercises in Analysis*, Cambridge University Press, Cambridge, pp. 89–112.

Davidson, Donald [1965] "Theories of meaning and learnable languages", in Yehoshua Bar-Hillel (ed.), *Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science*, North-Holland, Amsterdam; reprinted in Davidson [1984], pp. 3–15.

——[1967a] "Truth and meaning", *Synthese* 7, pp. 304–23; reprinted in Davidson [1984], pp. 17–36.

——[1967b] "The logical form of action sentences", in N. Rescher (ed.), *The Logic of Decision and Action*, University of Pittsburgh Press, Pittsburgh, Pa., pp. 115–20; reprinted in Davidson [1980], pp. 105–21.

——[1967c] "Causal relations", *Journal of Philosophy* 64, pp. 691–703; reprinted in E. Sosa (ed.), *Causation and Conditionals*, Oxford University Press, Oxford (1975), pp. 82–94, and in Davidson [1980], pp. 149–62.

——[1969] "On saying that", *Synthese* 19, pp. 130–46; reprinted in D. Davidson and J. Hintikka (eds), *Words and Objections*, Reidel, Dordrecht (1969), pp. 158–74; and in Davidson [1984], pp. 93–108.

——[1970a] "Action and reaction", *Inquiry* 13, pp. 140–8; reprinted as "Reply to Cargile", in Davidson [1984], pp. 137–46.

Davidson, Donald [1970b] "Semantics for natural languages", in *Linguaggi nella società e nella tecnica*, Milan; reprinted in Davidson [1984], pp. 55–64.
——[1973] "In defense of convention T", in H. Leblanc (ed.), *Truth, Syntax and Modality*, North-Holland, Amsterdam; reprinted in Davidson [1984], pp. 65–75.
——[1977] "The method of truth in metaphysics", in P. French, T. Uehling and H. Wettstein (eds), *Midwest Studies in Philosophy, II: Studies in the Philosophy of Language*, University of Minnesota Press, Minneapolis, pp. 244–54; reprinted in Davidson [1984], pp. 199–214.
——[1980] *Essays on Actions and Events*. Clarendon Press, Oxford.
——[1984] *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford.
Davies, Martin, K. [1981] *Meaning, Quantification, Necessity: Themes in Philosophical Logic*. Routledge and Kegan Paul, London.
——[1983] "Meaning and structure", *Philosophia* 13, pp. 13–33.
——[1987] "Tacit knowledge and semantic theory: Can a 5% difference matter?" *Mind* 96, pp. 441–62.
Delong, Howard [1970] *A Profile of Mathematical Logic*. Addison-Wesley Publishing Co., Reading, Mass.
Donnellan, Keith [1966] "Reference and definite descriptions", *Philosophical Review* 77, pp. 203–15.
Dudman, V. H. [1984a] "Conditional interpretations of 'if'-sentences", *Australian Journal of Linguistics* 4, pp. 143–204; reprinted in Jackson [1991], pp. 202–32.
——[1984b] "Parsing 'if'-sentences", *Analysis* 44, pp. 145–53.
——[1987] "Appiah on 'if'", *Analysis* 47, pp. 74–9.
——[1988] "Indicative and subjunctive", *Analysis* 48, pp. 113–22.
Dummett, Michael [1973] *Frege: Philosophy of Language*. Duckworth, London.
——[1981] *The Interpretation of Frege's Philosophy*. Duckworth, London.
——[1991] *The Logical Basis of Metaphysics*. Duckworth, London.
Edgington, Dorothy [1986] "Do conditionals have truth-conditions?" *Critica* 18, pp. 3–30; reprinted in Jackson [1991], pp. 176–201.
——[1988] "Review of Graeme Forbes: *Metaphysics of Modality*", *Philosophical Quarterly* 38, pp. 365–70.
——[1991] "Matter-of-fact conditionals", *Proceedings of the Aristotelian Society; Supplementary Volume* 65, pp. 185–209.
——[1995] "On conditionals", *Mind* 104, pp. 235–329.
——[1997] "Commentary", in Woods [1997], pp. 95–137.
——[2000] "General conditional statements: a response to Kölbel", *Mind* 109, pp. 109–16.
Eels, Ellery and Brian Skyrms (eds) [1994] *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge University Press, Cambridge.

Evans, Gareth [1975] "Identity and predication", *Journal of Philosophy* 72, pp. 343–63; reprinted in Evans [1985], pp. 25–48.
——[1976] "Semantic structure and logical form" in Evans and McDowell [1976], pp. 199–222; reprinted in Evans [1985], pp. 49–75.
——[1977a] "Pronouns, quantifiers and relative clauses (I)", *Canadian Journal of Philosophy* 7, pp. 467–536; reprinted in Evans [1985], pp. 76–152.
——[1977b] "Pronouns, quantifiers and relative clauses (II)", *Canadian Journal of Philosophy* 7, pp. 777–97; reprinted in Evans [1985], pp. 153–75.
——[1979] "Reference and contingency", *The Monist* 62, pp. 161–89; reprinted in Evans [1985], pp. 178–213.
——[1981] "Semantic theory and tacit knowledge", in S. Holzman and C. Leich (eds), *Wittgenstein: To Follow a Rule*, Routledge, London.
——[1982] *The Varieties of Reference*. Clarendon Press, Oxford.
——[1985] *Collected Papers*. Clarendon Press, Oxford.
Evans, Gareth and McDowell, John (eds) [1976] *Truth and Meaning*. Oxford University Press, Oxford.
Fine, Kit [1975] "Review of David Lewis: *Counterfactuals*", *Mind* 1975, 84, pp. 451–8.
——[1985] "Plantinga on the reduction of possibilist discourse", in James Tomberlin and Peter Van Inwagen (eds), *Alvin Plantinga: A Profile*, Reidel, Dordrecht, pp. 145–86.
Forbes, Graeme [1985] *The Metaphysics of Modality*. Clarendon Press, Oxford.
Foster, John [1976] "Meaning and truth theory", in Evans and McDowell [1976], pp. 1–32.
Frege, Gottlob [1879] *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Nebert, Halle; English translation in T. W. Bynum (ed. and trans.), *G. Frege, Conceptual Notation and Related Articles*, Clarendon Press, Oxford (1972), pp. 101–23.
——[1884] *Die Grundlagen der Arithmetik. Eine logisech-mathematische Untersuching über den Begreiff der Zahl*. Koebner, Breslau; reprinted with English translation by John L. Austin as *The Foundations of Arithmetic*, Blackwell, Oxford (1950).
——[1892a] "Über Begriff und Gegenstand", *Vierteljahrsschrift für wissenschaftliche Philosophie* 16, pp. 192–205; translated as "On concept and object" in P. Geach and M. Black (eds), *Translations from the Philosophical Writings of Gottlob Frege*, Blackwell, Oxford (1952), pp. 42–55.
——[1892b] "Über Sinn und Bedeutung", *Zeitschrift für Philosophie und philosophische Kritik* 100, pp. 25–50; translated as "On sense and reference" in P. Geach and M. Black (eds), *Translations from the Philosophical Writings of Gottlob Frege*, Blackwell, Oxford (1952), pp. 56–78.
Gabbay, Dov, M. and Guenthner, Franz [1983] *Handbook of Philosophical Logic, Vol. I*. Reidel, Dordrecht.

Gabbay, Dov, M. and Guenthner, Franz [1984] *Handbook of Philosophical Logic, Vol. II.* Reidel, Dordrecht.

——[1986] *Handbook of Philosophical Logic, Vol. III.* Reidel, Dordrecht.

——[1989] *Handbook of Philosophical Logic, Vol. IV.* Reidel, Dordrecht.

Gamut, L. T. F. [1991] *Logic, Language and Meaning.* Chicago University Press, Chicago.

Geach, Peter [1956] "Good and evil", *Analysis* 17, pp. 33–4.

——[1972] "A program for syntax" in D. Davidson and G. Harman (eds), *Semantics of Natural Languages*, Reidel, Dordrecht, pp. 483–97.

Gibbard, Allan [1975] "Contingent identity", *Journal of Philosophical Logic* 4, pp. 187–221.

——[1981] "Two recent theories of conditionals", in Harper, Stalnaker and Pearce [1981], pp. 211–47.

Goguen, J. A. [1969] "The logic of inexact concepts", *Synthese* 19, pp. 325–73.

Goodman, Nelson [1955] *Fact, Fiction and Forecast.* Harvard University Press, Cambridge, Mass.; 2nd edn, Bobbs-Merrill, Indianapolis, 1965.

Grice, H. P. [1961] "The causal theory of perception", *Aristotelian Society, Supplementary Volume* 35, pp. 121–54; reprinted in G. Warnock (ed.), *The Philosophy of Perception*, Oxford University Press, Oxford (1967), pp. 85–112.

——[1975] "Logic and conversation", in Peter Cole and Jerry L. Morgan (eds), *Syntax and Semantics: Vol. 3, Speech Acts*, Academic Press, New York, pp. 41–58.

——[1978] "Further notes on logic and conversation", in Peter Cole (ed.), *Syntax and Semantics: Vol. 9, Pragmatics*, Academic Press, New York, pp. 113–17.

Guttenplan, Samuel [1992] *The Languages of Logic.* Blackwell, Oxford; first published 1986.

Haack, Susan [1974] *Deviant Logic.* Cambridge University Press, Cambridge.

——[1978] *Philosophy of Logics.* Cambridge University Press, Cambridge.

Hale, Bob [1989] "Necessity, caution and scepticism", *Proceedings of the Aristotelian Society, Supplementary Volume* 63, pp. 175–202.

Harman, Gilbert [1986] *Change in View: Principles of Reasoning.* The MIT Press, Cambridge, Massachusetts.

Harper, W. L., Stalnaker, R. and Pearce, G. (eds) [1981] *Ifs: Conditionals, Belief, Decision, Chance, Time.* Reidel, Dordrecht.

Hazen, Allen [1976] "Expressive incompleteness in modal logic", *Journal of Philosophical Logic* 5, pp. 25–46.

——[1979] "Counterpart-theoretic semantics for modal logic", *Journal of Philosophy* 76, pp. 319–38.

Higginbotham, James [1986] "Linguistic theory and Davidson's program in semantics", in LePore [1986], pp. 29–48.

——[1988] "Is semantics necessary?" *Proceedings of the Aristotelian Society* 88, pp. 219–41.

Hochberg, Herbert [1957] "On Pegasizing", *Philosophy and Phenomenological Research* 17; reprinted in Hochberg [1984], pp. 101–4.

——[1984] *Logic, Ontology and Language: Essays on Truth and Reality.* Philosophia Verlag, Munich.

Hodges, Wilfrid [1977] *Logic.* Penguin, London.

——[1983] "Elementary predicate logic", in Gabbay and Guenthner [1983], pp. 1–131.

Hornsby, J. [1977] "Saying of", *Analysis* 37, pp. 177–85.

Hughes, G. E. and Cresswell, Max, J. [1984] *Companion to Modal Logic.* Methuen, London.

——[1996] *A New Introduction to Modal Logic.* Routledge, London.

Jackson, Frank [1979] "On assertion and indicative conditionals", *Philosophical Review* 87, pp. 567–89.

——[1981] "Conditionals and possibilia", *Proceedings of the Aristotelian Society* 81 (1980/1), pp. 125–37.

——[1987] *Conditionals.* Cambridge University Press, Cambridge.

——(ed.) [1991] *Conditionals.* Oxford University Press, Oxford.

Kamp, J. A. W. [1975] "Two theories of adjectives", in Edward L. Keenan (ed.), *Formal Semantics of Natural Language*, Cambridge University Press, Cambridge, pp. 123–55.

Kant, Immanuel [1787] *Critique of Pure Reason*, English translation by Norman Kemp Smith, Methuen, London (1929).

Kaplan, David [1979] "Transworld heir lines", in Loux [1979], pp. 88–109.

Kirwan, Christopher [1978] *Logic and Argument.* Duckworth, London.

Kneale, William and Kneale, Martha [1962] *The Development of Logic.* Oxford University Press, Oxford.

Kölbel, Max [2000] "Edgington on compounds of conditionals", *Mind* 109, pp. 97–107.

Kripke, Saul [1963] "Semantical considerations on modal logic", *Acta Philosophica Fennica* 16, pp. 83–94; reprinted in Linsky [1971], pp. 63–72.

——[1972] "Naming and necessity", in D. Davidson and G. Harman (eds), *Semantics of Natural Language*, Reidel, Dordrecht; reprinted with added preface (referred to as Kripke [1980]).

——[1976] "Is there a problem about substitutional quantification?" in Evans and McDowell [1976], pp. 325–419.

——[1980] *Naming and Necessity.* Blackwell, Oxford.

Lacey, Hugh M. [1971] "Quine on the logic and ontology of time", *Australasian Journal of Philosophy* 49, pp. 47–67.

Lambert, Karel [1965] "On logic and existence", *Notre Dame Journal of Formal Logic* 6, pp. 135–41.

——[1983] *Meinong and the Principle of Independence.* Cambridge University Press, Cambridge.

Lambert, Karel [1991a] *Philosophical Applications of Free Logic*. Oxford University Press, Oxford.

——[1991b] "A theory of definite descriptions", in Lambert [1991a], pp. 17–27.

Le Carré, John [1971] *The Naive and Sentimental Lover*. Hodder and Stoughton, London.

Lehman, Scott [1994] "Strict Fregean free logic", *Journal of Philosophical Logic* 23, pp. 307–36.

Lemmon, Edward J. [1965] *Beginning Logic*. Nelson, London.

Leonard, H. S. [1956] "The logic of existence", *Philosophical Studies* 5, pp. 49–64.

LePore, Ernest [1986] *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Blackwell, Oxford.

Lewis, C. I. and Langford, C. H. [1932] *Symbolic Logic*. Century Co., New York.

Lewis, David [1968] "Counterpart theory and quantified modal logic", *Journal of Philosophy* 65, pp. 17–25; reprinted with added postscripts in Lewis [1983], pp. 26–46.

——[1970] "General semantics", *Synthese* 22, pp. 18–67; reprinted with an added postscript in Lewis [1983], pp. 189–232.

——[1973a] "Counterfactuals and comparative possibility", *Journal of Philosophical Logic* 2, pp. 418–46; reprinted in Harper, Stalnaker and Pearce [1981], pp. 57–85.

——[1973b] *Counterfactuals*. Basil Blackwell, Oxford.

——[1973c] "Causation", *Journal of Philosophy* 70, pp. 556–67; reprinted in Lewis [1986c], pp. 159–213.

——[1976] "Probabilities of conditionals and conditional probabilities", *Philosophical Review* 85, pp. 297–315; reprinted in Harper, Stalnaker and Pearce [1981], pp. 129–47; and with a postscript in Lewis [1986b], pp. 133–56.

——[1979] "Counterfactual dependence and time's arrow", *Noûs* 13, pp. 455–76; reprinted with a postscript in Lewis [1986b], pp. 32–66; and in Jackson [1991], pp. 46–75.

——[1983] *Philosophical Papers I*. Oxford University Press, Oxford.

——[1986a] *On The Plurality Of Worlds*. Blackwell, Oxford and New York.

——[1986b] *Philosophical Papers II*. Oxford University Press, Oxford.

——[1986c] "Probabilities of Conditionals and Conditional Probabilities II", *The Philosophical Review* 95, pp. 581–9.

Linsky, Leonard (ed.) [1971] *Reference and Modality*. Oxford University Press, Oxford.

——[1977] *Names and Descriptions*. University of Chicago Press, Chicago.

Loux, Michael, J. (ed.) [1979] *The Possible and the Actual: Readings in the Metaphysics of Modality*. Cornell University Press, Ithaca.

Lucas, John [1961] "Minds, machines and Gödel", *Philosophy* 36, pp. 112–27.

Lycan, William, G. [1984] *Logical Form in Natural Language*. MIT Press, Cambridge Mass.

Mackie, John, L. [1974] "*De* what *re* is *de re* modality?" *Journal of Philosophy* 71, pp. 551–61.

Mackie, Penelope [1987] "Essence, origin and bare identity", *Mind* 96, pp. 173–201.

Marsh, R. C. (ed.) [1956] *Logic and Knowledge*. Arrowsmith, Bristol (a collection of essays by Russell).

McCawley, James, D. [1981] *Everything that Linguists have Always Wanted to Know about Logic but were Ashamed to Ask*. University of Chicago Press, Chicago.

McCulloch, Gregory [1989] *The Game of the Name: Introducing Logic, Language and Mind*. Clarendon Press, Oxford.

McDowell, John [1977] "On the sense and reference of a proper name", *Mind* 86, pp. 159–85; reprinted in M. de B. Platts (ed.), *Reference, Truth and Reality*, Routledge and Kegan Paul, London (1980), pp. 141–66.

McGinn, Colin [1981] "Modal reality", in R. Healey (ed.), *Reduction, Time and Reality*, Cambridge University Press, Cambridge, pp. 143–205.

Meinong, Alexius [1904] "The theory of objects", in Meinong (ed.), *Untersuchungen zur Gegandstandtheorie und Psychologie*, Leipzig; English translation by Isaac Levi, D. B. Terrell and Roderick M. Chisholm, in Roderick M. Chisholm (ed.), *Realism and the Background of Phenomenology*, The Free Press, New York (1960).

Melia, Joseph [1992] "Against modalism", *Philosophical Studies* 68, pp. 35–56.

Mill, John Stuart [1879] *System of Logic*. Longmans, London.

Mondadori, Fabrizio [1983] "Counterpartese, counterpartese*, counterpartese$_D$" *Histoire, Epistémologie, Langage* 5, pp. 69–94.

Neale, Stephen [1990] *Descriptions*. MIT Press, Cambridge, Mass., London.

——[1995] "The philosophical significance of Gödel's slingshot", *Mind* 104, pp. 761–825.

——[2000] "On a milestone of empiricism", in Alex Orenstein and Petr Kotatko (eds), *Knowledge, Language and Logic*, Kluwer, Dordrecht, pp. 237–346.

Oliver, Alex [1999] "A few more remarks on logical form", *Proceedings of the Aristotelian Society* 99, pp. 247–72.

Pap, Arthur [1958] *Semantics and Necessary Truth*. Yale University Press, New Haven.

Parsons, Terence [1972] "Some problems concerning the logic of grammatical modifiers", in Donald Davidson and Gilbert Harman (eds), *Semantics of Natural Languages*, Reidel, Dordrecht.

——[1974]: "Prolegomena to Meinongian semantics", *Journal of Philosophy* 71, pp. 561–80.

Peacocke, Christopher A. B. [1973] "What is a logical constant?" *Journal of Philosophy* 73, pp. 221–40.

——[1975] "Proper names, reference and rigid designation", in S. Blackburn (ed.), *Meaning, Reference and Necessity*, Cambridge University Press, Cambridge, pp. 109–32.

——[1987] "Understanding logical constants: a realist's account", *Proceedings of the British Academy*, Oxford University Press, Oxford (1988), pp. 153–200.

Plantinga, Alvin [1974] *The Nature of Necessity*. Oxford University Press, Oxford.

——[1976] "Actualism and possible worlds", *Theoria* 42, pp. 139–60; reprinted in Loux [1979] pp. 253–73.

——[1987] "Two concepts of modality: modal realism and modal reductionism", in James E. Tomberlin (ed.), *Philosophical Perspectives, I: Metaphysics*, Ridgeview, Atascadero, Calif., pp. 189–231.

Platts, Mark de Breton [1979] *Ways of Meaning*. Routledge and Kegan Paul, London.

Pollock, John, L. [1982] *Language and Thought*. Princeton University Press, New Jersey.

——[1976] *Subjunctive Reasoning*. Reidel, Dordrecht.

Popper, Karl [1959] *The Logic of Scientific Discovery*. Hutchinson, London.

Prawitz, Dag [1965] *Natural Deduction*. Almqvist and Wiksell, Stockholm.

——[1977] "Meaning and proofs: on the conflict between classical and intuitionistic logic", *Theoria* 43, pp. 2–40.

——[1979] "Proofs and the meaning and completeness of the logical constants", in Jaakko Hintikka, Ilkka Niiniluoto and Esa Saarinen (eds), *Essays on Mathematical and Philosophical Logic*, Reidel, Dordrecht, pp. 25–40.

Prior, Arthur [1960] "The runabout inference ticket", *Analysis* 21, pp. 38–9.

Putnam, Hilary [1980] "Models and reality", *Journal of Symbolic Logic* 45, pp. 464–82; reprinted in his *Realism and Reason*, Cambridge University Press, Cambridge (1983), pp. 1–25.

Quine, Willard van, O. [1936] "Truth by convention", in O. H. Lee (ed.), *Philosophical Essays for A. N. Whitehead*, Longmans, New York (1936); reprinted in Quine [1966], pp. 70–99.

——[1940] *Mathematical Logic* Harvard University Press, Cambridge Mass.

——[1948] "On what there is", *Review of Metaphysics* 2, pp. 21–38; reprinted in Quine [1953a], pp. 1–19.

——[1951] "Two dogmas of empiricism", *Philosophical Review* 60, pp. 20–43; reprinted in Quine [1953a], pp. 20–46.

——[1952] *Methods of Logic*. Routledge and Kegan Paul, London.

——[1953a] *From A Logical Point of View*. Harvard University Press, Cambridge, Mass.

——[1953b] "Logic and the reification of universals", in Quine [1953a], pp. 102–29.

——[1953c] "Reference and Modality", in Quine [1953a], pp. 139–59.

——[1953d] "Three grades of modal involvement", *Proceedings of the XIth International Congress of Philosophy*, North Holland, Amsterdam, pp. 65–81; reprinted in Quine [1966], pp. 156–74.

——[1956] "Quantifiers and propositional attitudes", *Journal of Philosophy* 53; reprinted in Quine [1966], pp. 183–94.

——[1960] *Word and Object*. MIT Press, Cambridge, Mass.

——[1961] *From a Logical Point of View*. Harvard University Press, Cambridge, Mass.

——[1963] "Carnap and logical truth" in P. A. Schilpp (ed.), *The Philosophy of Rudolph Carnap*, Open Court, La Salle; reprinted in Quine [1966], pp. 100–25.

——[1966] *Ways of Paradox and Other Essays*. Random House, New York.

——[1968] "Ontological relativity", *Journal of Philosophy* 65, pp. 185–212; reprinted in Quine [1969b], pp. 26–68.

——[1969a] "Existence and quantification", in Quine [1969b], pp. 91–113.

——[1969b] *Ontological Relativity and Other Essays*. Columbia University Press, New York.

——[1970] *Philosophy of Logic*. Prentice-Hall, New Jersey.

——[1976] "Worlds Away", *Journal of Philosophy* 73, pp. 859–63.

Quinton, Anthony [1963] "The a priori and the analytic", *Proceedings of the Aristotelian Society* 64 (1963/4), pp. 331–54; reprinted in P. F. Strawson (ed.), *Philosophical Logic*, Oxford University Press, Oxford (1967), pp. 107–28.

Ramachandran, Murali [1989] "An alternative translation scheme for counterpart theory", *Analysis* 49, pp. 131–41.

Ramsey, Frank, P. [1929] "General propositions and causality", in *Foundations*, revised edition, Humanities Press, New Jersey (1978), pp. 133–51.

Read, Stephen [1988] *Relevant Logic*. Blackwell, Oxford.

——[1994] *Thinking About Logic: An Introduction to the Philosophy of Logic*. Oxford University Press, Oxford.

Recanati, François [1989] "The pragmatics of what is said", *Mind and Language* 4, pp. 295–329.

——[1993] *Direct Reference*. Blackwell, Oxford.

Robertson, Teresa [1998] "Possibilities and the arguments for origin essentialism", *Mind* 107, pp. 729–49.

Routley, R. [1970] "Non-existence does not exist", *Notre Dame Journal of Formal Logic* 11, pp. 289–320.

Russell, Bertrand, A. W. [1905] "On Denoting", *Mind* 14, pp. 479–93; reprinted in R. C. Marsh (ed.), *Logic and Knowledge*, George Allen and Unwin, London (1956), pp. 442–54.

Russell, Bertrand, A. W. [1908] "Mathematical logic as based on the theory of types", *American Journal of Mathematics* 30, pp. 222–62; reprinted in R. C. Marsh (ed.), *Logic and Knowledge*, George Allen and Unwin, London (1956), pp. 59–102.

——[1912] *Problems of Philosophy*, reprinted by Oxford University Press, Oxford, 1967.

——[1914] *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*, Allen and Unwin, London.

——[1918] "Lectures on the philosophy of logical atomism", *Monist* 28 (1918), pp. 495–527, and *Monist* 29 (1919), pp. 32–63, 190–222, 345–80; reprinted in R. C. Marsh (ed.), *Logic and Knowledge*, George Allen and Unwin, London (1956), pp. 177–281.

——[1919] *Introduction to Mathematical Philosophy*. Allen and Unwin, London and New York.

——[1927] "Introduction to second edition" of *Principia Mathematica*. Cambridge University Press, Cambridge.

——[1959] *My Philosophical Development*. Allen and Unwin, London.

Russell, Bertrand, A. W. and Whitehead, Alfred, N. [1910] *Principia Mathematica*. Cambridge University Press, Cambridge (1910–13).

Sainsbury, R. M. [1979] *Russell*. Routledge and Kegan Paul, London.

——[1993] "Russell on names and communication", in Andrew Irvine and Gary Wedekind (eds), *Russell and Analytic Philosophy*, University of British Columbia Press, Vancouver, pp. 3–21.

——[1996] "Philosophical logic", in A. C. Grayling (ed.), *Philosophy: A Guide Through the Subject*, Oxford University Press, Oxford, pp. 61–122.

Salmon, Nathan [1981] *Reference and Essence*. Princeton University Press, Princeton.

——[1989] "Reference and information content: names and descriptions", in Gabbay and Guenthner [1989], pp. 409–61.

Schiffer, Stephen [1987] *Remnants of Meaning*. MIT Press, Cambridge, Mass.

Schock, R. [1968] *Logic Without Existence Assumptions*. Almqvist and Wiksell, Stockholm.

Scott, Dana [1967] "Existence and description in formal logic", in Lambert [1991a], pp. 28–48; originally published in Ralph Schoenman (ed.), *Bertrand Russell: Philosopher of the Century*, Little & Brown, Boston, pp. 181–200.

Searle, John [1958] "Proper names", *Mind* 67, pp. 166–73; reprinted in P. F. Strawson (ed.), *Philosophical Logic*, Oxford University Press, Oxford (1967), pp. 89–96.

Segal, Gabriel [1989] "A preference for sense and reference", *Journal of Philosophy* 86, pp. 73–89.

Skyrms, Brian [1966] *Choice and Chance*. Dickenson, Belmont, Calif.

Smiley, Timothy [1982] "The schematic fallacy", *Proceedings of the Aristotelian Society* 82 (1982/3), pp. 1–17.

Stalnaker, Robert, L. [1968] "A theory of conditionals", *Studies in Logical Theory, American Philosophical Quarterly*, Monograph Series No. 2, Blackwell, Oxford, pp. 98–112; reprinted in Harper, Stalnaker and Pearce [1981], pp. 41–55; and in Jackson [1991], pp. 28–45.

——[1975] "Indicative conditionals", *Philosophia* 5, pp. 269–86; reprinted in Harper, Stalnaker and Pearce [1981], pp. 193–210; and in Jackson [1991], pp. 136–54.

——[1981] "A defense of conditional excluded middle" in Harper, Stalnaker and Pearce [1981], pp. 67–104.

——[1984] *Inquiry*. MIT Press, Cambridge, Mass.

Strawson, P. F. [1950] "On referring", *Mind* 59, pp. 320–44; reprinted in Strawson [1971], pp. 1–27.

——[1952] *Introduction to Logical Theory*. Methuen, London.

——[1971] *Logico-linguistic Papers*. Methuen, London.

——[1974] "Positions for quantifiers", in M. K. Munitz and P. Unger (eds), *Semantics and Philosophy*, New York University Press, New York, pp. 63–79.

Tarski, A. [1937] "The concept of truth in formalized languages", in his *Logic, Semantics, Metamathematics*, Clarendon Press, Oxford (1956), pp. 152–278.

Taylor, Barry [1985] *Modes of Occurrence*. Blackwell, Oxford.

Teichmann, Roger [1990] "'Actually'", *Analysis* 50, pp. 16–19.

Tennant, Neil [1978] *Natural Logic*. Edinburgh University Press, Edinburgh.

Thomson, James [1967] "Is existence a predicate?", in P. F. Strawson (ed.), *Philosophical Logic*, Oxford University Press, Oxford, pp. 103–6.

Travis, Charles [1996] "Meaning's role in truth", *Mind* 105, pp. 451–66.

Trollope, A. [1864] *Can You Forgive Her?* Oxford University Press, Oxford.

Urbach, Peter [1988] "What is a law of nature? A Humean answer", *British Journal for the Philosophy of Science* 39, pp. 193–210.

Van Benthem, Johan [1983] "Correspondence theory", in Gabbay and Guenthner [1983], pp. 167–247.

Van Benthem, Johan and Doets, Kees [1983] "Higher-order logic", in Gabbay and Guenthner [1983], pp. 275–329.

Van Dijck, Teun, A. [1977] *Text and Context*. Longman, London.

Van Fraassen, Bas [1977] "The only necessity is verbal necessity", *Journal of Philosophy* 84, pp. 71–85.

Van Inwagen, Peter [1985] "Plantinga on trans-world identity", in James Tomberlin and Peter Van Inwagen (eds), *Alvin Plantinga: A Profile*, Reidel, Dordrecht, pp. 101–20.

——[1986] "Two concepts of possible worlds", in P. French, T. Uehling and H. Wettstein (eds), *Midwest Studies in Philosophy Volume XI: Studies in Essentialism*, University of Minnesota Press, Minneapolis, pp. 185–213.

Warmbrõd, Ken [1999] "Logical constants", *Mind* 108, pp. 503–38.
Wiggins, David [1976] "The *de re* 'must': a note on the logical form of essentialist claims", in Evans and McDowell [1976], pp. 285–312.
——[1980a] "'Most' and 'all': some comments on a familiar programme and on the logical form of quantified sentences", in M. de B. Platts (ed.), *Reference, Truth and Reality*, Routledge and Kegan Paul, London (1980), pp. 318–46.
——[1980b] *Sameness and Substance*. Blackwell, Oxford.
——[1985] "Verbs and adverbs, and some other modes of grammatical combination", *Proceedings of the Aristotelian Society* 86 (1985/6), pp. 273–304.
Wittgenstein, L. [1921] *Tractatus Logico-Philosophicus*, translated by C. K. Ogden, Routledge and Kegan Paul, London.
——[1932] "Elementary propositions", section 4 of appendix to part I of *Philosophical Grammar*, ed. Rush Rhees, tr. Anthony Kenny, Blackwell, Oxford (1974), pp. 210–14.
Woods, Michael [1997] *Conditionals*. Clarendon Press, Oxford.
Wright, Crispin [1981] "Rule-following, objectivity and the theory of meaning", in S. Holtzman and C. Leich (eds), *Wittgenstein: To Follow a Rule*, Routledge and Kegan Paul, London, pp. 99–117.
——[1986] "Inventing logical necessity", in Jeremy Butterfield (ed.), *Language Mind and Logic*, Cambridge University Press, Cambridge.
——[1989] "Necessity, caution and scepticism", *Proceedings of the Aristotelian Society, Supplementary Volume* 63, pp. 203–38.
Yourgrau, Palle [1987] "The dead", *Journal of Philosophy* 84, pp. 84–101.

Index

Note: an asterisk before a page number indicates a glossary entry.